Progress in Neurobiology 95 (2011) 448–490

Contents lists available at SciVerse ScienceDirect

Progress in Neurobiology

journal homepage: www.elsevier.com/locate/pneurobio

The neuronal encoding of information in the brain

Edmund T. Rolls [a,*], Alessandro Treves [b,c]

[a] Oxford Centre for Computational Neuroscience, Oxford, UK
[b] SISSA, Cognitive Neuroscience, via Bonomea 265, 34136 Trieste, Italy
[c] NTNU, Centre for the Biology of Memory, Trondheim, Norway

ARTICLE INFO

Article history:
Received 17 December 2010
Received in revised form 3 August 2011
Accepted 15 August 2011
Available online 2 September 2011

Keywords:
Information representation
Vision
Synchrony
Oscillation
Firing rate code
Inferior temporal visual cortex
Shannon information theory
Distributed encoding

ABSTRACT

We describe the results of quantitative information theoretic analyses of neural encoding, particularly in the primate visual, olfactory, taste, hippocampal, and orbitofrontal cortex. Most of the information turns out to be encoded by the firing rates of the neurons, that is by the number of spikes in a short time window. This has been shown to be a robust code, for the firing rate representations of different neurons are close to independent for small populations of neurons. Moreover, the information can be read fast from such encoding, in as little as 20 ms. In quantitative information theoretic studies, only a little additional information is available in temporal encoding involving stimulus-dependent synchronization of different neurons, or the timing of spikes within the spike train of a single neuron. Feature binding appears to be solved by feature combination neurons rather than by temporal synchrony. The code is sparse distributed, with the spike firing rate distributions close to exponential or gamma. A feature of the code is that it can be read by neurons that take a synaptically weighted sum of their inputs. This dot product decoding is biologically plausible. Understanding the neural code is fundamental to understanding not only how the cortex represents, but also processes, information.

Contents

Abbreviation: IT, inferior temporal visual cortex.

* Corresponding author.

0301-0082/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.pneurobio.2011.08.002

1. Introduction

Because single neurons are the computing elements of the brain and send the results of their processing by spiking activity to other neurons, we can analyze brain processing by understanding what is encoded by the neuronal firing at each stage of the brain (e.g. each cortical area), and determining how what is encoded changes from stage to stage. Each neuron responds differently to a set of stimuli (with each neuron tuned differently to the members of the set of stimuli), and it is this that allows different stimuli to be represented. We can only address the richness of the representation therefore by understanding the differences in the responses of different neurons, and the impact that this has on the amount of information that is encoded. These issues can only be adequately and directly addressed at the level of the activity of single neurons and of populations of single neurons, and understanding at this neuronal level (rather than at the level of thousands or millions of neurons as revealed by functional neuroimaging) is essential for understanding brain computation.

Information theory provides the means for quantifying how much neurons communicate to other neurons, and thus provides a quantitative approach to fundamental questions about information processing in the brain. To investigate what in neuronal activity carries information, one must compare the amounts of information carried by different codes, that is different descriptions of the same activity, to provide the answer. To investigate the speed of information transmission, one must define and measure information rates from neuronal responses. To investigate to what extent the information provided by different cells is redundant or instead independent, again one must measure amounts of information in order to provide quantitative evidence. To compare the information carried by the number of spikes, by the timing of the spikes within the response of a single neuron, and by the relative time of firing of different neurons reflecting for example stimulus-dependent neuronal synchronization, information theory again provides a quantitative and well-founded basis for the necessary comparisons. To compare the information carried by a single neuron or a group of neurons with that reflected in the behaviour of the human or animal, one must again use information theory, as it provides a single measure which can be applied to the measurement of the performance of all these different cases. In all these situations, there is no quantitative and well-founded alternative to information theory.

The overall aim of this paper is to describe the methods used for the analysis of neuronal activity in primates and other mammals, and to describe the main principles that have been discovered to date about the representation of information in the primate brain. Although there have been descriptions of some of the methods used to analyze cortical population encoding (Rolls et al., 1997b; Franco et al., 2004; Quian Quiroga and Panzeri, 2009), this is the first paper we know that provides a comprehensive account of the principles of information encoding by single neurons and populations of neurons in the mammalian and particularly primate cortex, together with the methods used to make these discoveries. We focus on work on the primate to make the findings very relevant to understanding neuronal encoding in the human brain; because primates can be trained to maintain visual fixation and attention in a way that allows reliable and repeated presentation of

stimuli, which is essential for information theoretic analysis; and because in primates it has been possible to analyze neuronal activity in connected brain areas in order to understand the difference in the representation at different stages of cortical processing, and in different sensory pathways (Rolls, 2008).

This paper first summarizes information theory used in the analysis of the responses of neurons in the primate brain. Information theory, developed by Shannon (1948), is presented formally elsewhere (Cover and Thomas, 1991; Hamming, 1990), and further descriptions of its use in the analysis of neuronal firing are provided by Rolls (2008), Quian Quiroga and Panzeri (2009) and Rieke et al. (1997). In this paper we focus on its use for the analysis of neuronal activity in primate brains. One reason is that we are especially interested in situations in which large numbers of neurons are involved in representing stimuli using a distributed code (in contrast to many invertebrates in which the focus is more on the information conveyed by individual specialized neurons (Rieke et al., 1997)). A second reason is that primates (in contrast to rodents) have a highly developed cortical visual system and an ability to perform attentional and visual search tasks similar to those performed by humans, so that answers to how information is represented in systems similar to those in humans can be obtained. Moreover, even the taste system is connected and therefore probably operates differently in primates and rodents (Rolls, 2008), and the hippocampus appears to contain different types of neurons in primates (Rolls et al., 1998; Rolls and Xiang, 2006), so we include analyses of the representation of information in these systems too in primates.

After reviewing the basic methodology for extracting information measures in the next section, the main findings on neuronal encoding, as well as some specialized methods, are described in Section 3, and the main conclusions are described in Section 4.

2. Information theory and its applications to the analysis of neural activity

2.1. Information channels and information measures

Let us consider an information *channel* that receives symbols s from an alphabet S and emits symbols s' from alphabet S'. The **mutual information** transmitted by the channel can be expressed by

$$I = \sum_{s} P(s) \sum_{s'} P(s'|s) \log_2 \frac{P(s'|s)}{P(s')} \tag{1}$$

$$= \sum_{s,s'} P(s,s') \log_2 \frac{P(s,s')}{P(s)P(s')}.$$

The **mutual information** can also be expressed as the entropy of the source using alphabet S minus the *equivocation* of S with respect to the new alphabet S' used by the channel, written

$$I = H(S) - H(S|S') \equiv H(S) - \sum_{s'} P(s') H(S|s'). \tag{2}$$

The *capacity* of the channel can be defined as the maximal mutual information across all possible sets of input probabilities $P(s)$.

2.2. The information carried by a neuronal response and its average

Considering the processing of information in the brain, we are often interested in the amount of information the response r of a neuron, or of a population of neurons, carries about an event happening in the outside world, for example a stimulus s shown to the animal. Once the inputs and outputs are conceived of as sets of symbols from two alphabets, the neuron(s) may be regarded as an information channel. We may denote with $P(s)$ the *a priori* probability that the particular stimulus s out of a given set was shown, while the conditional probability $P(s|r)$ is the *a posteriori* probability, that is updated by the knowledge of the response r. The Kullback–Leibler distance between these two probability distributions can be defined as the response-specific transinformation

$$I(r) = \sum_{s} P(s|r) \log_2 \frac{P(s|r)}{P(s)}, \tag{3}$$

which takes the extreme values of $I(r) = -\log_2 P(s(r))$ if r unequivocally determines $s(r)$ (that is, $P(s|r)$ equals 1 for that one stimulus and 0 for all others); and $I(r) = \sum_s P(s) \log_2 (P(s)/P(s)) = 0$ if there is no relation between s and r, that is they are independent, so that the response tells us nothing new about the stimulus and thus $P(s|r) = P(s)$.

This positive-definite quantity is one possible definition of the information conveyed by each particular response. One is usually interested in further averaging this quantity over all possible responses r,

$$\langle I \rangle = \sum_{r} P(r) \left[\sum_{s} P(s|r) \log_2 \frac{P(s|r)}{P(s)} \right]. \tag{4}$$

The angular brackets $\langle\rangle$ are used here to emphasize the averaging operation, in this case over responses. Denoting with $P(s, r)$ the *joint probability* of the pair of events s and r, and using Bayes' theorem, this reduces to the symmetric form (Eq. (1)) for the **mutual information** $I(S, R)$

$$\langle I \rangle = \sum_{s,r} P(s,r) \log_2 \frac{P(s,r)}{P(s)P(r)} \tag{5}$$

which emphasizes that responses tell us about stimuli just as much as stimuli tell us about responses. This is, of course, a general feature, independent of the two variables being in this instance stimuli and neuronal responses.

In fact, what is of interest, besides the mutual information of Eqs. (4) and (5), is often the information specifically conveyed about each stimulus, which can be defined, symmetrically to Eq. (3), as

$$I(s) = \sum_{r} P(r|s) \log_2 \frac{P(r|s)}{P(r)}. \tag{6}$$

This quantity, sometimes written $I(s, R)$ to draw attention to the fact that it is calculated across the full set of responses R, is again the positive-definite Kullback–Leibler divergence between two probability distributions. It has also been called the stimulus-specific surprise (DeWeese and Meister, 1999) to emphasize its being always a positive number. An alternative definition of the stimulus-specific information is additive rather than positive, but both definitions once averaged across stimuli yield the mutual information $I(S, R)$, which is both positive and additive.

All these information measures quantify the variability in the responses elicited by the stimuli, compared to the overall variability. Since $P(r)$ is the probability distribution of responses averaged across stimuli, it is evident that, for example, the stimulus-specific information measure of Eq. (6) depends not only on the stimulus s, but also on all other stimuli used. Likewise, the mutual information measure, despite being of an average nature, is dependent on what set of stimuli has been used in the average. This emphasizes again the relative nature of all information measures. More specifically, it underscores the relevance of using, while measuring the information conveyed by a given neuronal population, stimuli that are either representative of real-life stimulus statistics, or are of particular interest for the properties of the population being examined.

2.3. Quantifying the speed of information transfer

In Section 3 we shall discuss temporal aspects of neuronal codes observed in primates, including when they can be described as temporally modulated codes in contrast to plain firing rate codes. It is intuitive that if short periods of firing of single cells are considered, there is less time for temporal modulation effects. The information conveyed about stimuli by the firing rate and that conveyed by more detailed temporal codes become similar in value. When the firing periods analyzed become shorter than roughly the mean interspike interval, even the statistics of firing rate values on individual trials cease to be relevant, and the information content of the firing depends solely on the mean firing rates across all trials with each stimulus. This is expressed mathematically by considering the amount of information provided as a function of the length t of the time window over which firing is analyzed, and taking the limit for $t \to 0$ (Skaggs et al., 1993; Panzeri et al., 1996). To first order in t, only two responses can occur in a short window of length t: either the emission of an action potential, with probability tr_s, where r_s is the mean firing rate calculated over many trials using the same window and stimulus; or no action potential, with probability $1 - tr_s$. Inserting these conditional probabilities into Eq. (6), taking the limit and dividing by t, one obtains for the derivative of the stimulus-specific transinformation

$$\frac{dI(s)}{dt} = r_s \log_2\left(\frac{r_s}{\langle r\rangle}\right) + \frac{\langle r\rangle - r_s}{\ln 2}, \tag{7}$$

where $\langle r\rangle$ is the grand mean rate across stimuli. This formula thus gives the rate, in bits/s, at which information about a stimulus begins to accumulate when the firing of a cell is recorded. Such an information rate depends only on the mean firing rate to that stimulus and on the grand mean rate across stimuli. As a function of r_s, it follows the U-shaped curve in Fig. 1.

The curve is universal, in the sense that it applies irrespective of the detailed firing statistics of the cell, and it expresses the fact that the emission or not of a spike in a short window conveys information in as much as the mean response to a given stimulus is above or below the overall mean rate. No information is conveyed, over short times, about those stimuli the mean response to which is the same as the overall mean. In practice, although the curve describes only the universal behaviour of the initial slope of the specific information as a function of time, it approximates well the full stimulus-specific information $I(s, R)$ computed even over rather long periods (Rolls et al., 1996, 1997c).

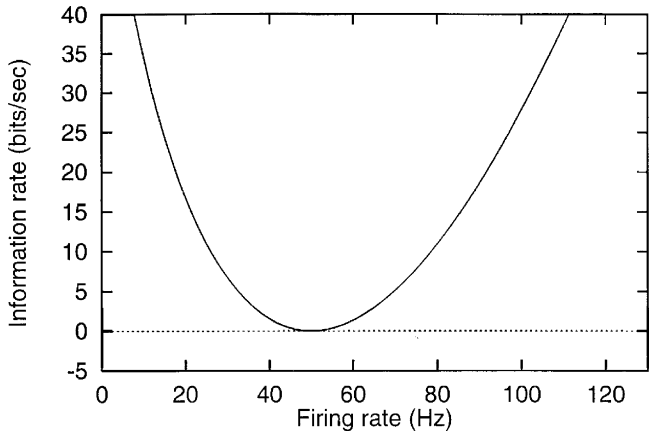

Fig. 1. Time derivative of the stimulus-specific information as a function of firing rate, for a cell firing at a grand mean rate of 50 Hz. For different grand mean rates, the graph would simply be rescaled.

Averaging Eq. (7) across stimuli one obtains the time derivative of the mutual information. Further dividing by the overall mean rate yields the adimensional quantity

$$\chi = \sum_s P(s)\left(\frac{r_s}{\langle r\rangle}\right)\log_2\left(\frac{r_s}{\langle r\rangle}\right) \tag{8}$$

which measures, in bits, the mutual information per spike provided by the cell (Bialek et al., 1991; Skaggs et al., 1993). One can prove that this quantity can range from 0 to $\log_2(1/a)$

$$0 < \chi < \log_2\left(\frac{1}{a}\right), \tag{9}$$

where a is the single neuron sparseness a^s defined in Section 3.1.1. For mean rates r_s distributed in a nearly binary fashion, χ is close to its upper limit $\log_2(1/a)$, whereas for mean rates that are nearly uniform, or at least unimodally distributed, χ is relatively close to zero (Panzeri et al., 1996). In practice, whenever a large number of more or less 'ecological' stimuli are considered, mean rates are not distributed in arbitrary ways, but rather tend to follow stereotyped distributions (which for some neurons approximate an exponential distribution of firing rates – see Section 3.1 (Treves et al., 1999b; Baddeley et al., 1997; Rolls and Treves, 1998; Rolls and Deco, 2002; Franco et al., 2007)), and as a consequence χ and a (or, equivalently, its logarithm) tend to covary (rather than to be independent variables (Skaggs and McNaughton, 1992)). Therefore, measuring sparseness is in practice nearly equivalent to measuring information per spike, and the rate of rise in mutual information, $\chi\langle r\rangle$, is largely determined by the sparseness a and the overall mean firing rate $\langle r\rangle$.

The important point to note about the single-cell information rate $\chi\langle r\rangle$ is that, to the extent that different cells express non-redundant codes, as discussed below, the instantaneous *information flow* across a population of C cells can be taken to be simply $C\chi\langle r\rangle$, and this quantity can easily be measured directly without major limited sampling biases, or else inferred indirectly through measurements of the sparseness a. Values for the information rate $\chi\langle r\rangle$ that have been published range from 2 to 3 bits/s for rat hippocampal cells (Skaggs et al., 1993), to 10–30 bits/s for primate temporal cortex visual cells (Rolls et al., 1997b), and could be compared with analogous measurements in the sensory systems of frogs and crickets, in the 100–300 bits/s range (Rieke et al., 1993).

If the first time-derivative of the mutual information measures information flow, successive derivatives characterize, at the single-cell level, different firing modes. This is because whereas the first derivative is universal and depends only on the mean firing rates to each stimulus, the next derivatives depend also on the variability of the firing rate around its mean value, across trials, and take different forms in different firing regimes. Thus they can serve as a measure of discrimination among firing regimes with limited variability, for which, for example, the second derivative is large and positive, and firing regimes with large variability, for which the second derivative is large and negative. Poisson firing, in which in every short period of time there is a fixed probability of emitting a spike irrespective of previous firing, is an example of large variability, and the second derivative of the mutual information can be calculated to be

$$\frac{d^2I}{dt^2} = \frac{[\ln a + (1-a)]\langle r\rangle^2}{(a\ln 2)}, \tag{10}$$

where a is the single neuron sparseness a^s defined in Section 3.1.1. This quantity is always negative. Strictly periodic firing is an example of zero variability, and in fact the second time-derivative of the mutual information becomes infinitely large in this case (although actual information values measured in a short time interval remain of course finite even for exactly periodic firing,

because there is still some variability, ±1, in the number of spikes recorded in the interval). Measures of mutual information from short intervals of firing of temporal cortex visual cells have revealed a degree of variability intermediate between that of periodic and of Poisson regimes (Rolls et al., 1997c). Similar measures can also be used to contrast the effect of the graded nature of neuronal responses, once they are analyzed over a finite period of time, with the information content that would characterize neuronal activity if it reduced to a binary variable (Panzeri et al., 1996). A binary variable with the same degree of variability would convey information at the same instantaneous rate (the first derivative being universal), but in for example 20–30% reduced amounts when analyzed over times of the order of the interspike interval or longer.

2.4. The limited sampling problem

With real neurophysiological data, because we typically have limited numbers of trials, it is difficult from the *frequency* of each possible neuronal response to accurately estimate its *probability*, and this limits our ability to estimate $\langle I \rangle$ correctly. We refer to this as the limited sampling problem. To elaborate, if the responses are continuous quantities, the probability of observing exactly the same response twice is infinitesimal. In the absence of further manipulation, this would imply that each stimulus generates its own set of unique responses, therefore any response that has actually occurred could be associated unequivocally with one stimulus, and the mutual information would always equal the entropy of the stimulus set. This absurdity shows that in order to estimate probability densities from experimental frequencies, one has to resort to some *regularizing* manipulation, such as smoothing the point-like response values by convolution with suitable kernels, or binning them into a finite number of discrete bins.

2.4.1. Smoothing or binning neuronal response data

The issue is how to estimate the underlying probability distributions of neuronal responses to a set of stimuli from only a limited number of trials of data (e.g. 10–30) for each stimulus. Several strategies are possible. One is to discretize the response space into bins, and estimate the probability density as the histogram of the fraction of trials falling into each bin. If the bins are too narrow, almost every response is in a different bin, and the estimated information will be overestimated. Even if the bin width is increased to match the standard deviation of each underlying distribution, the information may still be overestimated. Alternatively, one may try to 'smooth' the data by convolving each response with a Gaussian with a width set to the standard deviation measured for each stimulus. Setting the standard deviation to this value may actually lead to an underestimation of the amount of information available, due to oversmoothing. Another possibility is to make a bold assumption as to what the general shape of the underlying densities should be, for example a Gaussian. This may produce closer estimates. Methods for regularizing the data are discussed further by Rolls and Treves (1998) in their Appendix A2, where a numerical example is given.

2.4.2. The effects of limited sampling

The crux of the problem is that, whatever procedure one adopts, limited sampling tends to produce distortions in the estimated probability densities. The resulting mutual information estimates are intrinsically biased. The bias, or average error of the estimate, is upward if the raw data have not been regularized much, and is downward if the regularization procedure chosen has been heavier. The bias can be, if the available trials are few, much larger than the true information values themselves. This is intuitive, as fluctuations due to the finite number of trials available would tend, on average, to either produce or emphasize differences among the distributions corresponding to different stimuli, differences that are preserved if the regularization is 'light', and that are interpreted in the calculation as carrying genuine information. This is illustrated with a quantitative example by Rolls and Treves (1998) in their Appendix A2.

Choosing the right amount of regularization, or the best regularizing procedure, is not possible *a priori*. Hertz et al. (1992) have proposed the interesting procedure of using an artificial neural network to regularize the raw responses. The network can be trained on part of the data using backpropagation, and then used on the remaining part to produce what is in effect a clever data-driven regularization of the responses. This procedure is, however, rather computer intensive and not very safe, as shown by some self-evident inconsistency in the results (Heller et al., 1995). Obviously, the best way to deal with the limited sampling problem is to try and use as many trials as possible. The improvement is slow, however, and generating as many trials as would be required for a reasonably unbiased estimate is often, in practice, impossible.

2.4.3. Correction procedures for limited sampling

The above point, that data drawn from a single distribution, when artificially paired, at random, to different stimulus labels, results in 'spurious' amounts of apparent information, suggests a simple way of checking the reliability of estimates produced from real data (Optican et al., 1991). One can disregard the true stimulus associated with each response, and generate a randomly reshuffled pairing of stimuli and responses, which should therefore, being not linked by any underlying relationship, carry no mutual information about each other. Calculating, with some procedure of choice, the spurious information obtained in this way, and comparing with the information value estimated with the same procedure for the real pairing, one can get a feeling for how far the procedure goes into eliminating the apparent information due to limited sampling. Although this spurious information, I_s, is only indicative of the amount of bias affecting the original estimate, a simple heuristic trick (called 'bootstrap'[1]) is to subtract the spurious from the original value, to obtain a somewhat 'corrected' estimate. This procedure can result in quite accurate estimates (see Rolls and Treves (1998), Tovee et al. (1993))[2].

A different correction procedure (called 'jack-knife') is based on the assumption that the bias is proportional to $1/N$, where N is the number of responses (data points) used in the estimation. One computes, beside the original estimate $\langle I_N \rangle$, N auxiliary estimates $\langle I_{N-1} \rangle_k$, by taking out from the data set response k, where k runs across the data set from 1 to N. The corrected estimate

$$\langle I \rangle = N\langle I_N \rangle - \frac{1}{N}\sum_k (N-1)\langle I_{N-1} \rangle_k \tag{11}$$

is free from bias (to leading order in $1/N$), if the proportionality factor is more or less the same in the original and auxiliary estimates. This procedure is very time-consuming, and it suffers from the same imprecision of any algorithm that tries to determine a quantity as the result of the subtraction of two large and nearly equal terms; in this case the terms have been made large on purpose, by multiplying them by N and $N-1$.

A more fundamental approach (Miller, 1955) is to derive an analytical expression for the bias (or, more precisely, for its leading

[1] In technical usage bootstrap procedures utilize random pairings of responses with stimuli with replacement, while shuffling procedures utilize random pairings of responses with stimuli without replacement.

[2] Subtracting the 'square' of the spurious fraction of information estimated by this bootstrap procedure as used by Optican et al. (1991) is unfounded and does not work correctly (see Rolls and Treves (1998) and Tovee et al. (1993)).

terms in an expansion in 1/*N*, the inverse of the sample size). This allows the estimation of the bias from the data itself, and its subsequent subtraction, as discussed in Treves and Panzeri (1995) and Panzeri and Treves (1996). Such a procedure produces satisfactory results, thereby lowering the size of the sample required for a given accuracy in the estimate by about an order of magnitude (Golomb et al., 1997). However, it does not, in itself, make possible measures of the information contained in very complex responses with few trials. As a rule of thumb, the number of trials per stimulus required for a reasonable estimate of information, once the subtractive correction is applied, is of the order of the effectively independent (and utilized) bins in which the response space can be partitioned (Panzeri and Treves, 1996). This correction procedure is the one that we use standardly (Rolls et al., 1997c, 1996, 1998, 1999, 2006, 2010a; Booth and Rolls, 1998).

2.5. The information from multiple cells: decoding procedures

2.5.1. Decoding

The bias of information measures grows with the dimensionality of the response space, and for all practical purposes the limit on the number of dimensions that can lead to reasonably accurate direct measures, even when applying a correction procedure, is quite low, two to three. This implies, in particular, that it is not possible to apply equation 5 to extract the information content in the responses of several cells (more than two to three) recorded simultaneously. One way to address the problem is then to apply some strong form of regularization to the multiple cell responses. Smoothing has already been mentioned as a form of regularization that can be tuned from very soft to very strong, and that preserves the structure of the response space. Binning is another form, which changes the nature of the responses from continuous to discrete, but otherwise preserves their general structure, and which can also be tuned from soft to strong. Other forms of regularization involve much more radical transformations, or changes of variables.

Of particular interest for information estimates is a change of variables that transforms the response space into the stimulus set, by applying an algorithm that derives a predicted stimulus from the response vector, i.e. the firing rates of all the cells, on each trial. Applying such an algorithm is called decoding. Of course, the predicted stimulus is not necessarily the same as the actual one. Therefore the term decoding should not be taken to imply that the algorithm works successfully, each time identifying the actual stimulus. The predicted stimulus is simply a function of the response, as determined by the algorithm considered. Just as with any regularizing transform, it is possible to compute the mutual information between actual stimuli s and predicted stimuli s', instead of the original one between stimuli s and responses r. Since information about (real) stimuli can only be lost and not be created by the transform, the information measured in this way is bound to be lower in value than the real information in the responses. If the decoding algorithm is efficient, it manages to preserve nearly all the information contained in the raw responses, while if it is poor, it loses a large portion of it. If the responses themselves provided all the information about stimuli, and the decoding is optimal, then predicted stimuli coincide with the actual stimuli, and the information extracted equals the entropy of the stimulus set.

The procedure for extracting information values after applying a decoding algorithm is indicated in Fig. 2 (in which s? is s'). The underlying idea indicated in Fig. 2 is that if we know the average firing rate of each cell in a population to each stimulus, then on any single trial we can guess (or decode) the stimulus that was present by taking into account the responses of all the cells. The decoded stimulus is s', and the actual stimulus that was shown is s. [What

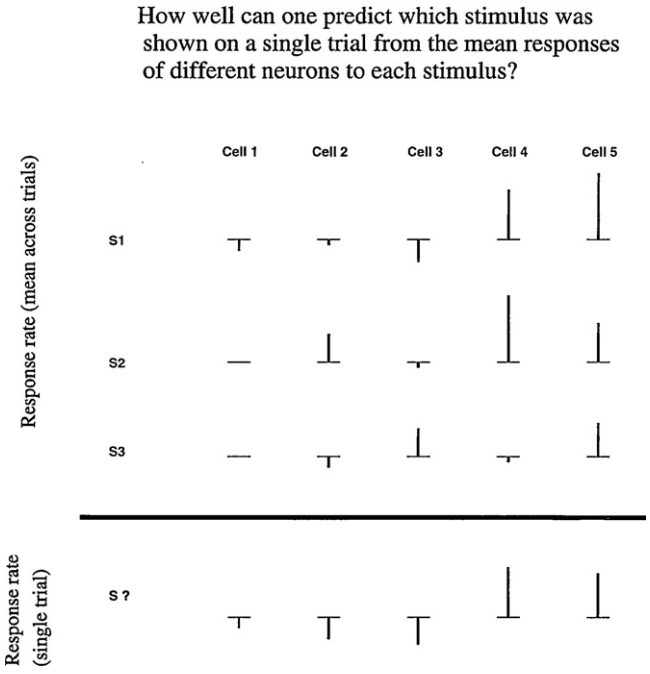

Fig. 2. This diagram shows the average response for each of several cells (Cell 1, etc.) to each of several stimuli (S1, etc.). The change of firing rate from the spontaneous rate is indicated by the vertical line above or below the horizontal line, which represents the spontaneous rate. We can imagine guessing or predicting from such a table the predicted stimulus S? (i.e. s') that was present on any one trial.

we wish to know is how the percentage correct, or better still the information, based on the evidence from any single trial about which stimulus was shown, increases as the number of cells in the population sampled increases. We can expect that the more cells there are in the sample, the more accurate the estimate of the stimulus is likely to be. If the encoding was local, the number of stimuli encoded by a population of neurons would be expected to rise approximately linearly with the number of neurons in the population. In contrast, with distributed encoding, provided that the neuronal responses are sufficiently independent, and are sufficiently reliable (not too noisy), information from the ensemble would be expected to rise linearly with the number of cells in the ensemble, and (as information is a log measure) 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 procedure is schematized in Table 1 where the double arrow indicates the transformation from stimuli to responses operated by the nervous system, while the single arrow indicates the further transformation operated by the decoding procedure. $I(s, s')$ is the mutual information between the actual stimuli s and the stimuli s' that are predicted to have been shown based on the decoded responses. The decoding procedure just described is called maximum likelihood decoding, because only the most likely stimulus on a given trial given the responses of the neurons on that

Table 1
Decoding. s' is the decoded stimulus, i.e. that predicted from the neuronal responses r.

$$s \quad \Rightarrow \quad r \quad \rightarrow \quad s'$$
$$I(s,r) \qquad\qquad I(s,s')$$

trial is decoded. We refer to the information estimated with this procedure as I_{ml}.

A slightly more complex variant of this procedure is a decoding step that extracts from the response on each trial not a single predicted stimulus, but rather probabilities that each of the possible stimuli was the actual one. The joint probabilities of actual and posited stimuli can be averaged across trials, and information computed from the resulting probability matrix ($S \times S$). Computing information in this way takes into account the relative uncertainty in assigning a predicted stimulus to each trial, an uncertainty that is instead not considered by the previous procedure based solely on the identification of the maximally likely stimulus (Treves, 1997). *Maximum likelihood* information values I_{ml} based on a single stimulus tend therefore to be higher than *probability* information values I_p based on the whole set of stimuli, although in very specific situations the reverse could also be true.

The same correction procedures for limited sampling can be applied to information values computed after a decoding step. Values obtained from maximum likelihood decoding, I_{ml}, suffer from limited sampling more than those obtained from probability decoding, I_p, since each trial contributes a whole 'brick' of weight $1/N$ (N being the total number of trials), whereas with probabilities each brick is shared among several slots of the ($S \times S$) probability matrix. The neural network procedure devised by Hertz et al. (1992) can in fact be thought of as a decoding procedure based on probabilities, which deals with limited sampling not by applying a correction but rather by strongly regularizing the original responses.

When decoding is used, the rule of thumb becomes that the minimal number of trials per stimulus required for accurate information measures is roughly equal to the size of the stimulus set, if the subtractive correction is applied (Panzeri and Treves, 1996). This correction procedure is applied as standard in our multiple cell information analyses that use decoding (Rolls et al., 1997b, 1998, 2006, 2009, 2010a; Booth and Rolls, 1998; Franco et al., 2004; Aggelopoulos et al., 2005).

2.5.2. Decoding algorithms

Any transformation from the response space to the stimulus set could be used in decoding, but of particular interest are the transformations that either approach optimality, so as to minimize information loss and hence the effect of decoding, or else are implementable by mechanisms that *could* conceivably be operating in the brain, so as to extract information values that could be extracted by the brain itself.

The optimal transformation is in theory well-defined: one should estimate from the data the conditional probabilities $P(r|s)$, and use Bayes' rule (see Glossary in Section 5) to convert them into the conditional probabilities $P(s'|r)$. Having these for any value of r, one could use them to estimate I_p, and, after selecting for each particular real response the stimulus with the highest conditional probability, to estimate I_{ml}. To avoid biasing the estimation of conditional probabilities, the responses used in estimating $P(r|s)$ should not include the particular response for which $P(s'|r)$ is going to be derived (jack-knife cross-validation). In practice, however, the estimation of $P(r|s)$ in usable form involves the fitting of some simple function to the responses. This need for fitting, together with the approximations implied in the estimation of the various quantities, prevents us from defining the really optimal decoding, and leaves us with various algorithms, depending essentially on the fitting function used, which are hopefully close to optimal in some conditions. We have experimented extensively with two such algorithms, which both approximate Bayesian decoding (Rolls et al., 1997b). Both these algorithms fit the response vectors produced over several trials by the cells being recorded to a product of conditional probabilities for the response of each cell given the stimulus. In one case, the single cell conditional probability is assumed to be Gaussian (truncated at zero); in the other it is assumed to be Poisson (with an additional weight at zero). Details of these algorithms are given by Rolls et al. (1997b).

Biologically plausible decoding algorithms are those that limit the algebraic operations used to types that could be easily implemented by neurons, e.g. dot product summations, thresholding and other single-cell non-linearities, and competition and contrast enhancement among the outputs of nearby cells. There is then no need for ever fitting functions or other sophisticated approximations, but of course the degree of arbitrariness in selecting a particular algorithm remains substantial, and a comparison among different choices based on which yields the higher information values may favour one choice in a given situation and another choice with a different data set.

To summarize, the key idea in decoding, in our context of estimating information values, is that it allows substitution of a possibly very high-dimensional response space (which is difficult to sample and regularize) with a reduced object much easier to handle, that is with a discrete set equivalent to the stimulus set. The mutual information between the new set and the stimulus set is then easier to estimate even with limited data, that is with relatively few trials. For each response recorded, one can use all the responses except for that one to generate estimates of the average response vectors (the average response for each neuron in the population) to each stimulus. Then one considers how well the selected response vector matches the average response vectors, and uses the degree of matching to estimate, for all stimuli, the probability that they were the actual stimuli. The form of the matching embodies the general notions about population encoding, for example the 'degree of matching' might be simply the dot product between the current vector and the average vector ($\mathbf{r}^{av}$), suitably normalized over all average vectors to generate probabilities

$$P(s'|\mathbf{r}(s)) = \frac{\mathbf{r}(s) \cdot \mathbf{r}^{av}(s')}{\sum_{s''} \mathbf{r}(s) \cdot \mathbf{r}^{av}(s'')} \tag{12}$$

where s'' is a dummy variable. (This is called dot product decoding.) One ends up, then, with a table of conjoint probabilities $P(s, s')$, and another table obtained by selecting for each trial the most likely (or predicted) single stimulus s^p, $P(s, s^p)$. Both s' and s^p stand for all possible stimuli, and hence belong to the same set S. These can be used to estimate mutual information values based on probability decoding (I_p) and on maximum likelihood decoding (I_{ml}):

$$\langle I_p \rangle = \sum_{s \in S} \sum_{s' \in S} P(s, s') \log_2 \frac{P(s, s')}{P(s)P(s')} \tag{13}$$

and

$$\langle I_{ml} \rangle = \sum_{s \in S} \sum_{s^p \in S} P(s, s^p) \log_2 \frac{P(s, s^p)}{P(s)P(s^p)}. \tag{14}$$

Examples of the use of these procedures are available (Rolls et al., 1997b, 1998, 2004, 2006; Booth and Rolls, 1998; Franco et al., 2004; Ince et al., 2010b), and some of the results obtained are described in Section 3.

2.6. Information in the correlations between the spikes of different cells

Simultaneously recorded neurons sometimes shows cross-correlations in their firing, that is the firing of one is systematically related to the firing of the other cell. One example of this is neuronal response synchronization. The cross-correlation, to be defined below, shows the time difference between the cells at which the systematic relation appears. A significant peak or trough in the cross-correlation function could reveal a synaptic connection from one cell

to the other, or a common input to each of the cells, or any of a considerable number of other possibilities. If the synchronization occurred for only some of the stimuli, then the presence of the significant cross-correlation for only those stimuli could provide additional evidence separate from any information in the firing rate of the neurons about which stimulus had been shown. Information theory in principle provides a way of quantitatively assessing the relative contributions from these two types of encoding, by expressing what can be learned from each type of encoding in the same units, bits of information.

Fig. 3 illustrates how synchronization occurring only for some of the stimuli could be used to encode information about which stimulus was presented. In the figure the spike trains of three neurons are shown after the presentation of two different stimuli on one trial. As shown by the cross-correlogram in the lower part of the figure, the responses of cell 1 and cell 2 are synchronized when stimulus 1 is presented, as whenever a spike from cell 1 is emitted, another spike from cell 2 is emitted after a short time lag. In contrast, when stimulus 2 is presented, synchronization effects do not appear. Thus, based on a measure of the synchrony between the responses of cells 1 and 2, it is possible to obtain some information about what stimulus has been presented. The contribution of this effect is measured as the stimulus-dependent synchronization information. Cells 1 and 2 have no information about what stimulus was presented from the number of spikes, as the same number is found for both stimuli. Cell 3 carries information in the spike count in the time window (which is also called the firing rate) about what stimulus was presented. (Cell 3 emits 6 spikes for stimulus 1 and 3 spikes for stimulus 2.)

2.6.1. A decoding approach

The example shown in Fig. 3 is for the neuronal responses on a single trial. Given that the neuronal responses are variable from trial to trial, we need a method to quantify the information that is gained from a single trial of spike data in the context of the measured variability in the responses of all of the cells, including how the cells' responses covary in a way that may be partly stimulus-dependent, and may include synchronization effects. The direct approach is to apply the Shannon mutual information measure (Shannon, 1948; Cover and Thomas, 1991)

$$I(s,\mathbf{r}) = \sum_{s\in S}\sum_{\mathbf{r}} P(s,\mathbf{r})\log_2 \frac{P(s,\mathbf{r})}{P(s)P(\mathbf{r})}, \tag{15}$$

where $P(s, \mathbf{r})$ is a probability table embodying a relationship between the variable s (here, the stimulus) and $\mathbf{r}$ (a vector where each element is the firing rate of one neuron).

However, because the probability table of the relation between the neuronal responses and the stimuli, $P(s, \mathbf{r})$, is so large (given that there may be many stimuli, and that the response space which has to include spike timing is very large), in practice it is difficult to obtain a sufficient number of trials for every stimulus to generate the probability table accurately, at least with data from mammals in which the experiment cannot usually be continued for many hours of recording from a whole population of cells. To circumvent this undersampling problem, Rolls et al. (1997b) developed a decoding procedure (described in Section 2.5), and a similar decoding process is used when measuring the information conveyed by cross-correlations between neurons, as described next.

The new step taken by Franco et al. (2004) is to introduce into the Table Data(s, r) shown in the upper part of Fig. 4 new columns, shown on the right of the diagram, containing a measure of the cross-correlation (averaged across trials in the upper part of the table) for some pairs of cells (labelled as Corrln Cells 1–2 and 2–3). The decoding procedure can then take account of any cross-correlations between pairs of cells, and thus measure any contributions to the information from the population of cells that arise from cross-correlations between the neuronal responses. If

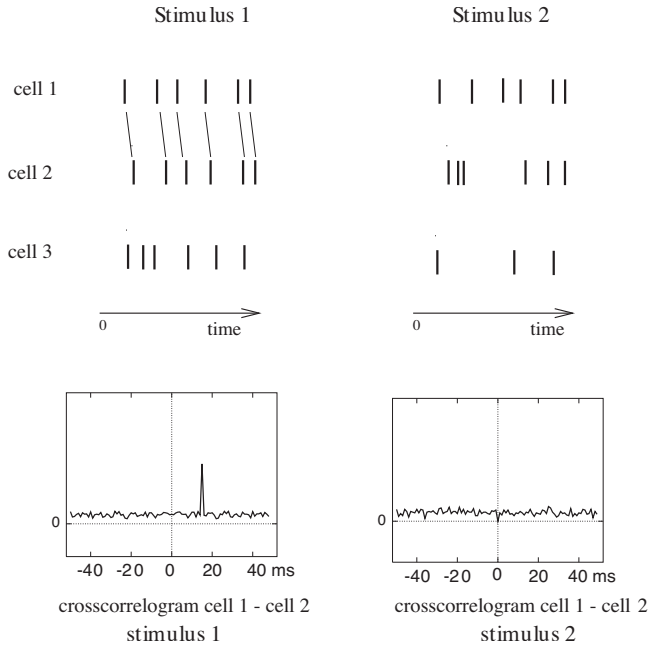

Fig. 3. Illustration of the information that could be carried by spike trains. The responses of three cells to two different stimuli are shown on one trial. Cell 3 reflects which stimulus was shown in the number of spikes produced, and this can be measured as spike count or rate information. Cells 1 and 2 have no spike count or rate information, because the number of spikes is not different for the two stimuli. Cells 1 and 2 do show some synchronization, reflected in the cross-correlogram, that is stimulus dependent, as the synchronization is present only when stimulus 1 is shown. The contribution of this effect is measured as the stimulus-dependent synchronization information.

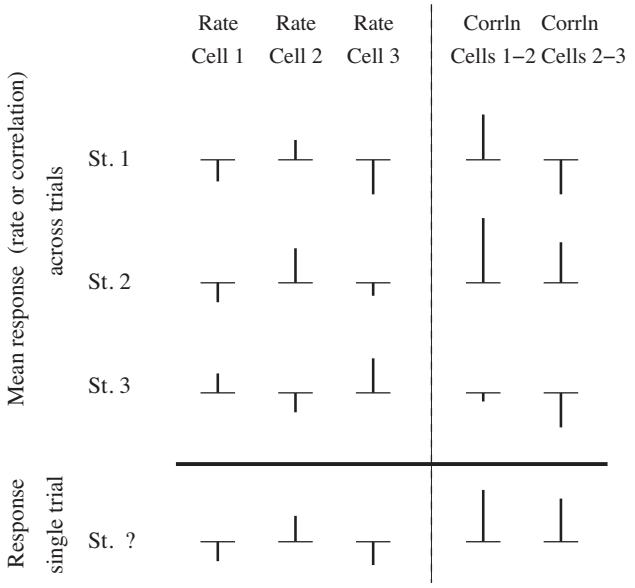

Fig. 4. Decoding neuronal firing rates and cross-correlations between neurons. The left part of the diagram shows the average firing rate (or equivalently spike count) responses of each of 3 cells (labelled as Rate Cell 1, 2, 3) to a set of 3 stimuli. The right two columns show a measure of the cross-correlation (averaged across trials) for some pairs of cells (labelled as Corrln Cells 1–2 and 2–3). The last row (labelled Response single trial) shows the data that might be obtained from a single trial and from which the stimulus that was shown (St. ? or s') must be estimated or decoded, using the average values across trials shown in the top part of the table. From the responses on the single trial, the most probable decoded stimulus is stimulus 2, based on the values of both the rates and the cross-correlations.
After Franco et al. (2004).

these cross-correlations are stimulus-dependent, then their positive contribution to the information encoded can be measured.

To test different hypotheses, the decoding can be based on all the columns of the Table (to provide the total information available from both the firing rates and the stimulus-dependent synchronization), on only the columns with the firing rates (to provide the information available from the firing rates), and only on the columns with the cross-correlation values (to provide the information available from the stimulus-dependent cross-correlations). Any information from stimulus-dependent cross-correlations will not necessarily be orthogonal to the rate information, and the procedures allow this to be checked by comparing the total information to that from the sum of the two components.

The measure of the synchronization introduced into the Table Data(s, r) on each trial is, for example, the value of the Pearson cross-correlation coefficient calculated for that trial at the appropriate lag for cell pairs that have significant cross-correlations (Franco et al., 2004). This value of this Pearson cross-correlation coefficient for a single trial can be calculated from pairs of spike trains on a single trial by forming for each cell a vector of 0s and 1s, the 1s representing the time of occurrence of spikes with a temporal resolution of 1 ms. Resulting values within the range −1 to 1 are shifted to obtain positive values. An advantage of basing the measure of synchronization on the Pearson cross-correlation coefficient is that it measures the amount of synchronization between a pair of neurons independently of the firing rate of the neurons. The lag at which the cross-correlation measure was computed for every single trial, and whether there was a significant cross-correlation between neuron pairs, can be identified from the location of the peak in the cross-correlogram taken across all trials. The cross-correlogram is calculated by, for every spike that occurred in one neuron, incrementing the bins of a histogram that correspond to the lag times of each of the spikes that occur for the other neuron. The raw cross-correlogram is corrected by subtracting the 'shift predictor' cross-correlogram (which is produced by random re-pairings of the trials), to produce the corrected cross-correlogram.

The decoding procedures used are similar to those described in Section 2.5 but applied to data of the type shown in Fig. 4, and further details of the decoding procedures are provided elsewhere (Rolls et al., 1997b; Franco et al., 2004). Examples of the use of these procedures are available (Franco et al., 2004; Aggelopoulos et al., 2005), and some of the results obtained are described in Section 3.

2.6.2. A second derivative approach

Another information theory-based approach to stimulus-dependent cross-correlation information has been developed as follows by Panzeri et al. (1999a) and Rolls et al. (2003b), extending the time-derivative approach of Section 2.3 (see also Ince et al. (2010b)).

This approach then addresses the limited sampling problem by taking short time epochs for the information analysis, in which low numbers of spikes, in practice typically 0, 1, or 2, spikes are likely to occur from each neuron.

Taking advantage of this, the response probabilities can be calculated in terms of pairwise correlations. These response probabilities are inserted into the Shannon information formula 16 to obtain expressions quantifying the impact of the pairwise correlations on the information $I(t)$ transmitted in a short time t by groups of spiking neurons:

$$I(t) = \sum_{s \in S} \sum_{\mathbf{r}} P(s, \mathbf{r}) \log_2 \frac{P(s, \mathbf{r})}{P(s)P(\mathbf{r})} \tag{16}$$

where $\mathbf{r}$ is the firing rate response vector comprised by the number of spikes emitted by each of the cells in the population in the short time t, and $P(s, \mathbf{r})$ refers to the joint probability distribution of stimuli with their respective neuronal response vectors.

The information depends upon the following two types of correlation:

The correlations in the neuronal response variability from the average to each stimulus (sometimes called "noise" correlations) γ:

$\gamma_{ij}(s)$ (for $i \neq j$) is the fraction of coincidences above (or below) that expected from uncorrelated responses, relative to the number of coincidences in the uncorrelated case (which is $\overline{n}_i(s)\overline{n}_j(s)$, the bar denoting the average across trials belonging to stimulus s, where $n_i(s)$ is the number of spikes emitted by cell i to stimulus s on a given trial)

$$\gamma_{ij}(s) = \frac{\overline{n_i(s)n_j(s)}}{(\overline{n}_i(s)\overline{n}_j(s))} - 1, \tag{17}$$

and is named the 'scaled cross-correlation density'. It can vary from −1 to ∞; negative $\gamma_{ij}(s)$'s indicate anticorrelation, whereas positive $\gamma_{ij}(s)$'s indicate correlation[3]. $\gamma_{ij}(s)$ can be thought of as the amount of trial by trial concurrent firing of the cells i and j, compared to that expected in the uncorrelated case. $\gamma_{ij}(s)$ (for $i \neq j$) is the 'scaled cross-correlation density' (Aertsen et al., 1989; Panzeri et al., 1999a), and is sometimes called the "noise" correlation (Gawne and Richmond, 1993; Shadlen and Newsome, 1995, 1998).

The correlations in the mean responses of the neurons across the set of stimuli (sometimes called "signal" correlations) ν:

$$\nu_{ij} = \frac{\langle \overline{n}_i(s)\overline{n}_j(s) \rangle_s}{\langle \overline{n}_i(s) \rangle_s \langle \overline{n}_j(s) \rangle_s} - 1 = \frac{\langle \overline{r}_i(s)\overline{r}_j(s) \rangle_s}{\langle \overline{r}_i(s) \rangle_s \langle \overline{r}_j(s) \rangle_s} - 1 \tag{19}$$

where $\overline{r}_i(s)$ is the mean rate of response of cell i (among C cells in total) to stimulus s over all the trials in which that stimulus was present. ν_{ij} can be thought of as the degree of similarity in the mean response profiles (averaged across trials) of the cells i and j to different stimuli. ν_{ij} is sometimes called the "signal" correlation (Gawne and Richmond, 1993; Shadlen and Newsome, 1995, 1998).

2.6.3. Information in the cross-correlations in short time periods

In the short timescale limit, the first (I_t) and second (I_{tt}) information derivatives describe the information $I(t)$ available in the short time t

$$I(t) = tI_t + \frac{t^2}{2} I_{tt}. \tag{20}$$

(The zeroth order, time-independent term is zero, as no information can be transmitted by the neurons in a time window

[3] $\gamma_{ij}(s)$ is an alternative, which produces a more compact information analysis, to the neuronal cross-correlation based on the Pearson correlation coefficient $\rho_{ij}(s)$ (Eq. (18)), which normalizes the number of coincidences above independence to the standard deviation of the number of coincidences expected if the cells were independent. The normalization used by the Pearson correlation coefficient has the advantage that it quantifies the strength of correlations between neurons in a rate-independent way. For the information analysis, it is more convenient to use the scaled correlation density $\gamma_{ij}(s)$ than the Pearson correlation coefficient, because of the compactness of the resulting formulation, and because of its scaling properties for small t. $\gamma_{ij}(s)$ remains finite as $t \to 0$, thus by using this measure we can keep the t expansion of the information explicit. Keeping the time-dependence of the resulting information components explicit greatly increases the amount of insight obtained from the series expansion. In contrast, the Pearson noise-correlation measure applied to short timescales approaches zero at short time windows:

$$\rho_{ij}(s) \equiv \frac{\overline{n_i(s)n_j(s)} - \overline{n}_i(s)\overline{n}_j(s)}{\sigma_{n_i(s)}\sigma_{n_j(s)}} \simeq t\gamma_{ij}(s)\sqrt{\overline{r}_i(s)\overline{r}_j(s)}, \tag{18}$$

where $\sigma_{n_i(s)}$ is the standard deviation of the count of spikes emitted by cell i in response to stimulus s.

of zero length. Higher order terms are also excluded as they become negligible.)

The instantaneous information rate I_t is[4]

$$I_t = \sum_{i=1}^{C} \left\langle \bar{r}_i(s) \log_2 \frac{\bar{r}_i(s)}{\langle \bar{r}_i(s') \rangle_{s'}} \right\rangle_s . \quad (21)$$

This formula, which is just the average across stimuli, summed across neurons of Eq. (7), shows that this information rate (the first time derivative) should not be linked to a high signal-to-noise ratio, but only reflects the extent to which the mean responses of each cell are distributed across stimuli. It does not reflect anything of the variability of those responses, that is of their noisiness, nor anything of the correlations among the mean responses of different cells.

The effect of (pairwise) correlations between the cells begins to be expressed in the second time derivative of the information. The expression for the instantaneous information 'acceleration' I_{tt} (the second time derivative of the information) breaks up into three terms:

$$\begin{aligned} I_{tt} = & \frac{1}{\ln 2} \sum_{i=1}^{C} \sum_{j=1}^{C} \langle \bar{r}_i(s) \rangle_s \langle \bar{r}_j(s) \rangle_s \left[\nu_{ij} + (1+\nu_{ij}) \ln \left(\frac{1}{1+\nu_{ij}} \right) \right] \\ & + \sum_{i=1}^{C} \sum_{j=1}^{C} \left[\left\langle \bar{r}_i(s) \bar{r}_j(s) \gamma_{ij}(s) \right\rangle_s \right] \log_2 \left(\frac{1}{1+\nu_{ij}} \right) \\ & + \sum_{i=1}^{C} \sum_{j=1}^{C} \left\langle \bar{r}_i(s) \bar{r}_j(s) (1+\gamma_{ij}(s)) \log_2 \left[\frac{(1+\gamma_{ij}(s)) \langle \bar{r}_i(s') \bar{r}_j(s') \rangle_{s'}}{\left\langle \bar{r}_i(s') \bar{r}_j(s') (1+\gamma_{ij}(s')) \right\rangle_{s'}} \right] \right\rangle_s . \end{aligned} \quad (22)$$

The first of these terms is all that survives if there is no noise correlation at all. Thus the *rate component* of the information is given by the sum of I_t (which is always greater than or equal to zero) and of the first term of I_{tt} (which is instead always less than or equal to zero).

The second term is non-zero if there is some correlation in the variance to a given stimulus, even if it is independent of which stimulus is present; this term thus represents the contribution of *stimulus-independent noise correlation* to the information.

The third component of I_{tt} represents the contribution of *stimulus-modulated noise correlation*, as it becomes non-zero only for stimulus-dependent noise correlations. These last two terms of I_{tt} together are referred to as the correlational components of the information.

The application of this approach to measuring the information in the relative time of firing of simultaneously recorded cells, together with further details of the method, are described by Panzeri et al. (1999b), Rolls et al. (2003b), and Rolls et al. (2004), and in Section 3.3.6.

2.6.4. Limitations of the derivative approach

The second derivative approach is elegant, but severely limited in its applicability to very short times, of order the mean interspike interval divided by the number of cells in the population. Over these short times, the dominant contribution is that of individual cells, summating linearly even if correlated, and pairwise correlations give a minor contribution. The reason for the limitation is that over longer times, when pairwise correlations begin to play a substantial role, also three-way and higher order correlations come to the fore, and assessing their contribution is intractable. In fact, one can consider for the sake of the argument a model of a large ensemble, in which correlations among the signal and noise components of neuronal firing are small in absolute value and entirely random in origin (Bezzi et al., 2002). Even such small random correlations are shown to lead to large possible synergy or redundancy, whenever the time window for extracting information from neuronal firing extends to the order of the mean interspike interval. Details of the argument are presented by Bezzi et al. (2002).

[4] Note that s' is used in Eqs. (21) and (22) just as a dummy variable to stand for s, as there are two summations performed over s.

2.7. Programs for information measurement from neuronal responses

Computer programs have been made available for the measurement of the information contained in neuronal responses (Ince et al., 2010a; Magri et al., 2009). We emphasize that care is needed in applying these to real neuronal data and interpreting the results, with many of the relevant issues described above.

3. Neuronal encoding: results obtained from information-theoretic analyses

How is information encoded in cortical areas such as the inferior temporal visual cortex? Can we read the code being used by the cortex? What are the advantages of the encoding scheme used for the neuronal network computations being performed in different areas of the cortex? These are some of the key issues considered in this Section (3). Because information is exchanged between the computing elements of the cortex (the neurons) by their spiking activity, which is conveyed by their axon to synapses onto other neurons, the appropriate level of analysis is how single neurons, and populations of single neurons, encode information in their firing. More global measures that reflect the averaged activity of large numbers of neurons (for example, PET (positron emission tomography) and fMRI (functional magnetic resonance imaging), EEG (electroencephalographic recording), and ERPs (event-related potentials)) cannot reveal how the information is represented, or how the computation is being performed (Rolls et al. (2009); Section 3.6).

In the treatment provided here, we focus on applications to the mammalian and especially the primate brain, using examples from a whole series of investigations on information representation in visual cortical areas, the hippocampus, and the taste and olfactory systems, the original papers on which refer to related publications. To provide an indication of the type of neuronal data that will be considered, Fig. 5 shows typical firing rate changes of a single neuron in the macaque inferior temporal visual cortex on different trials to each of several different faces (Tovee et al., 1993). This makes it clear that from the firing rate on any one trial, information is available about which stimulus was shown, and that the firing rate is graded, with a different firing rate response of the neuron to each stimulus.

3.1. The sparseness of the distributed encoding used by the brain

Some of the types of representation that might be found at the neuronal level are summarized next. 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. This is sometimes called a grandmother cell representation, because in a famous example, a single neuron might be active only if one's grandmother was being seen (see Barlow (1995)). 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 (for example, either active or not), the most distributed encoding is when half the neurons are active for any one stimulus or event. A **sparse distributed representation** is a distributed representation

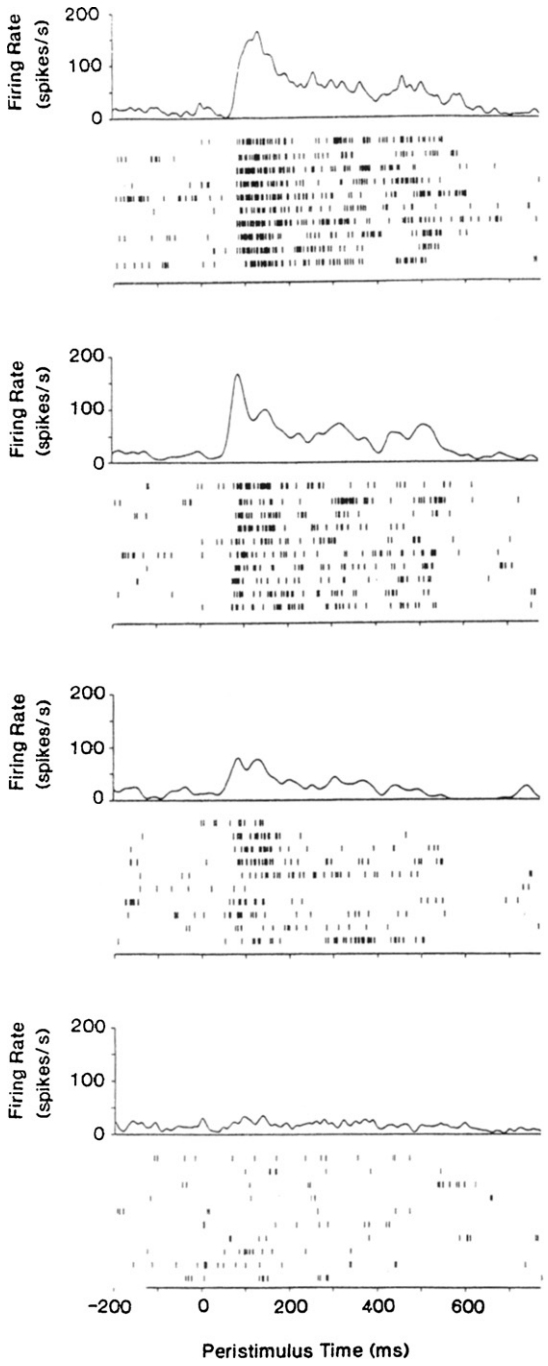

Fig. 5. Peristimulus time histograms and rastergrams showing the responses on different trials (originally in random order) of a face-selective neuron in the inferior temporal visual cortex to four different faces. (In the rastergrams each vertical line represents one spike from the neuron, and each row is a separate trial. Each block of the figure is for a different face.)
From Tovee et al. (1993).

in which a small proportion of the neurons is active at any one time. A local representation is sometimes termed a 'labelled line' representation, and a distributed representation is sometimes termed an 'across neuron' or 'across fiber' representation because the information can only be decoded by knowing the activity of an ensemble or population of neurons. A 'place' representation refers to the fact that the particular neurons that are active is important in encoding the information, and this in principle could apply to a local or distributed representation. In another type of encoding, the firing rate encodes the nature of the stimulus, as in the phase-locked encoding of frequency in the peripheral auditory system for stimuli below approximately 1 kHz. In most types of encoding, it is the relative firing rates of the particular ensemble of neurons that are firing that encodes which stimulus is present or its position in a topological space such as the retina or body surface as in distributed encoding, and the absolute firing rates of the active ensemble indicate the intensity of the stimulus.

3.1.1. Single neuron sparseness a^s

Eq. (23) defines a measure of the single neuron sparseness, a^s:

$$a^s = \frac{\left(\sum_{s=1}^{S} y_s/S\right)^2}{\left(\sum_{s=1}^{S} y_s^2\right)/S} \tag{23}$$

where y_s is the mean firing rate of the neuron to stimulus s in the set of S stimuli (Rolls and Treves, 1998). For a binary representation, a^s is 0.5 for a fully distributed representation, and $1/S$ if a neuron responds to one of the set of S stimuli. Another measure of sparseness is the kurtosis of the distribution, which is the fourth moment of the distribution. It reflects the length of the tail of the distribution. The distribution of the firing rates of a neuron in the inferior temporal visual cortex to a set of 65 stimuli is shown in Fig. 6. The sparseness a^s for this neuron was 0.69 (Rolls et al., 1997c).

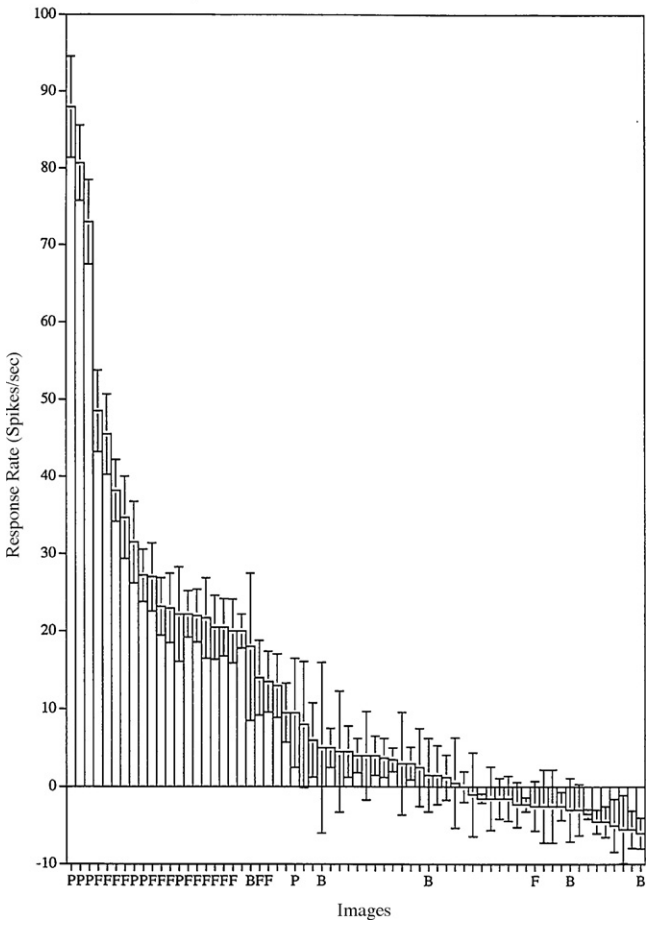

Fig. 6. Firing rate distribution of a single neuron in the temporal visual cortex to a set of 23 face (F) and 45 non-face images of natural scenes. The firing rate to each of the 68 stimuli is shown. The neuron does not respond to just one of the 68 stimuli. Instead, it responds to a small proportion of stimuli with high rates, to more stimuli with intermediate rates, and to many stimuli with almost no change of firing. This is typical of the distributed representations found in temporal cortical visual areas. The response, that is the firing rate minus the baseline spontaneous firing rate, is shown. After Rolls and Tovee (1995).

Table 2
Coding in associative memories.[a]

	Local	Sparse distributed	Fully distributed
Generalization, completion, graceful degradation	No	Yes	Yes
Number of patterns that can	N	of order $C/[a_o \log(1/a_o)]$	of order C
be stored	(large)	(can be larger)	(usually smaller than N)
Amount of information	Minimal	Intermediate	Large
in each pattern (values if binary)	($\log(N)$ bits)	($Na_o \log(1/a_o)$ bits)	(N bits)

[a] N refers here to the number of output units, and C to the average number of inputs to each output unit. a_o is the sparseness of output patterns, or roughly the proportion of output units activated by a UCS pattern. Note: logs are to the base 2.

It is important to understand and quantify the sparseness of representations in the brain, because many of the useful properties of neuronal networks such as generalization and completion only occur if the representations are distributed (Rolls, 2008), and because the value of the sparseness is an important factor in how many memories can be stored in such neural networks (Rolls and Treves, 1990; Treves and Rolls, 1991). Relatively sparse representations (low values of a^s) might be expected in memory systems as this will increase the number of different memories that can be stored and retrieved. Less sparse representations might be expected in sensory systems, as this could allow more information to be represented (see Table 2; and Rolls (2008)).

3.1.2. Grandmother cells vs. graded firing rates

Barlow (1972) proposed a single neuron doctrine for perceptual psychology. He proposed that sensory systems are organized to achieve as complete a representation as possible with the minimum number of active neurons. He suggested that at progressively higher levels of sensory processing, fewer and fewer cells are active, and that each represents a more and more specific happening in the sensory environment. He suggested that 1,000 active neurons (which he called cardinal cells) might represent the whole of a visual scene. An important principle involved in forming such a representation was the reduction of redundancy. The implication of Barlow's (1972) approach was that when an object is being recognized, there are, towards the end of the visual system, a small number of neurons (the cardinal cells) that are so specifically tuned that the activity of these neurons encodes the information that one particular object is being seen. (He thought that an active neuron conveys something of the order of complexity of a word.) The encoding of information in such a system is described as local, in that knowing the activity of just one neuron provides evidence that a particular stimulus (or, more exactly, a given 'trigger feature') is present. Barlow (1972) eschewed 'combinatorial rules of usage of nerve cells', and believed that the subtlety and sensitivity of perception results from the mechanisms determining when a single cell becomes active. In contrast, with distributed or ensemble encoding, the activity of several or many neurons must be known in order to identify which stimulus is present, that is, to read the code. It is the relative firing of the different neurons in the ensemble that provides the information about which object is present.

At the time Barlow (1972) wrote, there was little actual evidence on the activity of neurons in the higher parts of the visual and other sensory systems. There is now considerable evidence, which is now described.

First, it has been shown that the representation of which particular object (face) is present is actually 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 faces, with each neuron having a different profile of responses to each face (with an example for one neuron in Fig. 6) (Rolls and Tovee, 1995; Rolls, 2008). It would be difficult for most such single cells to tell which of a set of faces had been seen.

Second, the distributed nature of the representation can be further understood by the finding that the firing rate probability distribution of single neurons, when a wide range of natural visual stimuli are being viewed, is approximately exponential, with rather few stimuli producing high firing rates, and increasingly large numbers of stimuli producing lower and lower firing rates, as illustrated in Fig. 7a (Rolls and Tovee, 1995; Baddeley et al., 1997; Treves et al., 1999b; Franco et al., 2007).

For example, the responses of a set of temporal cortical neurons to 23 faces and 42 non-face natural images were measured, and a

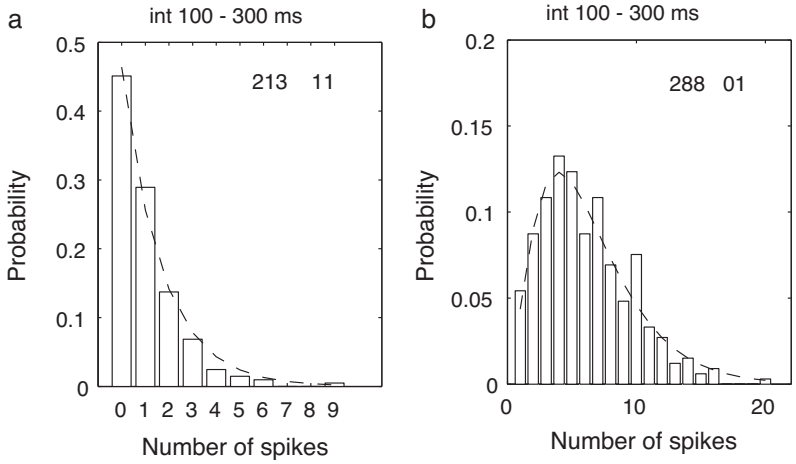

Fig. 7. Firing rate probability distributions for two neurons in the inferior temporal visual cortex tested with a set of 20 face and non-face stimuli. (a) A neuron with a good fit to an exponential probability distribution (dashed line). (b) A neuron that did not fit an exponential firing rate distribution (but which could be fitted by a gamma distribution, dashed line). The firing rates were measured in an interval 100–300 ms after the onset of the visual stimuli, and similar distributions are obtained in other intervals. After Franco et al. (2007).

distributed representation was found (Rolls and Tovee, 1995). The tuning was typically graded, with a range of different firing rates to the set of faces, and very little response to the non-face stimuli (see example in Fig. 6). The spontaneous firing rate of the neuron in Fig. 6 was 20 spikes/s, and the histogram bars indicate the change of firing rate from the spontaneous value produced by each stimulus. Stimuli that are faces are marked F, or P if they are in profile. B refers to images of scenes that included either a small face within the scene, sometimes as part of an image that included a whole person, or other body parts, such as hands (H) or legs. The non-face stimuli are unlabelled. 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 non-face stimuli. The sparseness value a^s for this cell across all 68 stimuli was 0.69, and the response sparseness a_r^s (based on the evoked responses minus the spontaneous firing of the neuron) was 0.19. It was found that the sparseness of the representation of the 68 stimuli by each neuron had an average across all neurons of 0.65 (Rolls and Tovee, 1995). This indicates a rather distributed representation. (If neurons had a continuum of firing rates equally distributed between zero and maximum rate, a^s would be 0.75, while if the probability of each response decreased linearly, to reach zero at the maximum rate, a^s would be 0.67). If the spontaneous firing rate was subtracted from the firing rate of the neuron to each stimulus, so that the changes of firing rate, that is the active responses of the neurons, were used in the sparseness calculation, then the 'response sparseness' a_r^s had a lower value, with a mean of 0.33 for the population of neurons, or 0.60 if calculated over the set of faces rather than over all the face and non-face stimuli. Thus the representation was rather distributed. (It is, of course, 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.) Thus we can reject a cardinal cell representation. As shown below, the readout of information from these cells is actually much better in any case than would be obtained from a local representation, and this makes it unlikely that there is a further population of neurons with very specific tuning that use local encoding.

These data provide a clear answer to whether these neurons are grandmother cells: they are not, in the sense that each neuron has a graded set of responses to the different members of a set of stimuli, with the prototypical distribution similar to that of the neuron illustrated in Fig. 6. On the other hand, each neuron does respond very much more to some stimuli than to many others, and in this sense is tuned to some stimuli.

Fig. 7 shows data of the type shown in Fig. 6 as firing rate probability density functions, that is as the probability that the neuron will be firing with particular rates. These data were from inferior temporal cortex neurons, and show when tested with a set of 20 face and non-face stimuli how fast the neuron will be firing in a period 100–300 ms after the visual stimulus appears (Franco et al., 2007). Fig. 7a shows an example of a neuron where the data fit an exponential firing rate probability distribution, with many occasions on which the neuron was firing with a very low firing rate, and decreasingly few occasions on which it fired at higher rates. This shows that the neuron can have high firing rates, but only to a few stimuli. Fig. 7b shows an example of a neuron where the data do not fit an exponential firing rate probability distribution, with insufficiently few very low rates. Of the 41 responsive neurons in this data set, 15 had a good fit to an exponential firing rate probability distribution; the other 26 neurons did not fit an exponential but did fit a gamma distribution in the way illustrated in Fig. 7b. For the neurons with an exponential distribution, the mean firing rate across the stimulus set was 5.7 spikes/s, and for the neurons with a gamma distribution was 21.1 spikes/s ($t = 4.5$, df = 25, $p < 0.001$). It may be that neurons with high mean rates to a stimulus set tend to have few low rates ever, and this accounts for their poor fit to an exponential firing rate probability distribution, which fits when there are many low firing rate values in the distribution as in Fig. 7a.

The large set of 68 stimuli used by Rolls and Tovee (1995) was chosen to produce an approximation to a set of stimuli that might be found to natural stimuli in a natural environment, and thus to provide evidence about the firing rate distribution of neurons to natural stimuli. Another approach to the same fundamental question was taken by Baddeley et al. (1997) who measured the firing rates over short periods of individual inferior temporal cortex neurons while monkeys watched continuous videos of natural scenes. They found that the firing rates of the neurons were again approximately exponentially distributed, providing further evidence that this type of representation is characteristic of inferior temporal cortex (and indeed also V1) neurons.

3.1.3. The typical shape of the firing rate distribution

The actual distribution of the firing rates to a wide set of natural stimuli is of interest, because it has a rather stereotypical shape, typically following a graded unimodal distribution with a long tail extending to high rates (see for example Fig. 7a). The mode of the distribution is close to the spontaneous firing rate, and sometimes it is at zero firing. If the number of spikes recorded in a fixed time window is taken to be constrained by a fixed maximum rate, one can try to interpret the distribution observed in terms of optimal information transmission (Shannon, 1948), by making the additional assumption that the coding is noiseless. An exponential distribution, which maximizes entropy (and hence information transmission for noiseless codes) is the most efficient in terms of energy consumption if its mean takes an optimal value that is a decreasing function of the relative metabolic cost of emitting a spike (Levy and Baxter, 1996). This argument would favour sparser coding schemes the more energy expensive neuronal firing is (relative to rest). Although the tail of actual firing rate distributions is often approximately exponential (see for example Fig. 7a; Baddeley et al. (1997); Rolls et al. (1997c); and Franco et al. (2007)), the maximum entropy argument cannot apply as such, because noise is present and the noise level varies as a function of the rate, which makes entropy maximization different from information maximization. Moreover, a mode at low but non-zero rate, which is often observed (see, e.g. Fig. 7b), is inconsistent with the energy efficiency hypothesis.

A simpler explanation for the characteristic firing rate distribution arises by appreciating that the value of the activation of a neuron across stimuli, reflecting a multitude of contributing factors, will typically have a Gaussian distribution; and by considering a physiological input–output transform (i.e. activation function), and realistic noise levels. In fact, an input–output transform that is supralinear in a range above threshold results from a fundamentally linear transform and fluctuations in the activation, and produces a variance in the output rate, across repeated trials, that increases with the rate itself, consistent with common observations. At the same time, such a supralinear transform tends to convert the Gaussian tail of the activation distribution into an approximately exponential tail, without implying a fully exponential distribution with the mode at zero. Such basic assumptions yield excellent fits with observed distributions (Treves et al., 1999b), which often differ from exponential in that there are too few very low rates observed, and too many low rates (Rolls et al., 1997c; Franco et al., 2007).

This peak at low but non-zero rates may be related to the low firing rate spontaneous activity that is typical of many cortical

neurons. Keeping the neurons close to threshold in this way may maximize the speed with which a network can respond to new inputs (because time is not required to bring the neurons from a strongly hyperpolarized state up to threshold). The advantage of having low spontaneous firing rates may be a further reason why a curve such as an exponential cannot sometimes be exactly fitted to the experimental data.

A conclusion of this analysis was that the firing rate distribution may arise from the threshold non-linearity of neurons combined with short-term variability in the responses of neurons (Treves et al., 1999b). It is worth noting, however, that for some neurons the firing rate distribution is approximately exponential, and that the sparseness of such an exponential distribution of firing rates is 0.5. It is interesting to realize that the representation that is stored in an associative network may be more sparse than the 0.5 value for an exponential firing rate distribution, because the non-linearity of learning introduced by the voltage dependence of the NMDA receptors effectively means that synaptic modification in, for example, an autoassociative network will occur only for the neurons with relatively high firing rates, i.e. for those that are strongly depolarized (Rolls, 2008).

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 non-face stimuli) do fit an exponential distribution, and others with higher spontaneous firing rates do not, as described above, 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. 8. An implication of the result shown in Fig. 8 is that a neuron with inputs from the inferior temporal visual cortex will receive an exponential distribution of firing rates on its afferents, and this is therefore the type of input that needs to be considered in theoretical models of neuronal network function in the brain (Rolls, 2008).

3.1.4. Population sparseness a^p

If instead we consider the responses of a population of neurons taken at any one time (to one stimulus), we might also expect a sparse graded distribution, with few neurons firing fast to a particular stimulus. It is important to measure the population

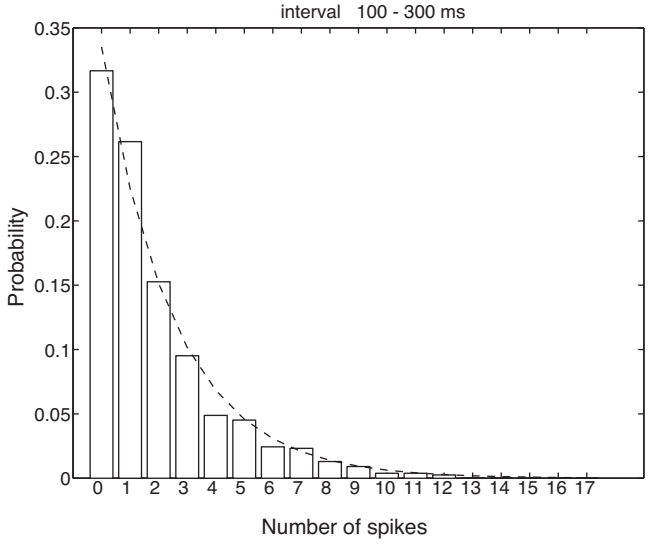

Fig. 8. An exponential firing rate probability distribution obtained by pooling the firing rates of a population of 41 inferior temporal cortex neurons tested to a set of 20 face and non-face stimuli. The firing rate probability distribution for the 100–300 ms interval following stimulus onset was formed 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 et al. (2007).

sparseness, for this is a key parameter that influences the number of different stimuli that can be stored and retrieved in networks such as those found in the cortex with recurrent collateral connections between the excitatory neurons, which can form autoassociation or attractor networks if the synapses are associatively modifiable (Hopfield, 1982; Treves and Rolls, 1991; Rolls and Treves, 1998; Rolls and Deco, 2002; Rolls, 2008). Further, in physics, if one can predict the distribution of the responses of the system at any one time (the population level) from the distribution of the responses of a component of the system across time, the system is described as ergodic, and a necessary condition for this is that the components are uncorrelated, that is, independent (Lehky et al., 2005). Considering this in neuronal terms, the average sparseness of a population of neurons over multiple stimulus inputs must equal the average selectivity to the stimuli of the single neurons within the population provided that the responses of the neurons are uncorrelated (Földiák, 2003).

The sparseness a^p of the population code may be quantified (for any one stimulus) as

$$a^p = \frac{\left(\sum_{n=1}^{N} y_n/N\right)^2}{\left(\sum_{n=1}^{N} y_n^2\right)/N} \tag{24}$$

where y_n is the mean firing rate of neuron n in the set of N neurons.

This measure, a^p, of the sparseness of the representation of a stimulus by a population of neurons has a number of advantages. One is that it is the same measure of sparseness that has proved to be useful and tractable in formal analyses of the capacity of associative neural networks and the interference between stimuli that use an approach derived from theoretical physics (Rolls and Treves, 1990, 1998; Treves, 1990; Treves and Rolls, 1991; Rolls, 2008). We note that high values of a^p indicate broad tuning of the population, and that low values of a^p indicate sparse population encoding.

Franco et al. (2007) measured the population sparseness of a set of 29 inferior temporal cortex neurons to a set of 20 stimuli that included faces and objects. Fig. 9a shows, for any one stimulus picked at random, the normalized firing rates of the population of neurons. The rates are ranked with the neuron with the highest rate on the left. For different stimuli, the shape of this distribution is on average the same, though with the neurons in a different order. (The rates of each neuron were normalized to a mean of 10 spikes/s before this graph was made, so that the neurons can be combined in the same graph, and so that the population sparseness has a defined value, as described by Franco et al. (2007).) The population sparseness a^p of this normalized (i.e. scaled) set of firing rates is 0.77.

Fig. 9b shows the probability distribution of the normalized firing rates of the population of (29) neurons to any stimulus from the set. This was calculated by taking the probability distribution of the data shown in Fig. 9a. This distribution is not exponential because of the normalization of the firing rates of each neuron, but becomes exponential as shown in Fig. 8 without the normalization step.

A very interesting finding of Franco et al. (2007) was that when the single cell sparseness a^s and the population sparseness a^p were measured from the same set of neurons in the same experiment, the values were very close, in this case 0.77. (This was found for a range of measurement intervals after stimulus onset, and also for a larger population of 41 neurons.)

The single cell sparseness a^s and the population sparseness a^p can take the same value if the response profiles of the neurons are uncorrelated, that is each neuron is independently tuned to the set of stimuli (Lehky et al., 2005). Franco et al. (2007) tested whether the response profiles of the neurons to the set of stimuli were

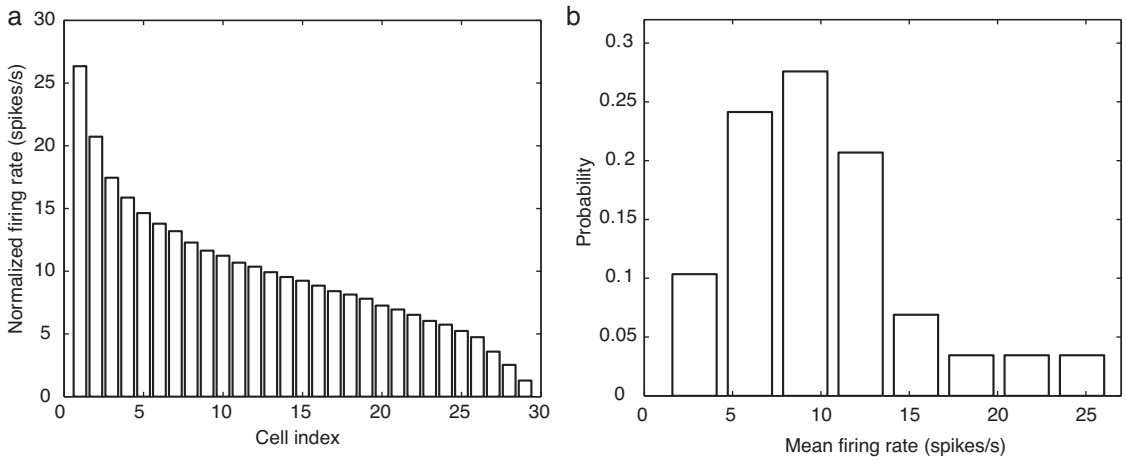

Fig. 9. Population sparseness. (a) The firing rates of a population of inferior temporal cortex neurons to any one stimulus from a set of 20 face and non-face stimuli. The rates of each neuron were normalized to the same average value of 10 spikes/s, then for each stimulus, the cell firing rates were placed in rank order, and then the mean firing rates of the first ranked cell, second ranked cell, etc. were taken. The graph thus shows, for any one stimulus picked at random, the expected normalized firing rates of the population of neurons. (b) The population normalized firing rate probability distributions for any one stimulus. This was computed effectively by taking the probability density function of the data shown in (a).
After Franco et al. (2007).

uncorrelated in two ways. In a first test, they found that the mean (Pearson) correlation between the response profiles computed over the 406 neuron pairs was low, 0.049 ± 0.013 (sem). In a second test, they computed how the multiple cell information available from these neurons about which stimulus was shown increased as the number of neurons in the sample was increased, and showed that the information increased approximately linearly with the number of neurons in the ensemble. The implication is that the neurons convey independent (non-redundant) information, and this would be expected to occur if the response profiles of the neurons to the stimuli are uncorrelated.

3.1.5. Ergodicity

We now consider the concept of ergodicity. The single neuron selectivity, a^s, reflects response distributions of individual neurons across time and therefore stimuli in the world (and has sometimes been termed "lifetime sparseness"). The population sparseness a^p reflects response distributions across all neurons in a population measured simultaneously (to for example one stimulus). The similarity of the average values of a^s and a^p (both 0.77 for inferior temporal cortex neurons (Franco et al., 2007)) indicates, we believe for the first time experimentally, that the representation (at least in the inferior temporal cortex) is ergodic. The representation is ergodic in the sense of statistical physics, where the average of a single component (in this context a single neuron) across time is compared with the average of an ensemble of components at one time (cf. Masuda and Aihara (2003) and Lehky et al. (2005)). This is described further next.

In comparing the neuronal selectivities a^s and population sparsenesses a^p, we formed a table in which the columns represent different neurons, and the stimuli different rows (Földiák, 2003). We are interested in the probability distribution functions (and not just their summary values a^s, and a^p), of the columns (which represent the individual neuron selectivities) and the rows (which represent the population tuning to any one stimulus). We could call the system strongly ergodic (cf. Lehky et al. (2005)) if the selectivity (probability density or distribution function) of each individual neuron is the same as the average population sparseness (probability density function). (Each neuron would be tuned to different stimuli, but have the same shape of the probability density function.) We have seen that this is not the case, in that the firing rate probability distribution functions of different neurons are different, with some fitting an exponential function, and some a gamma function (see Fig. 7). We can call the system weakly ergodic if individual neurons have different selectivities (i.e. different response probability density functions), but the average selectivity (measured in our case by $\langle a^s \rangle$) is the same as the average population sparseness (measured by $\langle a^p \rangle$), where $\langle\rangle$ indicates the ensemble average. We have seen that for inferior temporal cortex neurons the neuron selectivity probability density functions are different (see Fig. 7), but that their average $\langle a^s \rangle$ is the same as the average (across stimuli) $\langle a^p \rangle$ of the population sparseness, 0.77, and thus conclude that the representation in the inferior temporal visual cortex of objects and faces is weakly ergodic (Franco et al., 2007).

We note that weak ergodicity necessarily occurs if $\langle a^s \rangle$ and $\langle a^p \rangle$ are the same and the neurons are uncorrelated, that is each neuron is independently tuned to the set of stimuli (Lehky et al., 2005). The fact that both hold for the inferior temporal cortex neurons studied by Franco et al. (2007) thus indicates that their responses are uncorrelated, and this is potentially an important conclusion about the encoding of stimuli by these neurons. This conclusion is confirmed by the linear increase in the information with the number of neurons which is the case not only for this set of neurons (Franco et al., 2007), but also in other data sets for the inferior temporal visual cortex (Rolls et al., 1997b; Booth and Rolls, 1998). Both types of evidence thus indicate that the encoding provided by at least small subsets (up to, e.g. 20 neurons) of inferior temporal cortex neurons is approximately independent (non-redundant), which is an important principle of cortical encoding.

3.1.6. Comparisons of sparseness between areas: the hippocampus, insula, orbitofrontal cortex, and amygdala

In the study of Franco et al. (2007) on inferior temporal visual cortex neurons in macaques, the selectivity of individual cells for the set of stimuli, or single cell sparseness a^s, 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). Thus the single neuron probability density functions in these areas do not produce very sparse representations. Therefore the goal of the computations in the inferior temporal visual cortex may not be to produce sparse representations (as has been proposed for V1 (Field, 1994; Olshausen and Field, 1997, 2004; Vinje and Gallant, 2000)). Instead one of the goals of the

computations in the inferior temporal visual cortex may be to compute invariant representations of objects and faces (Rolls, 2000, 2007, 2008; Rolls and Deco, 2002; Rolls and Stringer, 2006), and to produce not very sparse distributed representations in order to maximize the information represented (see Table 2). In this context, it is very interesting that the representations of different stimuli provided by a population of inferior temporal cortex neurons are decorrelated, as shown by the finding that the mean (Pearson) correlation between the response profiles to a set of 20 stimuli computed over 406 neuron pairs was low, 0.049 ± 0.013 (sem) (Franco et al., 2007). The implication is that decorrelation is being achieved in the inferior temporal visual cortex, but not by forming a sparse code. It will be interesting to investigate the mechanisms for this.

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 described by Rolls et al. (1998) shows that for the representation of 64 locations around the walls of the room, the mean single cell sparseness $\langle a^s \rangle$ was 0.34 ± 0.13 (sd), and the mean population sparseness a^p was 0.33 ± 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. These sparseness values are for spatial view neurons, but it is possible that when neurons respond to combinations of spatial view and object (Rolls et al., 2005), or of spatial view and reward (Rolls and Xiang, 2005), the representations are more sparse. It is of interest that the mean firing rate of these spatial view neurons across all spatial views was 1.77 spikes/s (Rolls et al., 1998). (The mean spontaneous firing rate of the neurons was 0.1 spikes/s, and the average across neurons of the firing rate for the most effective spatial view was 13.2 spikes/s.) It is also notable that weak ergodicity is implied for this brain region too (given the similar values of $\langle a^s \rangle$ and $\langle a^p \rangle$), and the underlying basis for this is that the response profiles of the different hippocampal neurons to the spatial views are uncorrelated. Further support for these conclusions is that the information about spatial view increases linearly with the number of hippocampal spatial view neurons (Rolls et al., 1998), again providing evidence that the response profiles of the different neurons are uncorrelated.

Further evidence is now available on ergodicity in three further brain areas, the macaque insular primary taste cortex, the orbitofrontal cortex, and the amygdala. 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°C, 37°C and 42°C; flavour – blackcurrant juice; viscosity – carboxymethyl-cellulose 10 cPoise, 100 cPoise, 1000 cPoise and 10000 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 described by Verhagen et al. (2004) showed that in the primary taste cortex the mean value of a^s across 58 neurons was 0.745 and of a^p (normalized) was 0.708. Further analysis of data described by Rolls et al. (2003c), Verhagen et al. (2003), Kadohisa et al. (2004) and Kadohisa et al. (2005a) showed that in the orbitofrontal cortex the mean value of a^s across 30 neurons was 0.625 and of a^p was 0.611. Further analysis of data described by Kadohisa et al. (2005b) showed that in the amygdala the mean value of a^s across 38 neurons was 0.811 and of a^p was 0.813. Thus in all these cases, the mean value of a^s is close to that of a^p, and weak ergodicity is implied. The values of a^s and a^p are also 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.

3.1.7. Noise in the brain: the effects of sparseness and of graded representations

As we have seen, neuronal representations in the cortex have graded firing rates: the firing rate probability distribution of each neuron to a set of stimuli is often exponential or gamma. The graded nature of the representation is also evident in the range of firing rates in different neurons produced by a given stimulus (Fig. 9).

In processes in the brain such as decision-making, memory recall, and the maintenance of short-term memory, that are influenced by the noise produced by the close to random (Poisson) spike timings of each neuron for a given mean firing rate (Rolls and Deco, 2010; Rolls, 2008), the noise with this graded type of representation may be larger than with the uniform binary firing rate distribution that is usually investigated in theoretic analyses.

In integrate-and-fire simulations of an attractor decision-making network, Webb et al. (2011) showed that the noise is indeed greater for a given sparseness of the representation for graded, exponential, than for binary firing rate distributions. The greater noise was measured by faster escaping times from the spontaneous firing rate state when the decision cues are applied, and this corresponds to faster decision or reaction times. The greater noise was also evident as less stability of the spontaneous firing state before the decision cues are applied. The implication is that noise in the brain will continue to be a factor that influences processes such as decision-making, signal detection, short-term memory, and memory recall even with the quite large networks found in the cerebral cortex with several thousand recurrent collateral synapses onto each neuron. The greater noise with graded firing rate distributions has the advantage that it can increase the speed of operation of cortical circuitry. Noise in the brain has many other advantages (Rolls and Deco, 2010).

Conceptually, one can think that with graded firing rate distributions, a small number of neurons are made more important through their stronger weights and higher firing rates, noting that the variance of a Poisson process is equal to its mean. The influence of the few most highly firing neurons through their particularly strong synaptic weights on other neurons will have the effect of increasing the statistical fluctuations, which will be dominated by the relatively small number of highly firing neurons, and their possibly strong effects on a few other neurons with particularly strong synaptic weights from those highly firing neurons.

In the same way, making a representation more sparse (decreasing a) also increases the noise (stochastic fluctuations) due to finite size effects with spiking neurons (Webb et al., 2011). The diluted connectivity of the cerebral (including hippocampal) cortex has the effect of reducing the noise in integrate and fire attractor neuronal networks, but noise still remains in networks of biologically plausible size (Rolls and Webb, 2011).

We emphasize that it is important to understand the effects of noise in networks in the brain, and its implications for the stability of neuronal networks in the brain. For example, a stochastic neurodynamical approach to schizophrenia holds that there is less stability of cortical attractor networks involved in short-term memory and attention due to reduced functioning of the glutamate system, which decreases the firing rates of neurons in the prefrontal cortex, and therefore the depth of the basins of attraction, and thus the stability and signal-to-noise ratio given the spiking-related noise that is present. This it is suggested contributes to the cognitive changes in schizophrenia, which include impaired short-term memory and attention (Loh et al., 2007; Rolls et al., 2008b; Rolls and Deco, 2011). In another example, a stochastic neurodynamical approach to obsessive

compulsive disorder suggests that there is overstability in some networks in the prefrontal cortex and connected areas due to hyperglutamatergia (Rolls et al., 2008a; Rolls, 2011c). In both these cases, and also in normal brain function in relation to decision-making, memory recall, etc, it is important to know to what extent noise contributed by randomness in the spiking times of individual neurons for a given mean rate contributes to stochastic effects found in the brain which affect decision-making, stability, and which may if the stability is disturbed contribute to neuropsychiatric disorders (Rolls and Deco, 2010). In this context, the findings described in this paper are important for understanding normal and disordered brain function.

3.2. Sensory information from single neurons

An example of the responses of a single neuron (in this case in the inferior temporal visual cortex) to sets of objects and/or faces is shown in Fig. 6. We now consider how much information these types of neuronal response convey about the set of stimuli S, and about each stimulus s in the set. The mutual information $I(S, R)$ that the set of responses R encode about the set of stimuli S is calculated with Eq. (5) and corrected for the limited sampling using the analytic bias correction procedure described by Panzeri and Treves (1996) as described in detail by Rolls et al. (1997c). The information $I(s, R)$ about each single stimulus s in the set S, termed the stimulus-specific information (Rolls et al., 1997c) or stimulus-specific surprise (DeWeese and Meister, 1999), obtained from the set of the responses R of the single neuron is calculated with Eq. (6) and corrected for the limited sampling using the analytic bias correction procedure described by Panzeri and Treves (1996) as described in detail by Rolls et al. (1997c). (The average of $I(s, R)$ across stimuli is the mutual information $I(S, R)$.)

Fig. 10 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 65 stimuli is shown in Fig. 6.

Unless otherwise stated, the information measures given are for the information available on a single trial from the firing rate of the neuron in a 500 ms period starting 100 ms after the onset of the stimuli. It is shown in Fig. 10 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. 10 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. 10 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.

The information $I(s, R)$ about each stimulus in the set of 65 stimuli is shown in Fig. 11 for the same neuron, am242. The 23 face stimuli in the set are indicated by a diamond, and the 42 non-face stimuli by a cross. Using this much larger and more varied stimulus set, which is more representative of stimuli in the real world, a C-shaped function again describes the relation between the information conveyed by the cell about a stimulus and its firing rate to that stimulus.

In particular, this neuron reflected information about most, but not all, of the faces in the set, that is those faces that produced a higher firing rate than the overall mean firing rate to all the 65 stimuli, which was 31 spikes/s. In addition, it conveyed information about the majority of the 42 non-face stimuli by responding at a rate below the overall mean response of the neuron to the 65 stimuli. This analysis usefully makes the point that the information available in the neuronal responses about which stimulus was shown is relative to (dependent upon) the nature and range of stimuli in the test set of stimuli.

This evidence makes it clear that a single cortical visual neuron tuned to faces conveys information not just about one face, but about a whole set of faces, with the information conveyed on a single trial related to the difference in the firing rate response to a particular stimulus compared to the average response to all stimuli.

The analyses just described for neurons with visual responses are general, in that they apply in a very similar way to olfactory

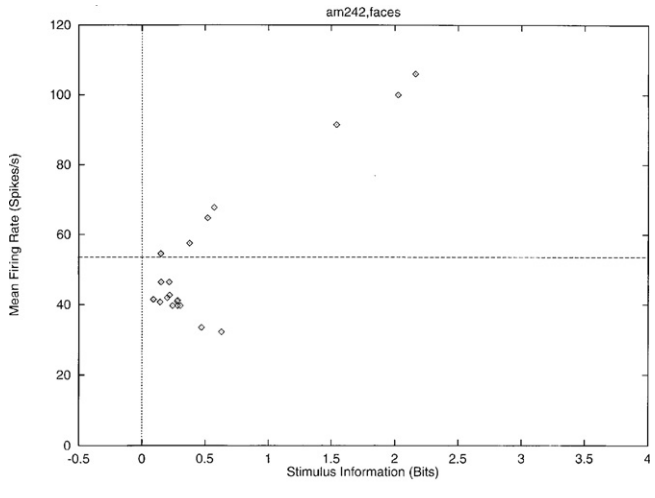

Fig. 10. The stimulus-specific information $I(s, R)$ available in the response of the same single neuron as in Fig. 6 about each of the stimuli in the set of 20 face stimuli (abscissa), with the firing rate of the neuron to the corresponding stimulus plotted as a function of this on the ordinate. The horizontal line shows the mean firing rate across all stimuli.
From Rolls et al. (1997c).

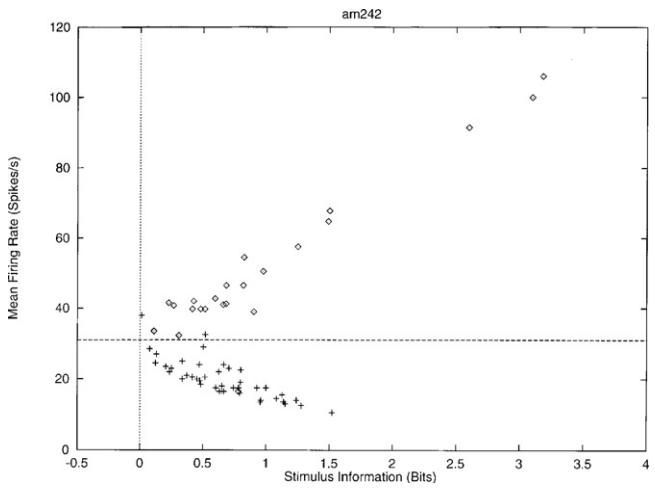

Fig. 11. The information $I(s, R)$ available in the response of the same neuron about each of the stimuli in the set of 23 face and 42 non-face stimuli (abscissa), with the firing rate of the neuron to the corresponding stimulus plotted as a function of this on the ordinate. The 23 face stimuli in the set are indicated by a diamond, and the 42 non-face stimuli by a cross. The horizontal line shows the mean firing rate across all stimuli.
After Rolls et al. (1997c).

neurons recorded in the macaque orbitofrontal cortex (Rolls et al., 1996).

The neurons in this sample reflected in their firing rates for the post-stimulus period 100–600 ms on average 0.36 bits of mutual information about which of 20 face stimuli was presented (Rolls et al., 1997c). Similar values have been found in other experiments (Tovee et al., 1993; Tovee and Rolls, 1995; Rolls et al., 1999, 2006). The information in short temporal epochs of the neuronal responses is described in Sections 2.3 and 3.2.5.

3.2.1. The information from single neurons: temporal codes vs. rate codes

With a firing rate code the number of spikes in a given time window is relevant. The temporal structure of the spikes within the time window might carry additional information (Rieke et al., 1997; Rolls, 2008; Panzeri et al., 2010). Although the timing of the spikes of neurons is known to code for time-varying features of a sensory stimulus (Panzeri et al., 2010), it is a more fundamental issue about whether time is exploited in the neuronal coding of (static) objects or the spatial structure of the environment, where non-time-varying stimulus features are present.

In the third of a series of papers that analyze the response of single neurons in the primate inferior temporal cortex to a set of static visual stimuli, Optican and Richmond (1987) applied information theory in a particularly direct and useful way. To ascertain the relevance of stimulus-locked temporal modulations in the firing of those neurons, they compared the amount of information about the stimuli that could be extracted from just the firing rate, computed over a relatively long interval of 384 ms, with the amount of information that could be extracted from a more complete description of the firing, which included temporal modulation. To derive this latter description (the temporal code within the spike train of a single neuron) they applied principal component analysis (PCA) to the temporal response vectors recorded for each neuron on each trial. The PCA helped to reduce the dimensionality of the neuronal response measurements. A temporal response vector was defined as a vector with as components the firing rates in each of 64 successive 6 ms time bins. The (64×64) covariance matrix was calculated across all trials of a particular neuron, and diagonalized. The first few eigenvectors of the matrix, those with the largest eigenvalues, are the principal components of the response, and the weights of each response vector on these four to five components can be used as a reduced description of the response, which still preserves, unlike the single value giving the mean firing rate during the entire interval, the main features of the temporal modulation within the interval. Thus a four- to five-dimensional temporal code could be contrasted with a one-dimensional rate code, and the comparison made quantitative by measuring the respective values for the mutual information with the stimuli.

Although the initial claim (Optican et al., 1991; Eskandar et al., 1992), that the temporal code carried nearly three times as much information as the rate code, was later found to be an artefact of limited sampling, and more recent analyses tend to minimize the additional information in the temporal description (Tovee et al., 1993; Heller et al., 1995), this type of application has immediately appeared straightforward and important, and it has led to many developments. By concentrating on the code expressed in the output rather than on the characterization of the neuronal channel itself, this approach is not affected much by the potential complexities of the preceding black box. Limited sampling, on the other hand, is a problem, particularly because it affects much more codes with a larger number of components, for example the four to five components of the PCA temporal description, than the one-dimensional firing rate code. This is made evident in the paper by Heller et al. (1995), in which the comparison is extended to several more detailed temporal descriptions, including a binary vector description in which the presence or not of a spike in each 1 ms bin of the response constitutes a component of a 320-dimensional vector. Obviously, this binary vector must contain at least all the information present in the reduced descriptions, whereas in the results of Heller et al. (1995), despite the use of a sophisticated neural network procedure to control limited sampling biases, the binary vector appears to be the code that carries the least information of all. In practice, with the data samples available in the experiments that have been done, and even when using analytic procedures to control limited sampling (Panzeri and Treves, 1996), reliable comparison can be made only with up to two- to three-dimensional codes.

Tovee et al. (1993) and Tovee and Rolls (1995) obtained further evidence that little information was encoded in the temporal aspects of firing within the spike train of a single neuron in the inferior temporal cortex by taking short epochs of the firing of neurons, lasting 20 ms or 50 ms, in which the opportunity for temporal encoding would be limited (because there were few spikes in these short time intervals). They found that a considerable proportion (30%) of the information available in a long time period of 400 ms utilizing temporal encoding within the spike train was available in time periods as short as 20 ms when only the number of spikes was taken into account.

Overall, the main result of these analyses applied to the responses to static stimuli in the temporal visual cortex of primates is that not much more information (perhaps only up to 10% more) can be extracted from temporal codes than from the firing rate measured over a judiciously chosen interval (Tovee et al., 1993; Heller et al., 1995). Indeed, it turns out that even this small amount of 'temporal information' is related primarily to the onset latency of the neuronal responses to different stimuli, rather than to anything more subtle (Tovee et al., 1993). In the primary visual cortex response latency and the number of spikes can similarly be partly independent, with latency more closely coding contrast or visibility (thus relating to the magnitude of the stimulus), and the number of spikes coding the stimulus orientation, or perhaps shape, that is the parameters of the stimuli to which the neuron is tuned (Richmond et al., 1997). A similar conclusion was reached about the information available about stimulus location in the rat somatosensory cortex (Panzeri et al., 2001a). In earlier visual areas than IT the additional 'temporally encoded' fraction of information may be larger, due especially to the increased relevance, earlier on, of precisely locked transient responses (Kjaer et al., 1994; Golomb et al., 1994; Heller et al., 1995; Victor, 2000; Panzeri et al., 2010). This is because if the responses to some stimuli are more transient and to others more sustained, or the onset latencies are different, this will result in more information if the temporal modulation of the response of the neuron is taken into account. This may however imply external knowledge of when a stimulus is presented, so that response latency can be used, and this is not likely to be available with natural visual stimuli in a natural setting.

In addition, the relevance for static (non-time-varying) visual stimuli of more substantial temporal codes involving for example particular spike patterns for particular stimuli remains to be demonstrated (Richmond, 2009). In particular, a strong null hypothesis that can be applied is that spike trains arise from stochastic sampling of an underlying deterministic temporally modulated rate function, that is, there is a time-varying rate function. In this view, order statistics seem to provide a sufficient theoretical construct to both generate simulated spike trains that are indistinguishable from those observed experimentally, and to evaluate (decode) the data recovered from experiments (Richmond, 2009). The implication of this view is that firing rates may

stochastically reflect a time-evolving process, but that there is nothing beyond this that has so far been demonstrated to be important about the temporal nature of the spikes elicited by a neuron to a set of stimuli, at least in well-studied systems in primates (Richmond, 2009).

For non-static visual stimuli and for other cortical systems, similar analyses have largely yet to be carried out, although clearly one expects to find more prominent temporal effects with at least the firing rate rapidly reflecting the changing stimuli in, e.g. the auditory system (Nelken et al., 1994; deCharms and Merzenich, 1996), for reasons similar to those just annunciated (Panzeri et al., 2010). However, because the firing rate code is fast in that much of the information available can be transmitted in time periods as short as 20 ms (Tovee and Rolls, 1995; Rolls et al., 1999, 2006), the firing rate code can follow stimuli that change rapidly in time, as in a movie, with different subsets of neurons active as the stimuli vary in time. In the auditory cortex, where time-related information is relevant to the time-varying sound, the pattern of spikes may carry extra information to that in the rates, though with biologically relevant timescales of postsynaptic potentials and membrane time constants, spike rate and phase rather than pattern are more informative (Kayser et al., 2009; Panzeri et al., 2010). Further discussion of coding in temporal response patterns of neurons to different stimuli is provided elsewhere (Victor, 2000; Panzeri et al., 2010).

3.2.2. Oscillations and phase coding

In this and the next section we consider how oscillations, either within a population of neurons, or between populations of neurons, influence the information that (as we have seen) is encoded and transmitted primarily in the number of spikes (Deco and Rolls, in preparation). The concept we develop is that because of the non-linear properties of neurons including the threshold for generating a spike, oscillations can influence the number of spikes that are produced within a population of neurons, or the speed with which they are transmitted between populations of neurons where the oscillations are in phase. We show how this can affect the speed of operation of neuronal networks, and information transmission between different networks.

The spike timing of hippocampal place neurons becomes earlier with respect to the theta oscillation cycle the further a rat has moved through the place field, and this phase precession can thus reflect the distance travelled through the place field (Huxter et al., 2003, 2008; Jensen and Lisman, 2000). Some information may in this way be encoded by the spike times of a neuron relative to the phase of an oscillation, a concept referred to as Phase-of-Firing Coding (O'Keefe and Burgess, 2005; Montemurro et al., 2008; Panzeri et al., 2010).

In the ventral visual cortical stream where short fixations are typically used to analyze stationery objects, the phase of firing in the gamma cycle appears to be redundant with respect to the firing rate (Vinck et al., 2010), but with rapidly timevarying stimuli in the auditory cortex and visual system, spike timing would be expected to be important, and indeed the phase of firing relative to slow (4–8 Hz) (Kayser et al., 2009; Panzeri et al., 2010) or faster (gamma, 60 Hz) (Koepsell et al., 2010) oscillations evident in local field potentials (LFPs) provides information additional to that in the firing rates.

In addition, short term memory encoding can be influenced by the phase of firing with respect to slow LFP oscillations (Lee et al., 2005; Siegel et al., 2009). The time of spike arrival relative to subthreshold membrane oscillations (SMO) in the postsynaptic neuron has been modeled as a possible way of encoding information (Nadasdy, 2010).

However, whether the neural firing relative to the phase angle is used to transmit that information for use by the brain remains to be shown, and if so, how much information this provides that is additional to the great deal of information in the neuronal population firing rate code (Section 3.3 and Fig. 18).

3.2.3. Oscillations and communication through coherence

Mechanisms through which oscillations that are coherent (i.e. of the same frequency and in a particular phase) can influence the speed of decision-making and of information transmission using "communication through coherence" (Fries, 2005, 2009) are illustrated in Fig. 12 and include the following (Deco and Rolls, in preparation).

One effect that oscillations can have is to speed information processing within a single network by increasing the mean spike count, that is the firing rate, of neurons, across both long time windows, and within shorter times within an oscillation period (Fig. 12a) (Smerieri et al., 2010). This was illustrated in an integrate-and-fire attractor neuronal network model of decision-making which normally operates without oscillations, and accounts for many aspects of decision-making in the brain at the neuronal and functional neuroimaging levels (Deco and Rolls, 2006; Wang, 2008, 2010; Rolls et al., 2010b,c). However, if the simulated network was made to oscillate at theta/delta frequencies (2–8 Hz) by introducing a second population of inhibitory interneurons with a longer synaptic time constant of 100 ms, then it was found that the decision times of the network were faster (Smerieri et al., 2010). One way in which the oscillations decreased the reaction times is illustrated in Fig. 12a. Because of the non-linearity of the neurons, which are typically held close to but on average a little below their firing threshold, the effect of the oscillations in the positive half cycle was to depolarize the neurons, which resulted in more spikes in total and bunched together within part of an oscillation cycle than would occur without oscillations. Given that with more spikes this class of attractor network tends to make decisions faster, as it is pushed more rapidly towards one of the high firing rate attractors due to the recurrent collateral positive feedback, the shorter reaction times are accounted for by the extra spike counts or firing rate (Rolls and Deco, 2010; Rolls et al., 2010c). Thus in this way oscillations can act through influencing firing rates to affect the properties of networks in the brain (Smerieri et al., 2010).

This type of effect applies not only to decision-making, but also to many operations supported by recurrent collateral excitatory connections in the neocortex, including memory recall, and the stability of short-term memory and attention (Rolls, 2008; Rolls and Deco, 2010). The effect of the oscillations in this case is like adding noise in the process known as stochastic resonance (Rolls and Deco, 2010). This leads to:

Principle 1: Oscillations interacting with the non-linearity of action potential generation can increase the number of spikes within short time intervals within an oscillation cycle and potentially in total, and this can increase the speed of processing in neural networks involved in decision-making, memory recall, etc, which are sensitive to the number of spikes received.

A second mechanism by which oscillations may affect information transmission is illustrated in Fig. 12b, and leads to:

Principle 2: Oscillations can increase the speed of processing by synchronizing spikes, leading to more rapid action potential generation.

This principle was illustrated in an integrate-and-fire biased competition model of attention in which it was shown that although modulation by the top-down bias of the firing rate was sufficient to implement attention, the reaction time was shorter if gamma oscillations (induced by altering the gAMPA /gNMDA conductance ratio for the synaptically activated ion channels) were present (Buehlmann and Deco, 2008). (The mechanism described

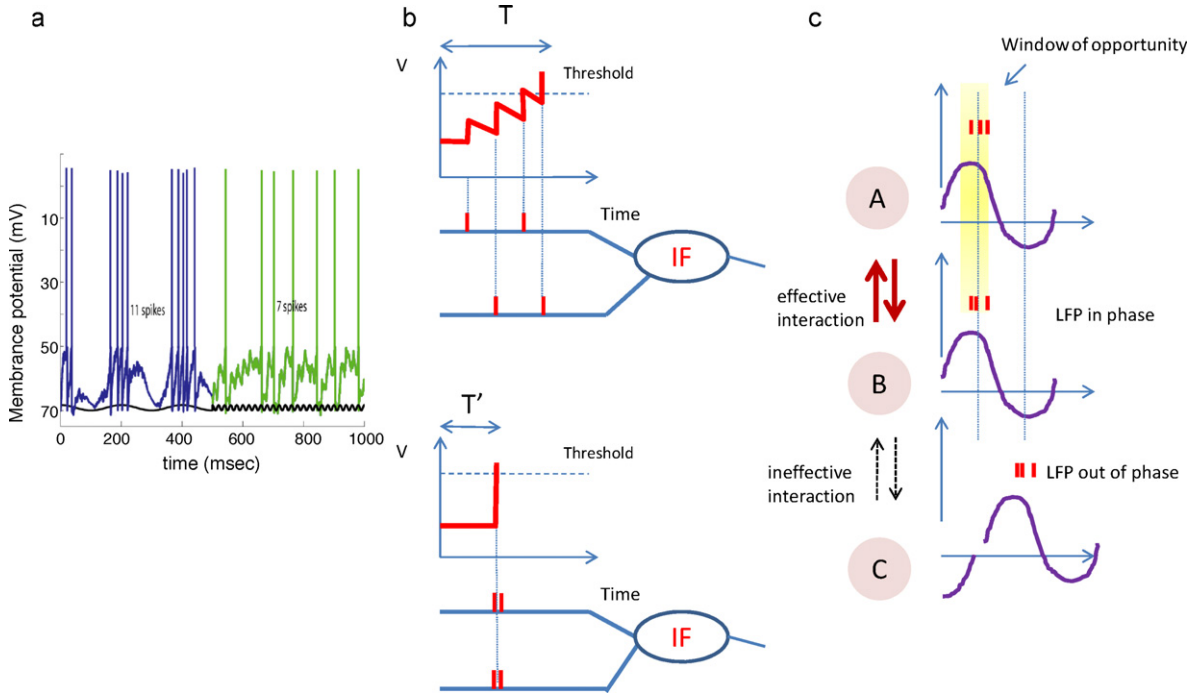

Fig. 12. Principles by which synchronicity and oscillations can influence neural processing. (a) Oscillations may lead to more spikes. The membrane potential of a single integrate-and-fire neuron in response to currents applied at slow (5 Hz) and at fast (50 Hz) frequencies and with identical amplitudes. The frequency was changed at time = 500 ms, and its time course is shown by the sinusoidal waveform (black line). The membrane time constant *gL* was 20 ms. The membrane potential shows a larger modulation with the low than with the high frequency input. This effect is produced by the filtering effect produced by the membrane time constant, which acts as a low pass filter, with smaller effects therefore produced by the higher frequency of 50 Hz. Depending on the average membrane potential produced by other inputs to the neuron, the larger modulation of the membrane potential produced by low frequencies may produce more action potentials with the low than with the high frequencies, as illustrated, or when the membrane potential oscillates than when there are no oscillations (after Smerieri et al., 2010). (b) Synchronous spikes from different sources (below) may speed neuronal responses compared to asynchronous spikes (above). *V* is the membrane potential of an integrate-and-fire neuron (IF). The time taken to reach the threshold for firing is T' vs. T (after Deco and Rolls, in preparation). (c) The communication-through-coherence (CTC) theory suggests that effective connections in a network can be shaped through phase relations (Fries, 2005, 2009). The neurons inside the pools A, B and C are rhythmically synchronized as indicated by the sinusoidal background LFP and the spikes (vertical red bars) around the peaks. Pools A and B are in phase and therefore the interchange of spikes is more effective, and more information is transmitted. On the other hand, pools B and C are in anti-phase and therefore fewer spikes are produced in the receiving population, and less information is transmitted (after Fries, 2005, 2009).

in Principle 1 may also have contributed to the effects found (Deco and Rolls, in preparation).)

A third mechanism is illustrated in Fig. 12c, and leads (Fries, 2005, 2009) to:

Principle 3: Cortical coherence is a mechanism that can influence the transmission of information between neuronal populations: the **communication-through-coherence** (CTC) hypothesis.

To investigate how the synchronization might influence communication, multi-unit activity (MUA) recorded simultaneously from 4 to 8 electrodes in cats' area 17, 18 and 21, and monkeys' area V1 and V4 was analyzed (Womelsdorf et al., 2007). For each pair of neuronal recordings, the synchronization was quantified by the MUA-MUA phase coherence spectrum, which showed a peak in the gamma frequency band close to 60 Hz. The strength of the interaction between two pairs, measured by the correlation between the two signals' power across trials, showed that this was on average across pairs highest when the phase lag between the pairs was zero. The authors suggested that the functional interactions between nodes in a network can be maximized if the phase relation is close to 0, and is lower at other phase relations, but did not elaborate on the mechanism (Womelsdorf et al., 2007).

One possibility we suggest is that if a synaptic input from a synchronized area arrives when a pyramidal cell is relatively depolarized because of an oscillation in-phase with the connected area, then the pyramidal cell might respond more, because of the threshold and then rapidly rising part of its activation function. (The activation function is the relation between the firing rate of the neuron and the depolarization caused by the synaptic inputs to the neuron, and this includes the non-linearity due to the threshold for firing of the neuron (Rolls, 2008).) If a synaptic input arrived out of phase, we suggest that its efficacy would be reduced because of shunting inhibition on the neurons receiving the input. Thus the efficacy of the spikes, which encode information as described above, in transmitting information may be influenced by synchronization, but the magnitude of the effect needs to be quantified, in ways that are suggested below.

The CTC hypothesis has been studied with a computational neuroscience approach with an integrate-and-fire neuronal network model with interacting populations of neurons (Buehlmann and Deco, 2010). The 'power correlation' had its maximum on average when the phase shift between the populations was 0. The transfer entropy (TE, an information theoretic measure described elsewhere (Buehlmann and Deco, 2010) which reflects the directionality of interactions by taking into account the firing rate in the other pool at a previous time step), also showed that on average there was strongest coupling between the populations at 0 phase shift. These effects could be obtained in the model in the beta as well as the gamma oscillation band.

The neurophysiological and modelling investigations just described suggest that neuronal populations influence one another most strongly if they oscillate at 0 phase shift, that is if they are synchronised as schematized in Fig. 12c (Fries, 2005, 2009; Womelsdorf et al., 2007; Buehlmann and Deco, 2010). We suggest that what needs to be tested next is whether information transmission between the two networks is facilitated if they are synchronized. This could be performed by using a set of stimuli *S* as

the input to one of the networks, and measuring whether the mutual information at the output of the second network about which stimulus was presented is greatest when the two networks are synchronized. It is likely that information transmission between neuronal populations is enhanced if both the frequency and phase of the two populations correspond, and this might make such transmission selective with respect to for example distracting stimuli (Akam and Kullmann, 2010). However, even this process may only work if the interfering stimuli are kept in separate frequency and phase bands (Akam and Kullmann, in press).

This raises the possibility, sometimes implied (Fries, 2005, 2009; Womelsdorf et al., 2007), that controlling which brain areas are synchronized might control the flow of information in the brain. That might be a mechanism of for example selective attention (Womelsdorf et al., 2007). However, what might be the external controller of synchronization? We are sceptical that there is one. Instead, we propose that oscillations arise as a result of the complex dynamics of cortical networks in which the ratios of synaptic conductances are a key to whether oscillations occur (Brunel and Wang, 2003; Smerieri et al., 2010; Buehlmann and Deco, 2010). Oscillations may then be more likely to occur synchronously when areas or neuronal populations are connected by strong synapses, and the neurons in both areas are simultaneously active, and especially when the firing is in phase the oscillations may reinforce each other because cortical areas typically have forward and backward connections (Rolls, 2008). Thus the synchronization may just result from the functional architecture of the cerebral cortex, and arise as a result of it being designed for other functions such as memory recall, short-term memory, attention, and decision-making, all of which can occur mechanistically in the brain without oscillations (Rolls, 2008; Rolls and Deco, 2010; Deco and Rolls, in preparation).

3.2.4. Oscillations can reset a network

A fourth mechanism by which oscillations may influence neuronal firing and neuronal encoding is by resetting neuronal activity. For example, a process such as memory recall may occur within a single theta cycle, and then be quenched so that a new attempt at recall can be made in the next theta cycle. This has the potential advantage that in a changing, ambiguous, or uncertain situation, several attempts can be made at the memory recall, without previous attempts dominating the memory state for a period due to attractor dynamics in autoassociation networks (Rolls and Treves, 1998, p. 118). Effects consistent with this prediction have recently been observed in the rat hippocampus (Jezek et al., 2011): in response to an instantaneous transition between two familiar and similar spatial contexts, hippocampal neurons in one theta cycle indicated one place, and in another theta cycle another place. These findings indicate that, in hippocampal CA3, pattern-completion dynamics can occur within each individual theta cycle. Reset, with potentially different recall in different theta cycles, may facilitate rapid updating and correction of recall. This leads to:

Principle 4: Autoassociative recall may occur in a single theta cycle, with reset and different recall in the next theta cycle.

3.2.5. The speed of information transfer by single neurons

Taking into account the points made in Section 2.3, Tovee et al. (1993) and Tovee and Rolls (1995) measured the information available in short epochs of the firing of single neurons, and found that a considerable proportion of the information available in a long time period of 400 ms was available in time periods as short as 20 ms and 50 ms. For example, in periods of 20 ms, 30% of the information present in 400 ms using temporal encoding with the first three principal components was available. Moreover, the exact time when the epoch was taken was not crucial, with the main effect being that rather more information was available if information was measured near the start of the spike train, when the firing rate of the neuron tended to be highest (see Fig. 13). The conclusion was that much information was available when temporal encoding could not be used easily, that is in very short time epochs of 20 or 50 ms.

It is also useful to note from Figs. 13 and 5 the typical time course of the responses of many temporal cortex visual neurons in the awake behaving primate. Although the firing rate and availability of information is highest in the first 50–100 ms of the neuronal response, the firing is overall well sustained in the 500 ms stimulus presentation period. Cortical neurons in the primate temporal lobe visual system, in the taste cortex (Rolls et al., 1990), and in the olfactory cortex (Rolls et al., 1996), do not in general have rapidly adapting neuronal responses to sensory stimuli. This may be important for associative learning: the outputs of these sensory systems can be maintained for sufficiently long while the stimuli are present for synaptic modification to occur. Although rapid synaptic adaptation within a spike train is seen in some experiments in brain slices (Markram and Tsodyks, 1996; Abbott et al., 1997), it is not a very marked effect in at least some brain systems in vivo, when they operate in normal physiological conditions with normal levels of acetylcholine, etc (Rolls, 2008).

3.2.6. Masking, information, and consciousness

To pursue this issue of the speed of processing and information availability even further, Rolls et al. (1994) and Rolls and Tovee (1994) limited the period for which visual cortical neurons could respond by using backward masking. In this paradigm, a short (16 ms) presentation of the test stimulus (a face) was followed after a delay of 0, 20, 40, 60 ms, etc. by a masking stimulus (which was a high contrast set of letters) (see Fig. 14). They showed that the mask did actually interrupt the neuronal response, and that at the shortest interval between the stimulus and the mask (a delay of 0 ms, or a 'Stimulus Onset Asynchrony' of 20 ms), the neurons in the temporal cortical areas fired for approximately 30 ms (see Fig. 15). Under these conditions, the subjects could identify which of five faces had been shown much better than chance. Interestingly, under these conditions, when the inferior temporal cortex neurons were firing for 30 ms, the subjects felt that they were guessing, and conscious perception was minimal (Rolls et al.,

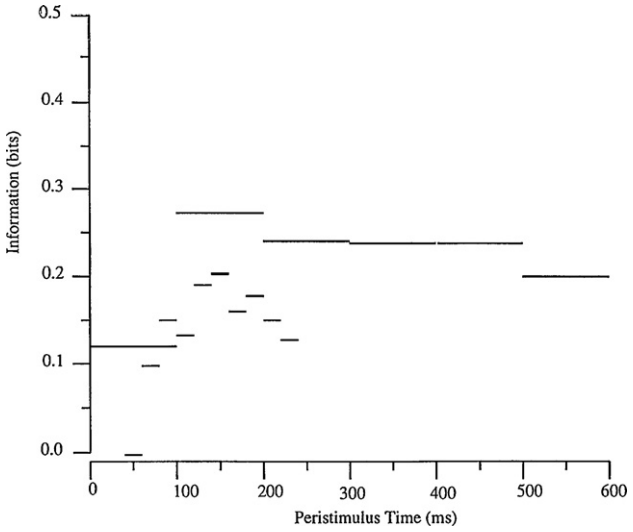

Fig. 13. The average information I(S,R) available in short temporal epochs (20 ms and 100 ms) of the spike trains of single inferior temporal cortex neurons about which face had been shown.
From Tovee and Rolls (1995).

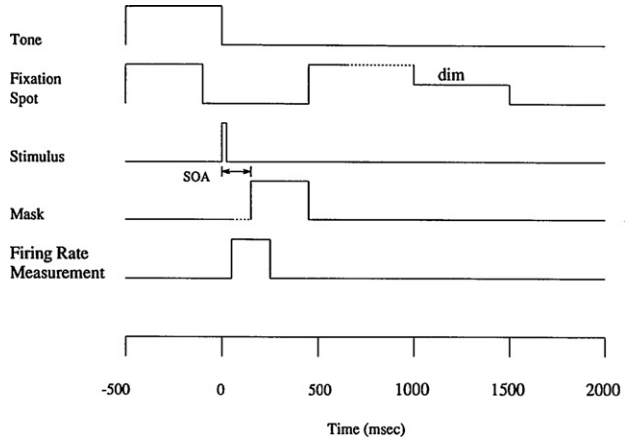

Fig. 14. Backward masking paradigm. The visual stimulus appeared at time 0 for 16 ms. The time between the start of the visual stimulus and the masking image is the Stimulus Onset Asynchrony (SOA). A visual fixation task was being performed to ensure correct fixation of the stimulus. In the fixation task, the fixation spot appeared in the middle of the screen at time −500 ms, was switched off 100 ms before the test stimulus was shown, and was switched on again at the end of the mask stimulus. Then when the fixation spot dimmed after a random time, fruit juice could be obtained by licking. No eye movements could be performed after the onset of the fixation spot.
After Rolls and Tovee (1994).

1994), the neurons conveyed on average 0.10 bits of information (Rolls et al., 1999). With a stimulus onset asynchrony of 40 ms, when the inferior temporal cortex neurons were firing for 50 ms, not only did the subjects' performance improve, but the stimuli were now perceived clearly, consciously, and the neurons conveyed on average 0.16 bits of information. This has contributed to the view that consciousness has a higher threshold of activity *in a given pathway*, in this case a pathway for face analysis, than does unconscious processing and performance using the same pathway (Rolls, 2003, 2006, 2011a).

3.2.7. First spike codes

The issue of how rapidly information can be read from neurons is crucial and fundamental to understanding how rapidly memory systems in the brain could operate in terms of reading the code from the input neurons to initiate retrieval, whether in a pattern associator or autoassociation network (Rolls, 2008; Rolls and Deco, 2010). This is also a crucial issue for understanding how any stage of cortical processing operates, given that each stage includes associative or competitive network processes that require the code to be read before it can pass useful output to the next stage of processing (see Rolls (2008); Rolls and Deco (2002); and Panzeri et al. (2001b)). For this reason, we have performed further analyses of the speed of availability of information from neuronal firing, and the neuronal code. A rapid readout of information from any one stage of for example visual processing is important, for the ventral visual system is organized as a hierarchy of cortical areas, and the neuronal response latencies are approximately 100 ms in the inferior temporal visual cortex, and 40–50 ms in the primary visual cortex, allowing only approximately 50–60 ms of processing time for V1–V2–V4–inferior temporal cortex (Baylis et al., 1987; Nowak and Bullier, 1997; Rolls and Deco, 2002). There is much evidence that the time required for each stage of processing is relatively short. For example, in addition to the evidence already presented, visual stimuli presented in succession approximately 15 ms apart can be separately identified (Keysers and Perrett, 2002); and the reaction time for identifying visual stimuli is relatively short and requires a relatively short cortical processing time (Rolls, 2003; Bacon-Mace et al., 2005).

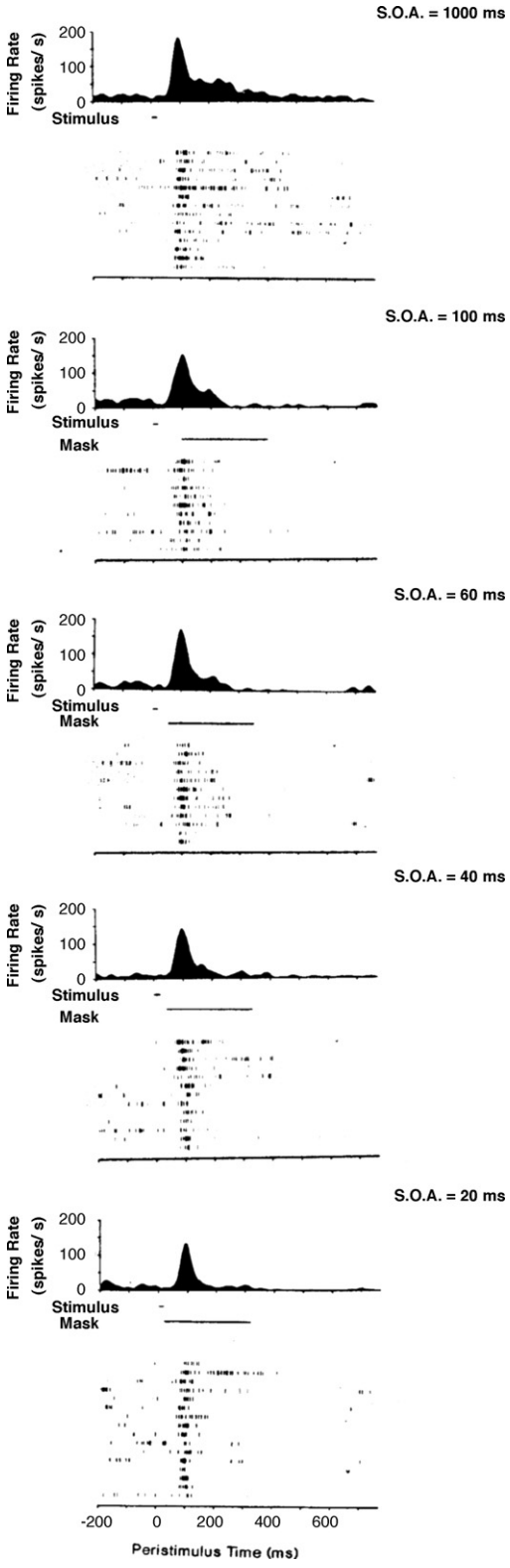

Fig. 15. Firing of a temporal cortex cell to a 20 ms presentation of a face stimulus when the face was followed with different stimulus onset asynchronies (SOAs) by a masking visual stimulus. At an SOA of 20 ms, when the mask immediately followed the face, the neuron fired for only approximately 30 ms, yet identification above change (by 'guessing') of the face at this SOA by human observers was possible.
After Rolls and Tovee (1994) and Rolls et al. (1994).

In this context, Delorme and Thorpe (2001) have suggested that just one spike from each neuron is sufficient, and indeed it has been suggested that the order of the first spike in different neurons may be part of the code (Delorme and Thorpe, 2001; Thorpe et al., 2001; VanRullen et al., 2005). (Implicit in the spike order hypothesis is that the first spike is particularly important, for it would be difficult to measure the order for anything other than the first spike.) An alternative view is that the number of spikes in a fixed time window over which a postsynaptic neuron could integrate information is more realistic, and this time might be in the order of 20 ms for a single receiving neuron, or much longer if the receiving neurons are connected by recurrent collateral associative synapses and so can integrate information over time (Deco and Rolls, 2006; Rolls and Deco, 2002; Panzeri et al., 2001b). Although the number of spikes in a short time window of, e.g. 20 ms is likely to be 0, 1, or 2, the information available may be more than that from the first spike alone, and Rolls et al. (2006) examined this by measuring neuronal activity in the inferior temporal visual cortex, and then applying quantitative information theoretic methods to measure the information transmitted by single spikes, and within short time windows.

The cumulative single cell information about which of the twenty stimuli was shown from all spikes and from the first spike starting at 100 ms after stimulus onset is shown in Fig. 16. A period of 100 ms is just longer than the shortest response latency of the neurons from which recordings were made, so starting the measure at this time provides the best chance for the single spike measurement to catch a spike that is related to the stimulus. The means and standard errors across the 21 different neurons are shown. The cumulated information from the total number of spikes is larger than that from the first spike, and this is evident and significant within 50 ms of the start of the time epoch. In calculating the information from the first spike, just the first spike in the analysis window starting in this case at 100 ms after stimulus onset was used.

Because any one neuron receiving information from the population being analyzed has multiple inputs, we show in Fig. 17 the cumulative information that would be available from multiple cells (21) about which of the 20 stimuli was shown, taking

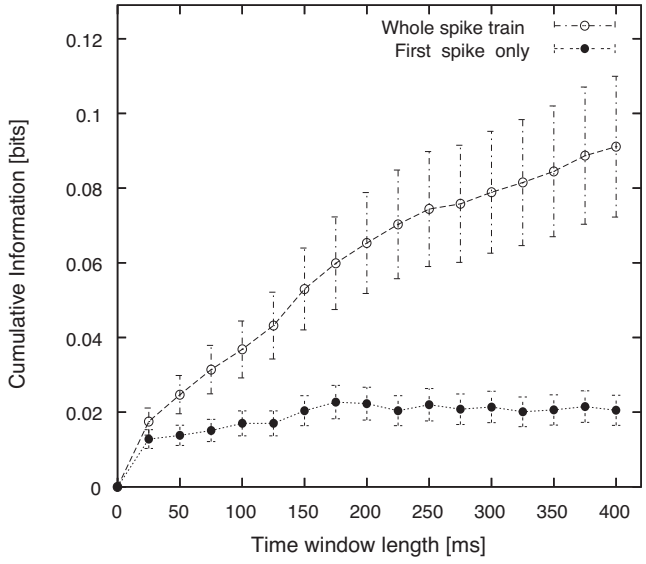

Fig. 16. Speed of information availability in the inferior temporal visual cortex. Cumulative single cell information from all spikes and from the first spike with the analysis starting at 100 ms after stimulus onset. The mean and sem over 21 neurons are shown.
After Rolls et al. (2006).

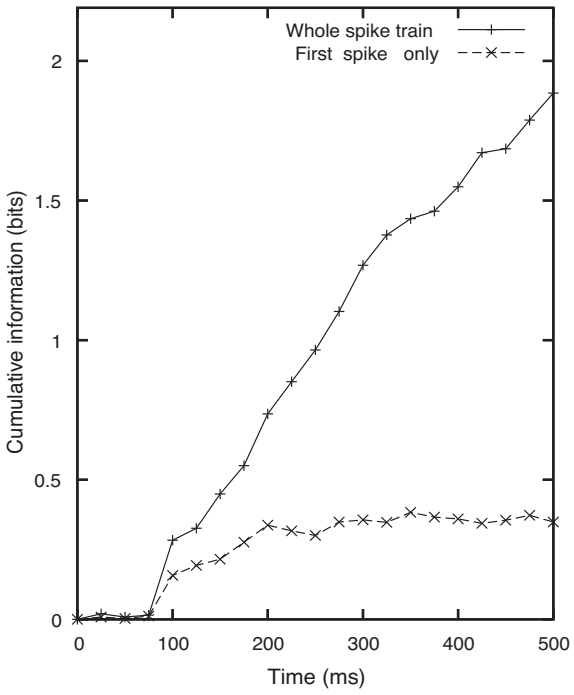

Fig. 17. Speed of information availability in the inferior temporal visual cortex. Cumulative multiple cell information from all spikes and first spike starting at the time of stimulus onset (0 ms) for the population of 21 neurons about the set of 20 stimuli.
After Rolls et al. (2006).

both the first spike after the time of stimulus onset (0 ms), and the total number of spikes after 0 ms from each neuron. The cumulative information even from multiple cells is much greater when all the spikes rather than just the first spike are used.

These and many further results thus show that although considerable information is present in the first spike, more information is available under the more biologically realistic assumption that neurons integrate spikes over a short time window (depending on their time constants) of for example 20 ms (Rolls et al., 2006). Moreover, the order of spike arrival times from different neurons did not convey significant extra information to that available from the firing rates in short periods (Rolls et al., 2006).

The conclusions from the single cell information analyses are thus that most of the information is encoded in the spike count; that large parts of this information are available in short temporal epochs of, e.g. 20 ms or 50 ms; and that any additional information which appears to be temporally encoded is related to the latency of the neuronal response, and reflects sudden changes in the visual stimuli. Therefore a neuron in the next cortical area would obtain considerable information within 20–50 ms by measuring the firing rate of a single neuron. Moreover, if it took a short sample of the firing rate of many neurons in the preceding area, then very much information is made available in a short time, as shown above and in Section 3.3.

3.3. Sensory information from multiple cells: independent information vs. redundancy

3.3.1. Overview of population encoding

The information conveyed by the firing rates of for example inferior temporal cortex (IT) neurons increases almost linearly with the number of different single neurons (up to reasonable numbers of neurons in the range 14–50), so that neurons encode information that is almost independent (Gawne and Richmond,

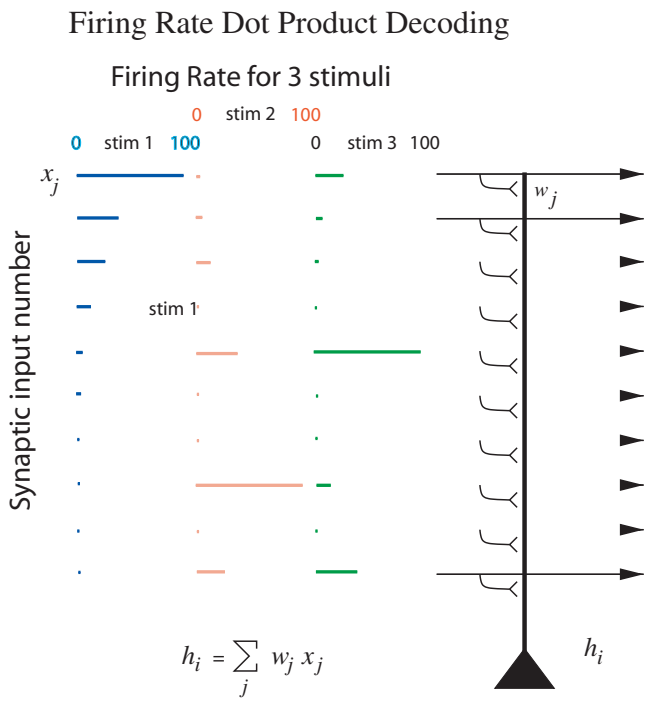

Fig. 18. Population encoding by firing rates. Each stimulus is encoded by an approximately exponential firing rate distribution of a population of neurons. The distribution is ordered to show this for stimulus 1, and other stimuli are represented by similar distributions with each neuron tuned independently of the others. The code can be read by a dot (inner) product decoding performed by any receiving neuron of the firing rates x_j with the synaptic weights w_j, with h_i being the depolarization of the neuron produced by the synaptic currents. The almost linear increase in information with the number of neurons is related to the finding that the tuning profiles to a set of stimuli are almost independent.

1993; Rolls et al., 1997b, 2004; Booth and Rolls, 1998; Aggelopoulos et al., 2005) (Fig. 18).

Similar results have been found for the representation of olfactory information in the orbitofrontal cortex (Rolls et al., 2010a), for hippocampal cortex neurons about spatial location (Rolls et al., 1998), and for head direction cells in the presubicular cortex (Robertson et al., 1999).

Because information is a log measure, the number of stimuli encoded by a population of neurons increases approximately exponentially with the number of neurons (Rolls et al., 1997b; Abbott et al., 1996).

Moreover, because the representation is distributed with a different subset of neurons responding to different stimuli (Fig. 18), and can be decoded quite efficiently by neuronally plausible dot-product decoding (Rolls et al., 1997b), the code is very robust and supports fundamental cortical computations including the recall of memories from associative networks (Rolls, 2008), and fast transmission across multilayer systems (Panzeri et al., 2001b).

The firing rate code is also fast: much of the information available in a long period of 400 ms can be transmitted in time periods as short as 20 ms (Rolls et al., 2006, 1999, 2004; Tovee and Rolls, 1995). In these short time windows, each neuron will transmit typically 0–3 spikes with a firing rate probability distribution that is frequently approximately exponential, that is, few neurons have high firing rates (Treves et al., 1999b; Franco et al., 2007; Rolls, 2008).

In this scenario, it is the numbers of spikes carried by each of the population of neurons that convey which stimulus is present, and those numbers are effectively randomly different for each stimulus (Fig. 18). More information is transmitted in a longer time window of 50 ms, showing that the number of spikes transmitted is important in transmitting information (Tovee and Rolls, 1995; Rolls, 2008).

This evidence that one or a few spikes are important in the firing rate code used by the cortex is relevant to understanding the mechanisms by which neuronal synchronization can influence information transmission in short time windows, by increasing the number of spikes in a short time window (Deco and Rolls, in preparation). These high information values are found with stimulus-locked data collection so that no account need be taken when neurons read the information of synchrony or oscillations, which in any case are relatively small effects as shown by the lack of oscillation evident in the autocorrelation functions of individual neurons in primates, and by the fact that to the extent that they are present they have little effect on these information measures as shown by shuffling the data from different simultaneously recorded neurons across trials (Aggelopoulos et al., 2005; Rolls et al., 2004; Rolls, 2008; Deco and Rolls, in preparation).

3.3.2. Population encoding with independent contributions from each neuron

The rate at which a single cell provides information translates into an instantaneous information flow across a population (with a simple multiplication by the number of cells) only to the extent that different cells provide different (independent) information. To verify whether this condition holds, one cannot extend to multiple cells the simplified formula for the first time-derivative, because it is made simple precisely by the assumption of independence between spikes, and one cannot even measure directly the full information provided by multiple (more than two to three) cells, because of the limited sampling problem discussed above. Therefore one has to analyze the degree of independence (or conversely of redundancy) either directly among pairs – at most triplets – of cells, or indirectly by using decoding procedures to transform population responses. Obviously, the results of the analysis will vary a great deal with the particular neural system considered and the particular set of stimuli, or in general of neuronal correlates, used. For many systems, before undertaking to quantify the analysis in terms of information measures, it takes only a simple qualitative description of the responses to realize that there is a lot of redundancy and very little diversity in the responses. For example, if one selects pain-responsive cells in the somatosensory system and uses painful electrical stimulation of different intensities, most of the recorded cells are likely to convey pretty much the same information, signalling the intensity of the stimulation with the intensity of their single-cell response. Therefore, an analysis of redundancy makes sense only for a neuronal system that functions to represent, and enable discriminations between, a large variety of stimuli, and only when using a set of stimuli representative, in some sense, of that large variety.

Rolls et al. (1997b) measured the information available from a population of inferior temporal cortex neurons using the decoding methods described in Section 2.5, and found that the information increased approximately linearly, as shown in Fig. 19. (It is shown below that the increase is limited only by the information ceiling of 4.32 bits necessary to encode the 20 stimuli. If it were not for this approach to the ceiling, the increase would be approximately linear (Rolls et al., 1997b).) To the extent that the information increases linearly with the number of neurons, the neurons convey independent information, and there is no redundancy, at least with numbers of neurons in this range.

Remembering that the information in bits is a logarithmic measure, this shows that the representational capacity of this population of cells increases exponentially (see Fig. 20). This is the case both when an optimal, probability estimation, form of decoding of the activity of the neuronal population is used, and also when the neurally plausible (though less efficient) dot product type of decoding is used (Fig. 19). By simulation of further neurons and

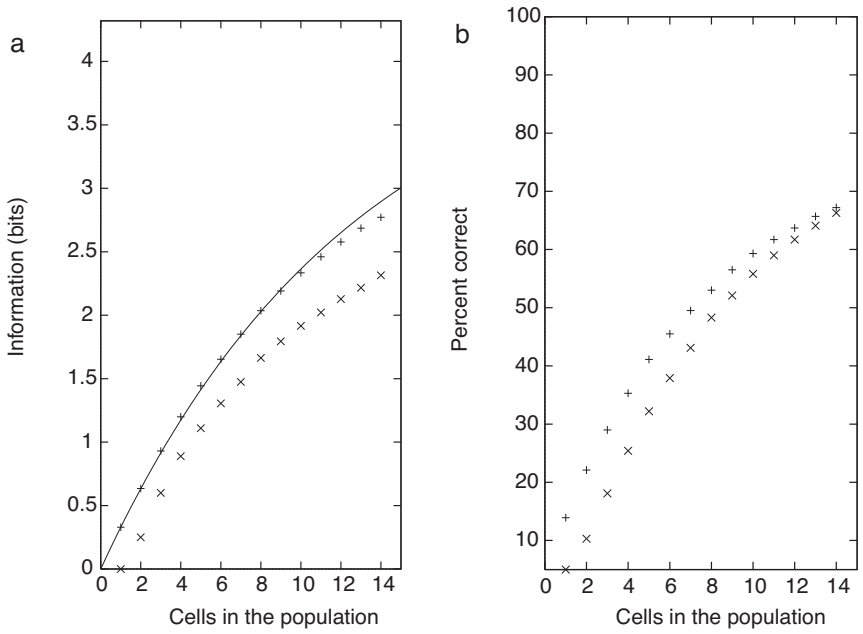

Fig. 19. (a) The values for the average information available in the responses of different numbers of these neurons on each trial, about which of a set of 20 face stimuli has been shown. The decoding method was Dot Product (DP, ×) or Probability Estimation (PE, +). The full line indicates the amount 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). The measurement period was 500 ms.
After Rolls et al. (1997b).

further stimuli, we have shown that the capacity grows very impressively, approximately as shown in Fig. 20 (Abbott et al., 1996).

Although these and some of the other results described here are for face-selective neurons in the inferior temporal visual cortex, similar results were obtained for neurons responding to objects in the inferior temporal visual cortex (Booth and Rolls, 1998), and for neurons responding to spatial view in the hippocampus (Rolls et al., 1998) (Fig. 21).

Although those neurons were not simultaneously recorded, a similar approximately linear increase in the information from *simultaneously* recorded cells as the number of neurons in the sample increased also occurs (Rolls et al., 2003b, 2004, 2006; Franco et al., 2004; Aggelopoulos et al., 2005). These findings imply little redundancy, and that the number of stimuli that can be encoded increases approximately exponentially with the number of neurons in the population, as illustrated in Figs. 20 and 19.

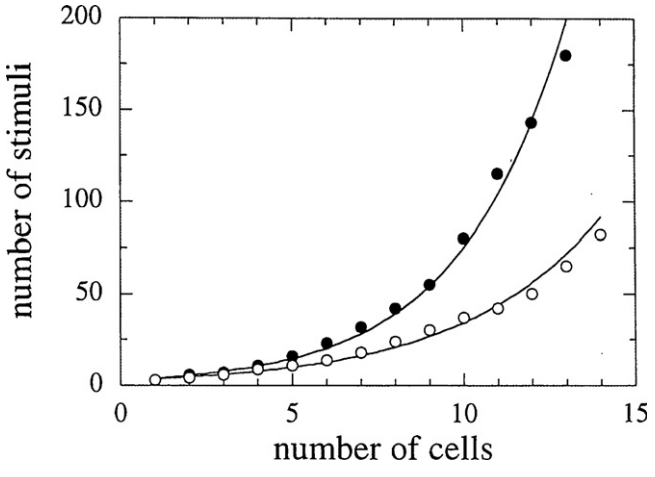

Fig. 20. The number of stimuli (in this case from a set of 20 faces) that are encoded in the responses of different numbers of neurons in the temporal lobe visual cortex, based on the results shown in Fig. 19. Filled circles: Bayesian Probability Estimation decoding. Open circles: Dot Product decoding.
After Rolls et al., 1997b; Abbott et al., 1996.

3.3.3. Quantifying redundancy

The issue of redundancy is considered in more detail now. Redundancy can be defined with reference to a multiple channel of capacity $T(C)$ which can be decomposed into C separate channels of capacities T_i, $i = 1, \ldots, C$:

$$R = 1 - \frac{T(C)}{\sum_i T_i} \tag{25}$$

so that when the C channels are multiplexed with maximal efficiency, $T(C) = \sum_i T_i$ and $R = 0$. What is measured more easily, in practice, is the redundancy defined with reference to a specific source (the set of stimuli with their probabilities). Then in terms of mutual information

$$R' = 1 - \frac{I(C)}{\sum_i I_i}. \tag{26}$$

Gawne and Richmond (1993) measured the redundancy R' among pairs of nearby primate inferior temporal cortex visual neurons, in their response to a set of 32 Walsh patterns. They found values with a mean $\langle R' \rangle = 0.1$ (and a mean single-cell transinformation of 0.23 bits). Since to discriminate 32 different patterns takes 5 bits of information, in principle one would need at least 22 cells each providing 0.23 bits of strictly orthogonal information to represent the full entropy of the stimulus set. Gawne and Richmond reasoned, however, that, because of the overlap, y, in the information they provided, more cells would be needed than if the redundancy had been zero. They constructed a simple model based on the notion that the overlap, y, in the information provided by any two cells in the population always corresponds to the average redundancy measured for nearby pairs. A redundancy $R' = 0.1$ corresponds to an overlap $y = 0.2$ in the information provided by the two neurons, since, counting the overlapping information only once, two cells would yield 1.8 times the amount

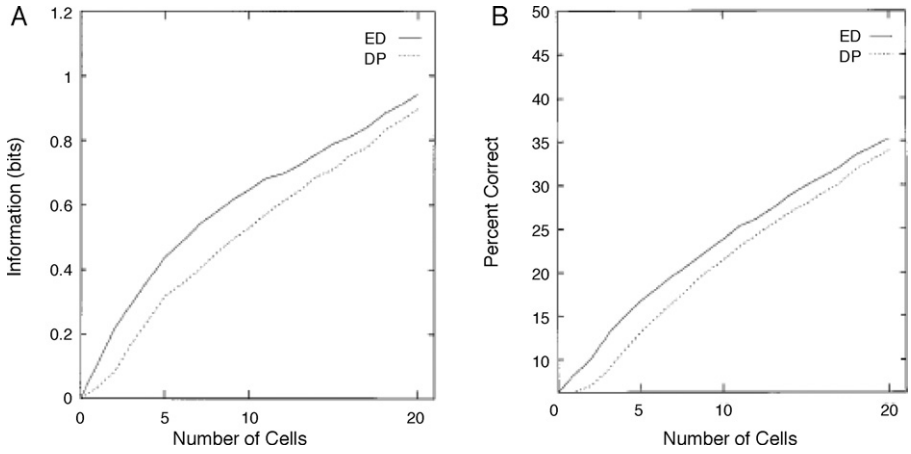

Fig. 21. Multiple cell information of spatial view cells in the primate hippocampus. (a) The values for the average information, $I(S, S')$, available in the responses of different numbers of hippocampal spatial view neurons on each trial, about which of the 16 'stimuli' (i.e. quarters of walls) is being looked at. The Euclidean Distance decoding algorithm was used for estimating the relative probability of posited stimuli s' (solid line); the Dot Product result is shown with the dashed line. The 20 cells were recorded from the same (av) animal. (b) The percent correct predictions based on the same data used in (a).
After Rolls et al. (1998).

transmitted by one cell alone. If a fraction of $1 - y = 0.8$ of the information provided by a cell is novel with respect to that provided by another cell, a fraction $(1 - y)^2$ of the information provided by a third cell will be novel with respect to what was known from the first pair, and so on, yielding an estimate of $I(C) = I(1)\sum_{i=0}^{C-1}(1 - y)^i$ for the total information conveyed by C cells. However such a sum saturates, in the limit of an infinite number of cells, at the level $I(\infty) = I(1)/y$, implying in their case that even with very many cells, no more than 0.23/0.2 = 1.15 bits could be read off their activity, or less than a quarter of what was available as entropy in the stimulus set! Gawne and Richmond (1993) concluded, therefore, that the average overlap among non-nearby cells must be considerably lower than that measured for cells close to each other.

The model above is simple and attractive, but experimental verification of the actual scaling of redundancy with the number of cells entails collecting the responses of several cells interspersed in a population of interest. Gochin et al. (1994) recorded from up to 58 cells in the primate temporal visual cortex, using sets of two to five visual stimuli, and applied decoding procedures to measure the information content in the population response. The recordings were not simultaneous, but comparison with simultaneous recordings from a smaller number of cells indicated that the effect of recording the individual responses on separate trials was minor. The results were expressed in terms of the *novelty N* in the information provided by C cells, which being defined as the ratio of such information to C times the average single-cell information, can be expressed as

$$N = 1 - R' \tag{27}$$

and is thus the complement of the redundancy. An analysis of two different data sets, which included three information measures per data set, indicated a behaviour $N(C) \approx 1/\sqrt{C}$, reminiscent of the improvement in the overall noise-to-signal ratio characterizing C independent processes contributing to the same signal. The analysis neglected however to consider limited sampling effects, and more seriously it neglected to consider saturation effects due to the information content approaching its ceiling, given by the entropy of the stimulus set. Since this ceiling was quite low, for 5 stimuli at $\log_2 5 = 2.32$ bits, relative to the mutual information values measured from the population (an average of 0.26 bits, or 1/9 of the ceiling, was provided by single cells), it is conceivable that the novelty would have taken much larger values if larger stimulus sets had been used.

A simple formula describing the approach to the ceiling, and thus the saturation of information values as they come close to the entropy of the stimulus set, can be derived from a natural extension of the Gawne and Richmond (1993) model. In this extension, the information provided by single cells, measured as a fraction of the ceiling, is taken to coincide with the average overlap among pairs of randomly selected, not necessarily nearby, cells from the population. The actual value measured by Gawne and Richmond would have been, again, 1/22 = 0.045, below the overlap among nearby cells, $y = 0.2$. The assumption that y, measured across any pair of cells, would have been as low as the fraction of information provided by single cells is equivalent to conceiving of single cells as 'covering' a random portion y of information space, and thus of randomly selected pairs of cells as overlapping in a fraction $(y)^2$ of that space, and so on, as postulated by the Gawne and Richmond (1993) model, for higher numbers of cells. The approach to the ceiling is then described by the formula

$$I(C) \approx H\{1 - \exp[C \ln(1 - y)]\} \tag{28}$$

that is, a simple exponential saturation to the ceiling. This simple law indeed describes remarkably well the trend in the data analyzed by Rolls et al. (1997b).

Although the model has no reason to be exact, and therefore its agreement with the data should not be expected to be accurate, the crucial point it embodies is that deviations from a purely linear increase in information with the number of cells analyzed are due solely to the ceiling effect. Aside from the ceiling, due to the sampling of an information space of finite entropy, the information contents of different cells' responses are independent of each other. Thus, in the model, the observed redundancy (or indeed the overlap) is purely a consequence of the finite size of the stimulus set. If the population were probed with larger and larger sets of stimuli, or more precisely with sets of increasing entropy, and the amount of information conveyed by single cells were to remain approximately the same, then the fraction of space 'covered' by each cell, again y, would get smaller and smaller, tending to eliminate redundancy for very large stimulus entropies (and a fixed number of cells). The actual data were obtained with limited numbers of stimuli, and therefore cannot probe directly the conditions in which redundancy might reduce to zero. The data are consistent, however, with the hypothesis embodied in the simple model, as shown also by the near exponential approach to lower ceilings found for information values calculated with reduced

subsets of the original set of stimuli (Rolls et al., 1997b) (see further Samengo and Treves (2000)).

The implication of this set of analyses, some performed towards the end of the ventral visual stream of the monkey, is that the representation of at least some classes of objects in those areas is achieved with minimal redundancy by cells that are allocated each to analyze a different aspect of the visual stimulus. This minimal redundancy is what would be expected of a self-organizing system in which different cells acquired their response selectivities through a random process, with or without local competition among nearby cells (Rolls, 2008). At the same time, such low redundancy could also very well result in a system that is organized under some strong teaching input, so that the emerging picture is compatible with a simple random process, but could be produced in other ways. The finding that, at least with small numbers of neurons, redundancy may be effectively minimized, is consistent not only with the concept of efficient encoding, but also with the general idea that one of the functions of the early visual system is to progressively minimize redundancy in the representation of visual stimuli (Attneave, 1954; Barlow, 1961). However, the ventral visual system does much more than produce a non-redundant representation of an image, for it transforms the representation from an image to an invariant representation of objects (Rolls, 2008). Moreover, what is shown in this section is that the information about objects can be read off from just the spike count of a population of neurons, using decoding as simple as the simplest that could be performed by a receiving neuron, dot product decoding. In this sense, the information about objects is made explicit in the firing rate of the neurons in the inferior temporal cortex, in that it can be read off in this way.

We consider in Section 3.3.6 whether there is more to it than this. Does the synchronization of neurons (and it would have to be stimulus-dependent synchronization) add significantly to the information that could be encoded by the number of spikes, as has been suggested by some?

Before this, we consider why encoding by a population of neurons is more powerful than the encoding than is possible by single neurons, adding to previous arguments that a distributed representation is much more computationally useful than a local representation, by allowing properties such as generalization, completion, and graceful degradation in associative neuronal networks (Rolls, 2008).

3.3.4. Should one neuron be as discriminative as the whole organism?

In the analysis of random dot motion with a given level of correlation among the moving dots, single neurons in area MT in the dorsal visual system of the primate can be approximately as sensitive or discriminative as the psychophysical performance of the whole animal (Zohary et al., 1994). The arguments and evidence presented here (e.g. in Section 3.3) suggest that this is not the case for the ventral visual system, concerned with object identification. Why should there be this difference?

Rolls and Treves (1998) suggest that the dimensionality of what is being computed may account for the difference. In the case of visual motion (at least in the study referred to), the problem was effectively one-dimensional, in that the direction of motion of the stimulus along a line in 2D space was extracted from the activity of the neurons. In this low-dimensional stimulus space, the neurons may each perform one of the few similar computations on a particular (local) portion of 2D space, with the side effect that, by averaging over a larger receptive field than in V1, one can extract a signal of a more global nature. Indeed, in the case of more global motion, it is the average of the neuronal activity that can be computed by the larger receptive fields of MT neurons that specifies the average or global direction of motion.

In contrast, in the higher dimensional space of objects, in which there are very many different objects to represent as being different from each other, and in a system that is not concerned with location in visual space but on the contrary tends to be relatively invariant with respect to location, the goal of the representation is to reflect the many aspects of the input information in a way that enables many different objects to be represented, in what is effectively a very high dimensional space. This is achieved by allocating cells, each with an intrinsically limited discriminative power, to sample as thoroughly as possible the many dimensions of the space. Thus the system is geared to use efficiently the parallel computations of all its neurons precisely for tasks such as that of face discrimination, which was used as an experimental probe. Moreover, object representation must be kept higher dimensional, in that it may have to be decoded by dot product decoders in associative memories, in which the input patterns must be in a space that is as high-dimensional as possible (i.e. the activity on different input axons should not be too highly correlated). In this situation, each neuron should act somewhat independently of its neighbours, so that each provides its own separate contribution that adds together with that of the other neurons (in a linear manner, see above and Figs. 19 and 20) to provide *in toto* sufficient information to specify which out of perhaps several thousand visual stimuli was seen. The computation involves in this case not an average of neuronal activity (which would be useful for, e.g. head direction (Robertson et al., 1999)), but instead comparing the dot product of the activity of the population of neurons with a previously learned vector, stored in, for example, associative memories as the weight vector on a receiving neuron or neurons.

Zohary et al. (1994) put forward another argument which suggested to them that the brain could hardly benefit from taking into account the activity of more than a very limited number of neurons. The argument was based on their measurement of a small (0.12) correlation between the activity of simultaneously recorded neurons in area MT. They suggested that there would because of this be decreasing signal-to-noise ratio advantages as more neurons were included in the population, and that this would limit the number of neurons that it would be useful to decode to approximately 100. However, a measure of correlations in the activity of different neurons depends entirely on the way the space of neuronal activity is sampled, that is on the task chosen to probe the system. Among face cells in the temporal cortex, for example, much higher correlations would be observed when the task is a simple two-way discrimination between a face and a non-face, than when the task involves finer identification of several different faces. (It is also entirely possible that some face cells could be found that perform as well in a given particular face/non-face discrimination as the whole animal.) Moreover, their argument depends on the type of decoding of the activity of the population that is envisaged (see further Robertson et al. (1999)). It implies that the average of the neuronal activity must be estimated accurately. If a set of neurons uses dot product decoding, and then the activity of the decoding population is scaled or normalized by some negative feedback through inhibitory interneurons, then the effect of such correlated firing in the sending population is reduced, for the decoding effectively measures the relative firing of the different neurons in the population to be decoded. This is equivalent to measuring the angle between the current vector formed by the population of neurons firing, and a previously learned vector, stored in synaptic weights. Thus, with for example this biologically plausible decoding, it is not clear whether the correlation Zohary et al. (1994) describe would place a severe limit on the ability of the brain to utilize the information available in a population of neurons.

The main conclusion from this and the preceding section is that the information available from a set or ensemble of temporal cortex visual neurons increases approximately linearly as more neurons are added to the sample. This is powerful evidence that distributed encoding is used by the brain; and the code can be read just by knowing the firing rates in a short time of the population of neurons. The fact that the code can be read off from the firing rates, and by a principle as simple and neuron-like as dot product decoding, provides strong support for the general approach taken to brain function (Rolls, 2008).

It is possible that more information would be available in the relative time of occurrence of the spikes, either within the spike train of a single neuron, or between the spike trains of different neurons, and it is to this that we now turn.

3.3.5. Information representation in the taste and olfactory systems

Similar principles to those described here for the representation of visual stimuli in the primate inferior temporal visual cortex also apply to the representation of taste and olfactory information in the cortical areas for taste and olfaction in primates (Rolls et al., 2010a). Information theory analysis shows a robust representation of taste in the primate orbitofrontal cortex, with an average mutual information of 0.45 bits for each neuron about which of 6 tastants (glucose (sweet), NaCl (salt), HCl (sour), quinine-HCl (bitter), monosodium glutamate (umami (Rolls, 2009)), and water) was present, averaged across 135 gustatory neurons. The information increased with the number of neurons in the ensemble, but less than linearly, reflecting some redundancy. There was less information per neuron about which of 6 odors was present from orbitofrontal cortex olfactory neurons, but the code was robust, in that the information increased linearly with the number of neurons, reflecting independent information encoded by different neurons (Rolls et al., 2010a).

3.3.6. The effects of cross-correlations between cells

Using the second derivative methods described in Section 2.6.2 (see Rolls et al. (2003b)), the information available from the number of spikes vs. that from the cross-correlations between simultaneously recorded cells has been analyzed for a population of neurons in the inferior temporal visual cortex (Rolls et al., 2004). The stimuli were a set of 20 objects, faces, and scenes presented while the monkey performed a visual discrimination task. If synchronization was being used to bind the parts of each object into the correct spatial relationship to other parts, this might be expected to be revealed by stimulus-dependent cross-correlations in the firing of simultaneously recorded groups of 2–4 cells using multiple single-neuron microelectrodes.

The results for the 20 experiments with groups of 2–4 simultaneously recorded inferior temporal cortex neurons are shown in Table 3. (The total information is the total from Eqs. (21) and (22) in a 100 ms time window, and is not expected to be the sum of the contributions shown in Table 3 because only the information from the cross terms (for $i \neq j$) is shown in the table for the contributions related to the stimulus-dependent contributions and the stimulus-independent contributions arising from the 'noise' correlations.) The results show that the greatest contribution to the information is that from the rates, that is from the numbers of spikes from each neuron in the time window of 100 ms. The average value of −0.05 bits for the cross term of the stimulus independent 'noise' correlation-related contribution is consistent with on average a small amount of common input to neurons in the inferior temporal visual cortex. A positive value for the cross term of the stimulus-dependent 'noise' correlation related contribution would be consistent with on average a small amount of stimulus-dependent synchronization, but the actual value found, 0.04 bits, is so small that for 17 of the 20 experiments it is less than that which can arise by chance statistical fluctuations of the time of arrival of the spikes, as shown by MonteCarlo control rearrangements of the same data. Thus on average there was no significant contribution to the information from stimulus-dependent synchronization effects (Rolls et al., 2004).

Table 3
The average contributions (in bits) of different components of equations 21 and 22 to the information available in a 100 ms time window from 13 sets of simultaneously recorded inferior temporal cortex neurons when shown 20 stimuli effective for the cells.

Rate	0.26
Stimulus-dependent "noise" correlation-related, cross term	0.04
Stimulus-independent "noise" correlation-related, cross term	−0.05
Total information	0.31

Thus, this data set provides evidence for considerable information available from the number of spikes that each cell produces to different stimuli, and evidence for little impact of common input, or of synchronization, on the amount of information provided by sets of *simultaneously recorded* inferior temporal cortex neurons. Further supporting data for the inferior temporal visual cortex are provided by Rolls et al. (2003b). In that parts as well as whole objects are represented in the inferior temporal cortex (Perrett et al., 1982), and in that the parts must be bound together in the correct spatial configuration for the inferior temporal cortex neurons to respond (Rolls et al., 1994), we might have expected temporal synchrony, if used to implement feature binding, to have been evident in these experiments.

3.3.7. Stimulus-dependent neuronal synchrony is not used for binding even with natural vision and attention

We have also explored neuronal encoding under natural scene conditions in a task in which top-down attention must be used, a visual search task. We applied the decoding information theoretic method of Section 2.6 to the responses of neurons in the inferior temporal visual cortex recorded under conditions in which feature binding is likely to be needed, that is when the monkey had to choose to touch one of the two simultaneously presented objects, with the stimuli presented in a complex natural background (Aggelopoulos et al., 2005). The investigation is thus directly relevant to whether stimulus-dependent synchrony contributes to encoding under natural conditions, and when an attentional task was being performed. In the attentional task, the monkey had to find one of the two objects and to touch it to obtain reward. This is thus an object-based attentional visual search task, where the top-down bias is for the object that has to be found in the scene (Aggelopoulos et al., 2005). The objects could be presented against a complex natural scene background. Neurons in the inferior temporal visual cortex respond in some cases to object features or parts, and in other cases to whole objects provided that the parts are in the correct spatial configuration (Perrett et al., 1982; Desimone et al., 1984; Rolls et al., 1994; Tanaka, 1996; Rolls, 2008, 2011b), and so it is very appropriate to measure whether stimulus-dependent synchrony contributes to information encoding in the inferior temporal visual cortex when two objects are present in the visual field, and when they must be segmented from the background in a natural visual scene, which are the conditions in which it has been postulated that stimulus-dependent synchrony would be useful (Singer, 1999, 2000).

Aggelopoulos et al. (2005) found that between 99% and 94% of the information was present in the firing rates of inferior temporal cortex neurons, and less that 5% in any stimulus-dependent synchrony that was present, as illustrated in Fig. 22. The implication of these results is that any stimulus-dependent synchrony that is present is not quantitatively important as measured by information theoretic analyses under natural scene conditions. This has been found for the inferior temporal visual

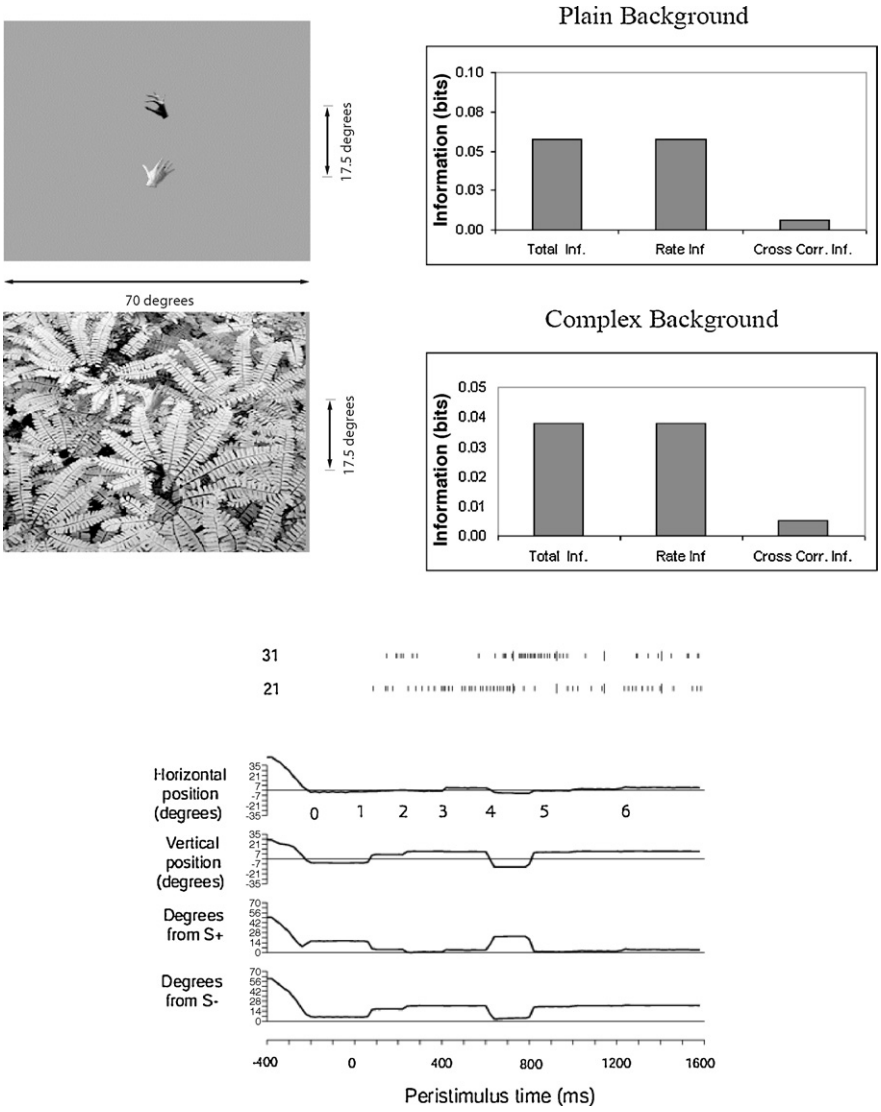

Fig. 22. Left: the objects against the plain background, and in a natural scene. Right: the information available from the firing rates (Rate Inf) or from stimulus-dependent synchrony (Cross-Corr Inf) from populations of simultaneously recorded inferior temporal cortex neurons about which stimulus had been presented in a complex natural scene. The total information (Total Inf) is that available from both the rate and the stimulus-dependent synchrony, which do not necessarily contribute independently. Bottom: eye position recordings and spiking activity from two neurons on a single trial of the task. (Neuron 31 tended to fire more when the macaque looked at one of the stimuli, S−, and neuron 21 tended to fire more when the macaque looked at the other stimulus, S+. Both stimuli were within the receptive field of the neuron). After Aggelopoulos et al. (2005).

cortex, a brain region where features are put together to form representations of objects (Rolls and Deco, 2002), where attention has strong effects, at least in scenes with blank backgrounds (Rolls et al., 2003a), and in an object-based attentional search task.

The finding as assessed by information theoretic methods of the importance of firing rates and not stimulus-dependent synchrony is consistent with previous information theoretic approaches (Rolls et al., 2003b, 2004; Franco et al., 2004). It would of course also be of interest to test the same hypothesis in earlier visual areas, such as V4, with quantitative, information theoretic, techniques. In connection with rate codes, it should be noted that the findings indicate that the number of spikes that arrive in a given time is what is important for very useful amounts of information to be made available from a population of neurons; and that this time can be very short, as little as 20–50 ms (Tovee and Rolls, 1995; Rolls and Tovee, 1994; Rolls et al., 1999, 1994, 2006; Rolls and Deco, 2002; Rolls, 2003). Further, it was shown that there was little redundancy (less than 6%) between the information provided by the spike counts of the simultaneously recorded neurons, making spike counts an efficient population code with a high encoding capacity.

3.3.8. Conclusions on feature binding in vision

The findings (Aggelopoulos et al., 2005; Rolls et al., 2004) are consistent with the hypothesis that feature binding is implemented by neurons that respond to features in the correct relative spatial locations (Rolls and Deco, 2002; Elliffe et al., 2002; Rolls, 2008), and not by temporal synchrony and attention (Malsburg, 1990; Singer et al., 1990; Abeles, 1991; Hummel and Biederman, 1992; Singer and Gray, 1995; Singer, 1999, 2000).

In any case, the computational point is that even if stimulus-dependent synchrony was useful for grouping, it would not without much extra machinery be useful for binding the relative spatial positions of features within an object (Rolls, 2008), or for that matter of the positions of objects in a scene which appears to be encoded in a different way, by neurons that respond to combinations of stimuli when they are in the correct relative spatial positions (Aggelopoulos and Rolls, 2005; Rolls et al., 2008c;

Rolls, 2008). The greatest computational problem is that synchronization does not by itself define the spatial relations between the features being bound, so is not just as a binding mechanism adequate for shape recognition. For example, temporal binding might enable features 1, 2 and 3, which might define one stimulus to be *grouped* together and kept separate from for example another stimulus consisting of features 2, 3 and 4, but would require a further temporal binding (leading in the end potentially to a combinatorial explosion) to indicate the relative spatial positions of the 1, 2 and 3 in the 123 stimulus, so that it can be discriminated from, e.g. 312 (Rolls, 2008).

Similar conclusions have been reached in visual motion area MT, where synchrony in spiking activity shows little dependence on feature grouping, whereas gamma band synchrony in local field potentials (LFP) can be significantly stronger when features are grouped (Palanca and DeAngelis, 2005). However, these changes in gamma band synchrony are small relative to the variability of synchrony across recording sites and do not provide a robust population signal for feature grouping. Moreover, these effects are reduced when stimulus differences nearby the receptive fields are eliminated using partial occlusion. These findings suggest that synchrony does not constitute a general mechanism of visual feature binding (Palanca and DeAngelis, 2005). Further, in MT coherent plaids (for which binding may be needed) elicited less stimulus-dependent neuronal synchrony than did non-coherent plaids (Thiele and Stoner, 2003). Similarly in V1, recordings from pairs of V1 recording sites while presenting either single or separate bar stimuli indicated that between 89% and 96% of the information was carried by firing rates; correlations contributed only 4–11% extra information (Golledge et al., 2003). The distribution across the population of either correlation strength or correlation information did not co-vary systematically with changes in perception predicted by Gestalt psychology. These results suggest that firing rates, rather than correlations, are the main element of the population code for feature binding in primary visual cortex (Golledge et al., 2003).

If there is synchrony that is present caused for example by oscillations (Deco and Rolls, in preparation) and that is stimulus-independent, this can, because of redundancy, decrease the information available in the ensemble of neurons (Oram et al., 1998; Rolls et al., 2004). In practice, such trial-by-trial stimulus-independent "noise" correlations that include any effects of synchrony lead to the loss of just a few percent of the information that would otherwise be available from IT neurons (Rolls et al., 2004).

3.4. Information about physical space

We have seen in the preceding sections that it is difficult to extract information measures from the activity of populations of neurons, because a large response space implies an enormous (exponentially large) number of repetitions of the same conditions, to be adequately sampled. Then, when more than 2–3 neurons (or 2–3 aspects of the activity of a single neuron) are to be considered together, the two effective approaches are either to consider the information conveyed in very short time periods, through the derivative approach (Section 2.6.2), or the information present in the confusion matrix, obtained after a decoding step, which brings us from the response space back to the stimulus or external correlate space (Section 2.6). Neither of these two approaches suffices, however, when the space of stimuli (or, in general, external correlates) is itself high-dimensional, or even low dimensional but effectively large or continuous. A prominent example is physical space.

When the information encoded by the neuronal population is the position of an object in space, or of the animal in space, as with 'place cells' (O'Keefe and Dostrovsky, 1971), or of the animal's gaze in space, as in 'spatial view cells' (Rolls et al., 1997a, 1998; Rolls and Xiang, 2006), there are very many positions potentially discriminable by decoding population activity. For example, a rat free foraging in a 1 sq m box can be, even discretizing space at the relatively gross resolution of 5 cm, in $20 \times 20 = 400$ distinct positions. The confusion matrix generated by a straightforward decoding algorithm has $400 \times 400 = 160,000$ elements. If hippocampal activity is sampled once per theta cycle (Section 3.2.4, Jezek et al. (2011)), sufficient sampling of the confusion matrix for the purposes of extracting information measures requires of the order of 1 million temporal samples (population vectors in a theta cycle), that is 1–2 days of continuous recording! This is feasible with neural network simulations (Cerasti and Treves, 2010) but not with recordings in vivo. Halving the linear spatial resolution to a still moderate 2.5 cm quadruples the number of positions and multiplies the recording time required by 16.

Clearly, the curse of dimensionality has moved from the response space to physical space, when physical space is what is encoded by the neuronal responses. What approaches are available to still address quantitatively this important correlate of neural activity?

3.4.1. Information about spatial context

One possible approach is to restrict the analysis to the encoding of spatial context, intended as a larger portion of space than the exact position of the object, or animal, or spatial view. If there are few possible contexts, the curse of dimensionality does not apply. Several experimental paradigms lend themselves naturally to a discretization of continuous space. For example, with a rodent foraging in a few different boxes, one can ask how well place cell activity discriminates among boxes, irrespective of the position of the animal within each box. With 3 boxes, for example, the confusion matrix reduces to 3×3, which in terms of size is easy to sample. The challenge, however, lies in the fact that place cell activity in a single box is highly inhomogeneous, with different ensembles of units active at different positions within a box. If one thinks in terms of population vectors, the population vectors sampled at one position in a box may form a more or less loose cluster, but those sampled at all the different positions in a box are likely to be scattered in a diffuse cloud, which may be hard to separate from the distribution of population vectors occurring in a different box. Here, too, a way to address the challenge is to consider temporal bins short relative to the inverse of the average firing rates of the population of units being decoded, as in the procedure developed by Fyhn et al. (2007).

In that analysis, population vectors were extracted with 150 ms bins, as the list of neurons that fired at least 1 spike in the bin (in a control analysis, the units that fired exactly 1 spike and those that fired 2 or more spikes were considered separately). Recording from 5 to 7 grid cells in medial entorhinal cortex (mEC) and some 25–30 place cells in CA3, typically each spatial context was represented by of order a hundred distinct binary population vectors (or double or so ternary ones). This is because 10–12 units may fire somewhere within a box, and rarely more than two in the same 150 ms temporal bin. Allowing rats to forage for 10 min in each box, or roughly 3600 temporal bins, is sufficient to sample adequately a hundred or so distinct possible responses. Therefore the question of whether distinct spatial contexts are represented in neural activity reduces to determining to what extent the manifolds spanned by population vectors in the different boxes overlap with each other, or in other words to what extent different boxes produce distinct coactivity patterns. With place cells, given the tens of units in practice sampled in a typical recording session and their sparseness, it is the single and pair-wise coactivity patterns that largely determine the outcome of the analysis.

In rats with simultaneous recordings from mEC and CA3, the mean Shannon mutual information between recording session and distribution of population vectors demonstrated strikingly diverging encoding schemes. In mEC, the mutual information was no larger for comparisons of sessions in different boxes than for comparisons of repeated sessions in the same box (and also not larger even for sessions run in entirely different rooms), indicating that grid cells that are coactive in one context remain coactive in others, too. In CA3, the mutual information about which box the rat was foraging in a session was above 0.5 bits when the boxes (or the rooms) were different, well above the values obtained for repeated sessions in the same spatial context (Fyhn et al., 2007). A shuffling procedure allowed for bootstrapping validation of the result: information values in mEC were much lower than when cell identities were randomly shuffled between sessions, whereas in CA3 they were not statistically different from the values obtained after random shuffling of cell identities. Therefore coactivity patterns are largely overlapping, or covering roughly the same manifold, in mEC, whereas they differentiate in CA3 (Fig. 23, Leutgeb et al. (2004)), supporting the notion that the dentate gyrus acts as a sort of random spatial pattern generator, to decorrelate spatial information as it enters the hippocampal system and establishes representations in CA3 (Cerasti and Treves, 2010; Treves and Rolls, 1992; Rolls, 1989, 2010).

3.4.2. Information about position from individual cells

The above observations indicate that in order to understand how spatial information is transformed in different regions of the brain, the coactivity patterns among different units have to be taken into account. If one is focusing on spatial codes expressed over very short times, however, the derivative approach of Sections 2.3 and 2.6.2 shows that even pair-wise coactivity pattern can be neglected: the first order terms in a Taylor expansion in the length of the temporal window, taken to be vanishingly short, are simply the contributions from individual units.

One can then use Eq. (21), with the linear summation of the contribution of different units. It is customary to divide the contribution of each unit to I_t by its mean firing rate, obtaining what is called the information per spike, introduced first by Bill Skaggs (Skaggs and McNaughton, 1992). This is nothing but a scaled version of the individual terms in the time derivative, but one should bear in mind that in order to add several individual terms, one must use bit/s and not bit/spike.

Contributions of order a few bits/s are typical of place cells in rodents. The fact that place cells are commonly assumed to code

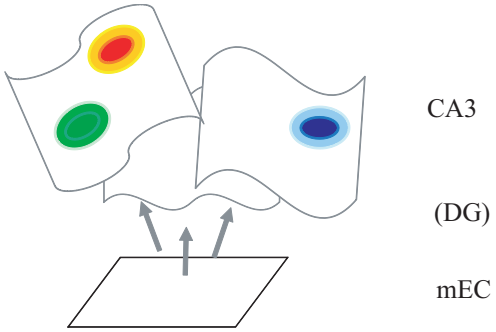

Fig. 23. Schematic of the decorrelation of spatial manifolds, observed as spatial information is recoded from rodent medial entorhinal cortex (mEC), where for all practical purposes one can think of allocentic space as being represented by a single map, to area CA3 of the hippocampus, where neuronal activity can be described as spanning multiple, uncorrelated maps (Leutgeb et al., 2004). Each map represents several locations in one physical environment, ideally through a continuous 2D attractor. Colored areas represent two place fields in one map and one in another map. The crucial decorrelation operation is ascribed to the randomizing effect of the dentate gyrus (DG) as in the model by Cerasti and Treves (2010).

information relatively independently of each other, aside from common modulation by theta and other rhythms, makes neglecting pairwise and higher-order correlations a very reasonable approximation.

3.4.3. Information about position, transparent and dark

In contrast to neuronal recordings, computer simulations can be carried on indefinitely, and they allow the quantification of information about exact position in space. They also allow, in principle, a comparison with model-based analytical calculations, although the comparison is not completely straightforward (Cerasti and Treves, 2010).

An interesting result emerging from computer simulations is the distinction between spatial information that is transparent, i.e. expressed in purely spatial terms and easily read out, and spatial information that is implicit in more convoluted codes, admixed with and contaminated by non-spatial information. The measures of mutual information that can be extracted from the simulations are, in fact, strongly dependent on the method used, in the decoding step, to construct the *localization matrix*, i.e. the confusion matrix specialized to the case of information about position, which compiles the frequency with which actual position x_0 was decoded as position $x_0 + \Delta x$. In the general case, applicable also to non-spatial codes, information measures are obtained constructing the full confusion matrix $Q(x_0, x_0 + \Delta x)$ which, if again one considers a square environment discretized into 20×20 spatial bins, is a large 400×400 matrix, which requires of order hundreds of thousands of decoding events to be effectively sampled, even after applying a correction for limited sampling. An alternative available in the spatial case, that allows extracting unbiased measures from much shorter simulations, is to construct a simplified matrix $Q(\Delta x)$, which averages over decoding events with the same vector displacement between actual and decoded positions. $Q(\Delta x)$ is easily constructed, and if the simulated environment is a torus, i.e. with periodic boundary conditions, it ends up being a much smaller 20×20 matrix which is effectively sampled in just a few thousand steps.

The two decoding procedures, given that the simplified matrix is the shifted average of the rows of the full matrix, might be expected to yield similar measures, but they do not, as shown in Fig. 24. The simplified matrix, by assuming translation invariance of the errors in decoding, is unable to quantify the information implicitly present in the full distribution of errors around each actual position. Such errors are of an *episodic* nature: the local view from position $x_0 + \Delta x$ might happen to be similar to that from position x_0, hence neural activity reflecting in part local views might lead to confuse the two positions, but this does not imply that another position z_0 has anything in common with $z_0 + \Delta x$. Fig. 24 shows that for any actual position there are a few selected positions that are likely to be erroneously decoded from the activity of a given sample of units; when constructing instead the translationally invariant simplified matrix, all average errors are distributed smoothly around the correct position (zero error), in a roughly Gaussian bell. The upper right panel in Fig. 24, reproduced from the simulations by Cerasti and Treves (2010), shows that such episodic information prevails. The lower right panel in the figure compares, instead, the entropies of the decoded positions with the two matrices, conditioned on the actual position – that is, the equivocation values. Unlike the mutual information, such equivocation is much higher for the simplified matrix; for this matrix, it is simply a measure of how widely displaced are decoded positions, with respect to the actual positions, represented at the center of the square; and for small samples of units, which are not very informative, the *displacement* entropy approaches that of a flat distribution of decoded positions, i.e. $\log_2(400) \approx 8.64$ bits. For larger samples, which enable better localization, the simplified

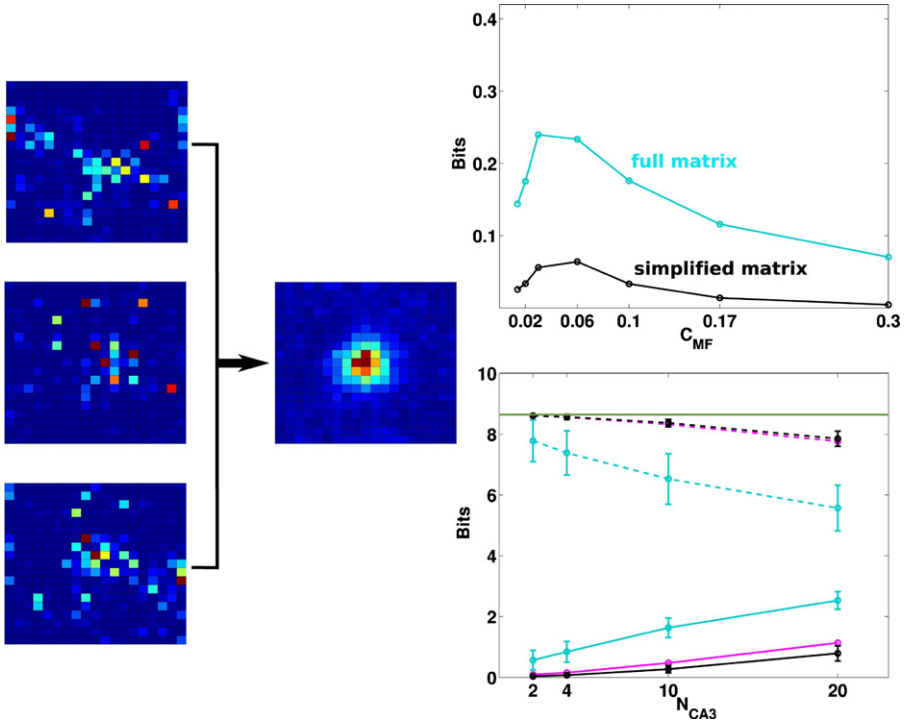

Fig. 24. Localization matrices, after Cerasti and Treves (2010). Left: the rows of the full matrix represent the actual positions of the virtual rat while its columns represent decoded positions (the full matrix is actually 400×400); three examples of rows are shown, rendered here as 20×20 squares, all from decoding a given sample of 10 units. The simplified matrix is a single 20 × 20 matrix obtained (from the same sample) as the average of the full matrix taking into account translation invariance. Right, top: the two procedures lead to large quantitative differences in information (here, the measures from samples of 10 units, divided by 10, from the full matrix, cyan, and from the simplified matrix, black), but with the same dependence on the number of mossy fiber connections per CA3 cell C_{MF}. Right, bottom: The conditional entropies of the full and simplified localization matrices (cyan and black, dashed, calculated over different samples of N_{CA3} units) in both cases add up to the respective mutual information measure (cyan and black, solid) to give the full entropy of log (400) ≈ 8.64 bits (green line). The conditional entropy calculated from the full matrix averaged across samples (red, dashed) is equivalent to that calculated from the displacements, for each sample (black, dashed).

localization matrix begins to be clustered in a Gaussian bell around zero displacement, so that the equivocation gradually decreases (the list of displacements, with their frequencies, is computed for each sample, and it is the equivocation, not the list itself, which is averaged across samples).

In contrast, the entropy of each row of the full localization matrix, i.e. the entropy of decoded positions conditioned on any actual position, is lower, and also decreasing more steeply with sample size; it differs from the full entropy, in fact, by the mutual information between decoded and actual positions, which increases with sample size. The two equivocation measures therefore both add up to the two mutual information measures to yield the same full entropy of about 8.64 bits (a bit less in the case of the full matrix, where the sampling is more limited), and thus serve as controls that the difference in mutual information is not due, for example, to inaccuracy. As a third crucial control, also the average conditional entropy of the full localization matrix was calculated, when the matrix is averaged across samples of a given size: the resulting entropy is virtually identical to the displacement entropy (which implies instead an average of the full matrix across rows, i.e. across actual positions). This indicates that different samples of units express distinct episodic content at each location, such that averaging across samples is equivalent to averaging across locations.

These distinctions do not alter the other results of the study by Cerasti and Treves (2010), since they affect the height of the curves, not their dependence, e.g. on the connectivity, however they have important implications. The simplified matrix has the advantage of requiring much less data, i.e. less simulation time, but also less real data if applied to neurophysiological recordings, than the full matrix, and in most situations it might be the only feasible measure of spatial information (analytical estimates are not available of course for real data). So in most cases it is only practical to measure spatial information with methods that, the model suggests, miss out much of the information present in neuronal activity, what we may refer to as *dark information*, not easily revealed. One might conjecture that in the specific case analyzed by Cerasti and Treves (2010), the prevalence of dark information is linked to the random nature of the spatial code established by DG inputs. It might be that additional stages of hippocampal processing, either with the refinement of recurrent CA3 connections or in CA1, are instrumental in making dark information more transparent.

3.5. Information in virtual space

When the external correlates encoded in neural activity do not span physical space, or a manifold with a natural metric, one may still study the metric of the virtual manifold established by the patterns of neuronal activity with which they are associated. If a number of discrete stimuli, faces for example, elicits on repeated trials a distribution of response population vectors in a particular cortical area, one may define distances between pairs of stimuli in terms of the overlaps in the corresponding distributions of population vectors, and analyze the overall structure of such pair-wise distances in geometric terms. A specific aspect of this general approach focuses solely on extracting a summary index quantifying the average overlap between patterns elicited by different stimuli. Since the absence of any overlap implies that no genuine metric structure can be defined (all stimuli would be effectively at maximum distance from each other, because differences in distances between mean population vectors, if their

distributions do not overlap, are irrelevant), such a quantity can be called an index of metric content. Conversely, any overlap results in the possibility of errors in decoding, so that one can formulate this type of analysis as first decoding neuronal activity, and then analyzing the confusion matrix expressing relations between actual and decoded stimuli.

3.5.1. The metric content index

An information theoretical approach focusing on an index of metric content index requires extracting, from a given neuronal representation of a discrete set of stimuli, the percent correctly decoded f_{corr} and the mutual information I between actual and decoded stimuli. Both quantities are obtained from the confusion matrix (f_{corr} is the sum of its diagonal elements), hence they do not have to be really based on an underlying neuronal representation: the confusion matrix can for example be obtained directly from behavioural responses, or from some other psychophysical measurement. In view of the application used for illustration below, and to avoid mentioning all the time the decoding step, we shall refer here to a confusion matrix obtained from the behaviour of human subjects, who classify items into categories. Nevertheless the framework remains completely general, and categories can be replaced with stimuli, and individual items with individual trials, as in our analysis of the metric content of hippocampal spatial view cells (Treves et al., 1999a).

For a given f_{corr}, I takes its theoretically minimum value when incorrect responses are evenly distributed among stimuli (Treves, 1997). Assuming the confusion matrix Q to have been constructed with sufficient statistics, this means that

$$Q(s, s' \neq s) = \frac{1 - f_{corr}}{S - 1}, \tag{29}$$

and

$$I_{min} = \log_2 S + f_{corr} \log_2 f_{corr} + (1 - f_{corr}) \log_2 \left[\frac{1 - f_{corr}}{S - 1} \right]. \tag{30}$$

The absolute maximum value of I for a given f_{corr}, on the other hand, is attained when all incorrect responses are grouped into a single category s' (different for each correct category s), in which case

$$I' = \log_2 S + f_{corr} \log_2 f_{corr} + (1 - f_{corr}) \log_2 (1 - f_{corr}). \tag{31}$$

This maximum however would correspond to a perverse systematic misclassification by the subject. A more useful reference value can be obtained by assuming unbiased classification (incorrect categories can at most be chosen as frequently as the correct one) and, for mathematical simplicity, a real (not integer) number of categories. Then the largest information value corresponds to the case in which all pictures are categorized in clusters of size $1/f_{corr}$ and, for each item, the cluster is correctly identified but the category inside it is selected at random (Treves, 1997). One finds in this case

$$I_{max} = \log_2 S + \log_2 f_{corr}. \tag{32}$$

If one then interprets the probability of misclassification as a monotonically decreasing function of some underlying perceived distance between the categories, in the minimum information scenario categories can be thought of as drawn from a space of extremely high dimensionality, so that they all tend to be at the same distance from each other; while in the maximum information case (which can be realized for example by an ultrametric or taxonomic classification (Treves, 1997)) categories which are at a distance less than some critical value from each other form clusters, while the distance between any two members of different clusters is above the critical value; therefore errors are more concentrated. This increases the mutual information value of the categorization for a given f_{corr}. Intermediate situations can be conveniently described by quantifying the relative amount of information for a given f_{corr} with the parameter

$$\lambda = \frac{I - I_{min}}{I_{max} - I_{min}}. \tag{33}$$

This *metric content* index ranges from 0 to about 1 (occasionally taking values above 1, see Fig. 26), and in all generality it quantifies the degree to which relationships of being 'close' or 'distant' among stimuli have been relevant to their perception and classification (Treves, 1997). For $\lambda = 0$ such relationships are irrelevant, and if a stimulus is misclassified the probability of assigning it to any of the wrong categories is the same. For $\lambda = 1$, categories can be thought of as clustering into an arbitrary but systematic semantic structure, while the particular category within each cluster is chosen at random.

To visualize the metric content measure, it is useful to plot f_{corr} and I with the additional lines indicating $I_{min}(f_{corr})$ and $I_{max}(f_{corr})$ (see Fig. 26a). The relative vertical excursion of a data point between these two lines represents the metric content of the classification by that subject. This kind of representation is particularly suitable to compare and analyze the performance of groups of subjects in the test described below, in that quantitative differences in the joint $f_{corr} - \lambda$ distribution are reflected in the different positions they occupy in the 'leaf' diagram.

3.5.2. Estimating metric content from human subjects' behaviour

The Famous Faces Multiple Choice Task (FFMCT (Lauro-Grotto et al., 1997a); see Fig. 25) requires the subject to classify a set of 54 pictures of famous people into 9 disjoint categories according to nationality (Italian, Other European and American) and field of activity (Sports people, Politicians, Actors and Singers). The picture categories are the 9 combinations of nationality by field of activity. Each category includes 6 famous faces from across the 20th century, 2 of whom became famous roughly in the 40s–50s, 2 in the 60s–70s and 2 in the 80s–90s. Prominent personalities were chosen on the basis of their fame, whereas they were portrayed in pictures spanning a range from easily recognizable to quite difficult. As a result, famous faces from the 50s were in principle recognizable even by subjects who were not alive when they became famous, and at the same time it was very difficult for any subject to achieve nearly perfect performance. Ensuring a substantial number of errors is of course a prerequisite in order to examine their distribution.

The performance of each subject can be described directly by the matrix $Q(s, s')$, the confusion matrix, reporting the frequency with which an image belonged to category s and was classified by the subject as s'. From $Q(s, s')$, one extracts $f_{corr} = \Sigma_s Q(s, s)$ and

$$I = \Sigma_{s,s'} Q(s, s') \log_2 \left[\frac{Q(s, s')}{P(s) Q(s')} \right] - C_1, \tag{34}$$

where $P(s) = 1/9$ is the *a priori* frequency of each category, $Q(s')$ is the marginal frequency of responses in category s' (cumulated over the actual category of each picture), and C_1 is a correction term that removes most of the bias due to using frequencies rather than probabilities (Panzeri and Treves, 1996).

When a picture is misclassified by the subject, it can still be assigned to the correct nationality or to the correct field of activity; it is also possible that the subject has a tendency to confuse, solely among Politicians, Americans with Other Europeans, or else, solely among Americans, Politicians with Actors and Singers; more in general, errors can be entirely random or they can be concentrated, to a varying degree, by incomplete semantic cueing. Unlike f_{corr}, I is sensitive to the concentration of the categories s' mistakenly

Fig. 25. The Famous Faces Memory Classification Task, developed by Lauro-Grotto et al. (2007).

assigned to each actual category s. However, since I measures the total (average) concentration of responses s' for each category s, it largely co-varies with f_{corr}, which measures their average concentration in the correct category s itself. Thus, to turn it into an effective measure of the concentration of errors only, the main dependence of I on f_{corr} can be removed by using, as explained above, the metric content index λ, which simply reflects the range of values I can take for a fixed f_{corr} (Treves, 1997). High levels of metric content indicate strong dependence of the classification performance on perceived relations among the set of stimuli, and therefore a preferred semantic access mode.

The metric content index is therefore in this case a measure of the amount of structure embedded in the neural representations that inform subject behaviour. It is high when individual memory items are classified using semantic cues, which leads to a more concentrated distribution of errors. It is low either when performance is random (in which case performance measures are also low), or when episodic access to the identity of each famous face is prevalent, semantic relationships remain largely unused, and errors, when made, tend to be more randomly distributed. It is important to note here that any tendency towards systematic misclassification, not only the correct identification of super-ordinates, is reflected in an increased metric content. For example, if a subject systematically confuses American Politicians with Italian Actors and Singers, due to their good looks, the corresponding λ value will be larger. Furthermore, subjects might be able to detect similarities in the data set that are more fine-grained than the explicit super-ordinates of nationality and field of activity: for example they could be prone to confuse Italian Politicians with Other European Politicians, but not with American Politicians. For these reasons, λ appears to represent a more effective and model-free measure of perceived semantic structure than the mere access to super-ordinate information.

We have found a significant effect of age on the metric content, shown in Fig. 26b, indicative of a shift from episodic to semantic access in older subjects; as well as a significant correlation between the metric content and relevant measures assessing episodic and semantic retrieval mode in the Remember (R)/Know (K) paradigm introduced by Tulving (1985) (Ciaramelli et al., 2006).

Metric content analysis can also be applied to neuronal representations, for example of spatial view by neurons in the hippocampus (Treves et al., 1999a). It would be of interest to apply it to neuronal representations in the inferior temporal visual cortex, in which while responding differently to different members within a category of, e.g. faces, also reflect categorical structure by for example not responding to non-faces, or to inanimate objects, etc (Rolls and Tovee, 1995; Rolls et al., 1997c; Kiani et al., 2007) (see example in Fig. 6).

3.5.3. Metric content increases with Alzheimer's but not with semantic dementia

A shifting balance between semantic and episodic memory can be assessed also by studying different groups of brain injured patients, who have been shown to present with distinct memory impairments. Patients with Alzheimer's Disease (AD), with often salient medial-temporal lobe (MTL) damage, typically show a shift in the character of their autobiographical memories, about which they provide semantic information more easily than they can re-access the contextual details (Piolino et al., 2003; Westmacott et al., 2001), reinforcing the notion that the MTL contributes the episodic flavour to such memories (Eichenbaum, 2006; Moscovitch et al., 2006). Semantic dementia patients, instead, with usually more lateral temporal cortex atrophy (Graham and Hodges, 1997; Lauro-Grotto et al., 1997b), typically show more impairment in their knowledge of common facts, with relatively preserved autobiographic and episodic information.

The metric content analysis was applied to AD patients and to a group of herpes simplex encephalitis (HES) patients by Lauro-Grotto et al. (2007). AD patients as well as HSE patients demonstrated both lower mean mutual information and a lower mean percent correct than their control subjects. Mean values for I and f_{corr} remained almost unvaried from the first to a second testing session (administered 10 months later) for AD patients, whereas HSE patients got significantly worse in the second compared to the first session. As the test includes famous faces from different epochs (the 50s, the 70s and the 90s), while AD patients as expected remembered older faces relatively better than HSE patients who were on average younger, neither group performed differently in this respect from their age-matched controls. In contrast, the metric content λ was significantly higher in AD patients compared to controls at the first testing session, and it tended to increase from the first to the second session, emphasizing the difference between patients and age-matched controls at the second testing session. As Fig. 26c shows, while all the control subjects and 7 of the AD patients are placed in the median area of the leaf diagram, 4 patients are placed even above the upper reference value $\lambda = 1$. In contrast to AD patients, HSE patients showed a marginally lower metric content than their controls, in both testing sessions (Fig. 26d).

The two patient groups were therefore both characterized by a decrease in person-related knowledge compared to their controls, in line with previous evidence. A dissociation was however observed with respect to metric content, indicating that their preferred access mode was quite different. AD patients, usually characterized by a precocious involvement of hippocampal cortices in the neurodegenerative process, showed a marked increase in metric content, indicating a shift to semantic access mode, supporting the notion that episodic access to knowledge, perhaps the default mode, is mediated by MTL structures. The metric content further increased over time, indicating that this measure can track the progress of the disease, which is known to result in progressive hippocampal volume loss and related episodic impairment (Gilboa et al., 2005). Metric content then provides a description of the memory changes associated with AD, as it correlates with measures of the severity of dementia (note that the sample was rather heterogeneous with respect to the severity of the degenerative process, resulting in a scattered distribution of results). It could be that the increase in metric content observed in AD patients may result not only from a shift to semantic access, but

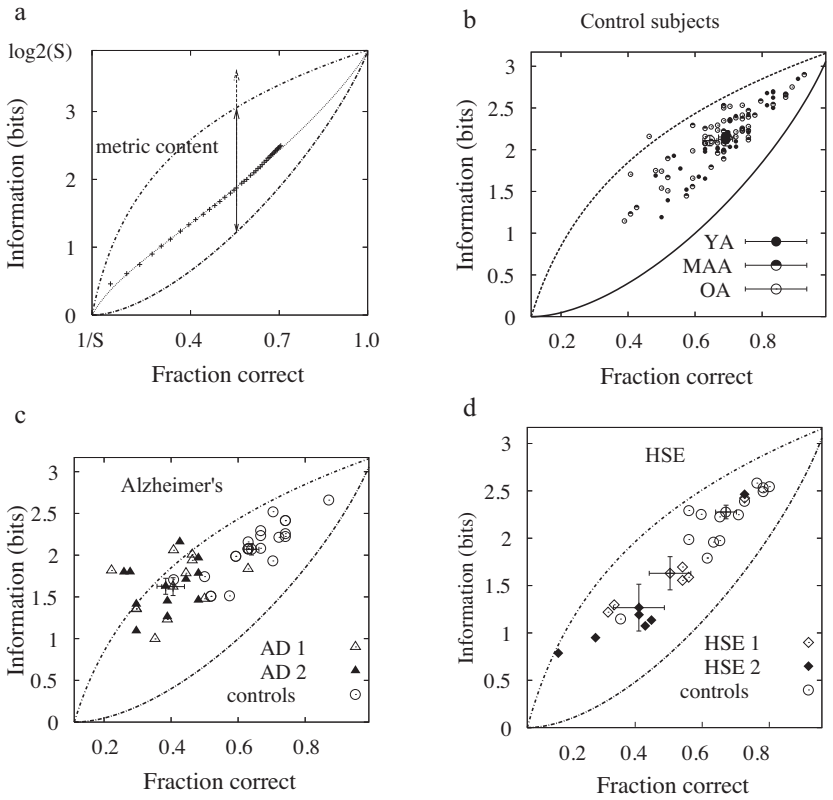

Fig. 26. The responses by human subjects can be conveniently displayed on the leaf diagram (a) indicating metric content. Control subjects (b) respond with considerable scatter, but average values show that older adults (OA, in their 60s) while significantly impaired with respect to young adults (YA, in their 40s) and middle-aged adults (MAA, in their 50s) both in terms of information and percent correct recognition, demonstrate significantly higher metric content in their memory for famous faces. Alzheimer patients (c) are significantly impaired with respect to an age- and education-matched control group, but even higher in metric content; upon retesting 10 months later their metric content increases further (average values for each cohort are those with error bars). Herpes Simplex Encephalitis patients (d, HSE) are significantly more impaired at retesting, and their metric content is slightly below that of matched controls (Lauro-Grotto et al., 2007).

also from a progressive loss of subordinate information in semantic memory representations, which is also typical of the disease.

Herpetic patients, while showing a similarly poor long-term memory performance, decreased somewhat in metric content compared to normal controls. Thus, HSE patients seem to resort to a preferred episodic access mode to person-related knowledge, consistent with the idea that the lateral temporal neocortex, commonly damaged in this condition, contributes to the retrieval of semantic information. This is consistent with evidence that these patients, with typically preserved MTL structures, show a preferential sparing of information about personally known individuals relative to equally famous celebrities of no personal significance (Westmacott et al., 2001). These findings suggest that person-related knowledge after temporal neocortex damage becomes more episodic than semantic in nature, and together with those on AD patients (Lauro-Grotto et al., 2007) and on normal ageing control subjects (Ciaramelli et al., 2006) they provide an information-theoretic description of the role of the medial temporal lobes and of the lateral temporal cortex, respectively, for the episodic and semantic routes to memory retrieval.

3.6. Predictions of decisions or subjective states from fMRI activations and local field potentials

3.6.1. The information from multiple voxels with functional neuroimaging

Analogous questions to those addressed above about neuronal encoding are now being asked with respect to data from functional neuroimaging investigations. These questions include how well it is possible to predict which stimulus has been shown, or which decision will be taken, by measuring the activity in the voxels of activity typically 1 mm^3 or larger which are usually analyzed in humans (Haynes and Rees, 2005a,b, 2006; Pessoa and Padmala, 2005; Lau et al., 2006; Haynes et al., 2007; Hampton and O'Doherty, 2007). Some of the findings are that, for example, when subjects held in mind in a delay period which of two tasks, addition or subtraction, they intended to perform, then it was possible to decode or predict whether addition or subtraction would be performed from a set of medial prefrontal cortex voxels within a radius of 3 voxels with a linear support vector classifier with accuracies in the order of 70%, where chance was 50% (Haynes et al., 2007).

Most of these studies have used the percentage of correct predictions as the measure. However, percentage correct does not allow quantitative and well founded approaches to fundamental issues of the type that can be addressed with information theory. These issues include the amount of information provided by any one voxel in a metric, of mutual information, that can be quantitatively compared with measurements at other levels such as the behavioural level, and the performance of single neurons or populations of neurons; whether each voxel carries independent information or whether there is redundancy; how the information obtained scales with the number of voxels considered; whether combining voxels from different brain areas yields more information than taking the same number of voxels from one brain area; and whether there is significant information about the stimulus or subjective state or prospective rating in the stimulus-dependent cross-correlations between the voxels, i.e. in the higher order statistics. An example of the latter might be that independently of

the mean level of activation of a set of voxels, if some voxels varied together for one event, but not for another, then that could potentially encode information about which event was present.

To bring information theory to bear on these issues, we have adapted our decoding approach illustrated in Fig. 4 to the analysis of neuroimaging and related types of data. In Fig. 4 the rates and correlations for cells are replaced by the activations of selected voxels, and the cross-correlations between the voxels, measured on every trial (Rolls et al., 2009). Then the same information analysis methods used for single neurons can be applied to the activations of voxels in an fMRI study, or to local field potentials, or to other measures of brain activity. The details of the methods are described by Rolls et al. (2009).

We applied this information theoretic approach to investigate how well one can predict the subjective state that will be reported later in a trial from the fMRI BOLD (functional magnetic resonance imaging blood oxygenation-level dependent) activations earlier in the trial to a set of different affective stimuli (Rolls et al., 2009). The subjective pleasantness produced by warm and cold applied to the hand could be predicted on single trials with typically in the range 60–80% correct from the activations of groups of voxels in the orbitofrontal and medial prefrontal cortex and pregenual cingulate cortex, and the information available was **typically in the range 0.1–0.2 bits** (with a maximum of 0.6 bits in the example in Fig. 27). The prediction was typically only a little better with multiple voxels than with one voxel, with the **information increasing very sublinearly with the number of voxels** up to typically 7 voxels. Thus the information from different voxels was not independent, and there was considerable redundancy across voxels. This redundancy was present even when the voxels were from different brain areas. The pairwise stimulus-dependent correlations between voxels, reflecting higher order interactions, did not encode significant information.

For comparison, we showed that the activity of a single neuron in the orbitofrontal cortex can predict with 90% correct and encode 0.5 bits of information about whether an affectively positive or negative visual stimulus has been shown (Rolls et al., 2009), and the information encoded by small numbers of neurons is typically independent.

3.6.2. The information from neurons vs. that from voxels

What is the fundamental difference underlying the different encoding by neurons and by voxels, and the ability to predict from these? The fundamental difference it is proposed is that the neurons, as the information processing computational elements of the brain, each with one output signal, its spike train, use a code to transmit information to other neurons that is rather powerful, in that each neuron, at least up to a limited number of neurons,

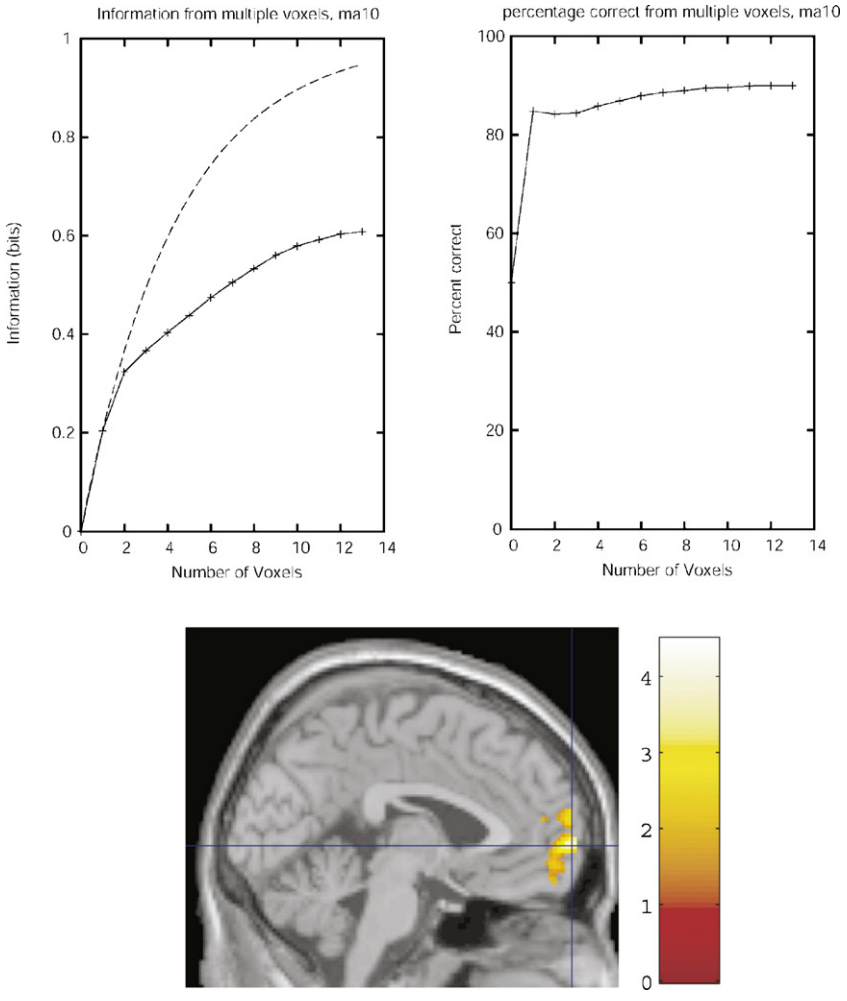

Fig. 27. Information from voxels in functional neuroimaging. (Top) The information available about whether the stimuli were pleasant (>0 on a scale from −2 to +2) or unpleasant (≤0) (left), together with the curve that would be produced if the voxels provided independent information (dashed line), and the percentage correct predictions (right) based on the activations in different numbers of voxels from the medial prefrontal cortex area 10 centred at [−4 66 2]. The prediction was for the ratings that would be made by participant 1. Probability estimation was used for the information analysis shown, and the information based on maximum likelihood decoding produced the same asymptotic value. (Bottom) The medial prefrontal cortex area 10 region from which the voxels centred at [−4 66 2] were obtained.
After Rolls et al. (2009).

carries independent information. This is achieved in part by the fact that the response profile of each neuron to a set of stimuli is relatively uncorrelated with the response profiles of other neurons. So at the neuron level, as this is how the information is transmitted between the computing elements of the brain, there is a great advantage to using an efficient code for the information transmission, and this means that relatively large amounts of information can be decoded from populations of single neurons, and can be used to make good predictions.

However, there is no constraint of this type at all on the activation of one voxel reflecting the activation of hundreds of thousands of neurons, compared to the activation of another voxel, as the average activity of vast numbers of neurons is not how information is transmitted between the computing elements of the brain. (If the neuronal density is taken as say 30,000 neurons/mm^3 (Abeles, 1991; Rolls, 2008), then a $3 \times 3 \times 3$ mm^3 voxel would contain 810,000 neurons.) Instead of the average activation (a single scalar quantity), it is the direction of the vector comprised by the firing of a population of neurons where the activity of each neuron is one element of the vector that transmits the information (Rolls, 2008). It is a vector of this type that each neuron receives, with the length of the vector, set by the number of synapses onto each neurons, typically of the order 10,000 for cortical pyramidal cells. Now of course different voxels in a cortical area will tend to have somewhat different activity, partly as a result of the effect of self-organizing maps in the cortex which tends to place neurons with similar responses close together in the map, and neurons with different responses further apart in the map (Rolls, 2008). So some information will be available about which stimulus was shown by measuring the average activation in different parts of the map. But the reason that this information is small in comparison to that provided by neurons is that the voxel map (reflecting averages of the activity of many hundreds of thousands of neurons) is not the way that information is transmitted between the computing elements of the brain. Instead it is the vector of neuronal activity (where each element of the vector is the firing of a different neuron) within each cortical area that is being used to transmit information round the brain, and in which therefore an efficient code is being used.

We also found that there was no significant information in the stimulus-dependent cross-correlations between voxels. Given the points made in the preceding paragraph, such higher order encoding effects across voxels, where each voxel contains hundreds of thousands of neurons, would not be expected. Even at the neuronal level, under natural visual conditions when attention is being paid and the brain is working normally to segment and discriminate between stimuli embedded in complex natural scenes, almost all the information, typically >95%, is encoded in the firing rates, with very little in stimulus-dependent cross-correlations between inferior temporal cortex neurons (Aggelopoulos et al., 2005; Rolls, 2008).

Because the code provided by the firing rate of single neurons is relatively independent, the code can never be read adequately by any process that averages across many neurons (and synaptic currents (Logothetis, 2008)), such as fMRI, local field potentials (LFP), magnetoencephalography, etc (Magri et al., 2009; Ince et al., 2010a; Quian Quiroga and Panzeri, 2009). Further, because it is a major principle of brain function that information is carried by the spiking of individual neurons each built to carry as independent information as possible from the other neurons, and because brain computation relies on distributed representations for generalization, completion, maintaining a memory, etc (Rolls, 2008), methods that average across many let alone hundreds of thousands of neurons will never reveal how information is actually encoded in the brain, the subject of this paper.

It is this independence of the information transmitted by individual neurons that enables a population of neurons to encode which individual face (Rolls et al., 1997b), which particular object (Booth and Rolls, 1998), which particular spatial view (Rolls et al., 1998), which particular head direction (Robertson et al., 1999) etc has been shown. (If just categorization is the measure, e.g. was it a face, object or spatial scene, then LFPs may reflect this, as they represent local activity, and we know that there is localization on a scale of 0.5–1 mm in what category is represented in IT cortex, with these categories represented in spatially separate neuronal clusters due to cortical self-organizing map principles (Rolls, 2008), which can be detected by LFP recording.)

If the average firing rate of a neuronal population, as it might be reflected in a LFP from a single electrode in thalamic visual processing, varies across time, for example when a movie is shown, then a dynamical network may go into different states, with for example the average firing rate of the input influencing gamma-range oscillations generated by inhibitory-excitatory neural interactions; and slow dynamic features – the time varying structure – of the input reflected in slow LFP fluctuations (Mazzoni et al., 2008). (The LFPs may reflect temporal changes of this type, but information transmission in the brain relies on spikes travelling along the axons of individual neurons to transmit encoded information, not on local field potentials.)

In the primary visual cortex when a movie with changing scenes is shown, the high gamma frequency power (60–100 Hz) from a single electrode was correlated with the multiunit activity (reflecting overall spiking from many neurons) recorded from the same electrode and was influenced by which 2 s period of the movie was being shown (0.22 bits of signal information); and low frequency power in the LFP (<24 Hz) reflected the trial-by-trial variability in the recordings (i.e. the noise correlation, which is not related to the signal in the different movie scenes) and also reflected some information about the stimulus (i.e. which 2 s period of the movie was present) (Belitski et al., 2008, 2010). (This is a small amount of information: less than the average that is conveyed by each single IT face-selective neuron about which particular face in a set of 20 faces was shown (0.36 bits (Rolls et al., 1997c)), or about which of 65 face and non-face stimuli was seen (0.58 bits (Rolls et al., 1997c) Fig. 1b), and much less than the 2.77 bits encoded by a population of 14 such single neurons about which particular face in a set of 20 faces was shown (Rolls et al., 1997b).) Although the neuronal firing and the LFP can vary independently (Gieselmann and Thiele, 2008), there have been few attempts to compare the information obtained from the population code provided by a set of separate single neurons with that in the LFP, though decoding accuracy even into semantic categories (face, animal, place, known to be differentially localized, let alone into the particular face etc) was increased very little by adding LFPs to spike data from single temporal cortex neurons in humans except for time windows so short that individual neurons rarely spiked (Kraskov et al., 2007). (For a 200 ms time window, where chance categorization was 33%, the decoding accuracy was approximately 40% correct from the LFP, 54% from the spikes, and 54% from the spikes and the LFP: see their Fig. 8.)

In summary, the information from single neurons can add independently so that 14 neurons may convey several bits of information, and this independence of the information provided by different neurons is unlikely to be a property of how information from different LFPs adds, and is not a property of how information from voxels in neuroimaging data add (Rolls et al., 2009). It is the spiking of large numbers of different neurons that connect to a receiving neuron by separate synaptic weights that provides the brain's basis for information coding and exchange between neurons (Fig. 18) (Rolls et al., 1997b; Rolls, 2008), not LFPs or the activations of voxels, which reflect an average activity of

hundreds or hundreds of thousands of neurons, not the independent information that neurons encode (Rolls et al., 2009). We note that LFPs can be used to measure the level of coherence between populations, and that both firing rates and coherence complement each other in the underlying neurodynamics (Deco and Rolls, in preparation).

4. Conclusions on cortical neuronal encoding

The conclusions emerging from this set of information theoretic analyses, many in cortical areas towards the end of the ventral visual stream of the monkey, and others in the hippocampus for spatial view cells (Rolls et al., 1998), in the presubiculum for head direction cells (Robertson et al., 1999), in the insular taste cortex for taste neurons, and in the orbitofrontal cortex for olfactory and taste neurons (Rolls et al., 1996, 2010a), are as follows.

The representation of at least some classes of objects in those areas is achieved with minimal redundancy by cells that are allocated each to analyze a different aspect of the visual stimulus (Abbott et al., 1996; Rolls et al., 1997b) (as shown in Sections 3.3 and 3.3.6). This minimal redundancy is what would be expected of a self-organizing system in which different cells acquired their response selectivities through processes that include some randomness in the initial connectivity, and local competition among nearby cells (Rolls, 2008). Towards the end of the ventral visual stream redundancy may thus be effectively minimized, a finding consistent with the general idea that one of the functions of the early visual system is indeed that of progressively minimizing redundancy in the representation of visual stimuli (Attneave, 1954; Barlow, 1961). Indeed, the evidence described in Sections 3.3, 3.3.6 and 2.3 shows that the exponential rise in the number of stimuli that can be decoded when the firing rates of different numbers of neurons are analyzed indicates that the encoding of information using firing rates (in practice the number of spikes emitted by each of a large population of neurons in a short time period) is a very powerful coding scheme used by the cerebral cortex, and that the information carried by different neurons is close to independent provided that the number of stimuli being considered is sufficiently large.

Quantitatively, the encoding of information using firing rates (in practice the number of spikes emitted by each of a large population of neurons in a short time period) is likely to be far more important than temporal encoding, in terms of the number of stimuli that can be encoded. Moreover, the information available from an ensemble of cortical neurons when only the firing rates are read, that is with no temporal encoding within or between neurons, is made available very rapidly (see Fig. 13 and Section 2.3). Further, the neuronal responses in most ventral or 'what' processing streams of behaving monkeys show sustained firing rate differences to different stimuli (see for example Fig. 5 for visual representations, for the olfactory pathways Rolls et al. (1996), for spatial view cells in the hippocampus Rolls et al. (1998), and for head direction cells in the presubiculum Robertson et al. (1999)), so that it may not usually be necessary to invoke temporal encoding for the information about the stimulus. Further, as indicated in Section 3.3.6, information theoretic approaches have enabled the information that is available from the firing rate and from the relative time of firing (synchronization) of inferior temporal cortex neurons to be directly compared with the same metric, and most of the information appears to be encoded in the numbers of spikes emitted by a population of cells in a short time period, rather than by the temporal synchronization of the responses of different neurons when certain stimuli appear (see Section 3.3.6 and Aggelopoulos et al. (2005)).

Information theoretic approaches have also enabled different types of readout or decoding that could be performed by the brain of the information available in the responses of cell populations to be compared (Rolls et al., 1997b; Robertson et al., 1999). It has been shown for example that the multiple cell representation of information used by the brain in the inferior temporal visual cortex (Rolls et al., 1997b; Aggelopoulos et al., 2005), olfactory cortex (Rolls et al., 1996), hippocampus (Rolls et al., 1998), and presubiculum (Robertson et al., 1999) can be read fairly efficiently by the neuronally plausible dot product decoding, and that the representation has all the desirable properties of generalization and graceful degradation, as well as exponential coding capacity (see Sections 3.3 and 3.3.6).

Information theoretic approaches have also enabled the information available about different aspects of stimuli to be directly compared. For example, it has been shown that inferior temporal cortex neurons make explicit much more information about what stimulus has been shown rather than where the stimulus is in the visual field (Tovee et al., 1994), and this is part of the evidence that inferior temporal cortex neurons provide translation invariant representations. In a similar way, information theoretic analysis has provided clear evidence that view invariant representations of objects and faces are present in the inferior temporal visual cortex, in that for example much information is available about what object has been shown from any single trial on which any view of any object is presented (Booth and Rolls, 1998).

Information theory has also helped to elucidate the way in which the inferior temporal visual cortex provides a representation of objects and faces, in which information about which object or face is shown is made explicit in the firing of the neurons in such a way that the information can be read off very simply by memory systems such as the orbitofrontal cortex, amygdala, and perirhinal cortex/hippocampal systems. The information can be read off using dot product decoding, that is by using a synaptically weighted sum of inputs from inferior temporal cortex neurons (Rolls, 2008). Moreover, information theory has helped to show that for many neurons considerable invariance in the representations of objects and faces are shown by inferior temporal cortex neurons (e.g. Booth and Rolls (1998)). Information theory has also helped to show that inferior temporal cortex neurons maintain their object selectivity even when the objects are presented in complex natural backgrounds (Aggelopoulos et al., 2005; Rolls, 2008).

Information theory has also enabled the information available in neuronal representations to be compared with that available to the whole animal in its behaviour (Zohary et al., 1994) (but see Section 3.3.4).

Finally, information theory also provides a metric for directly comparing the information available from neurons in the brain with that available from single neurons and populations of neurons in simulations of visual information processing (Rolls, 2008, Chapter 4).

In summary, the evidence from the application of information theoretic and related approaches to how information is encoded in the visual, hippocampal, and olfactory cortical systems described during behaviour leads to the following working hypotheses:

1. Much information is available about the stimulus presented in the number of spikes emitted by single neurons in a fixed time period, the firing rate.
2. Much of this firing rate information is available in short periods, with a considerable proportion available in as little as 20 ms. This rapid availability of information enables the next stage of processing to read the information quickly, and thus for multistage processing to operate rapidly. This time is the

order of time over which a receiving neuron might be able to utilize the information, given its synaptic and membrane time constants. In this time, a sending neuron is most likely to emit 0, 1, or 2 spikes.

3. This rapid availability of information is confirmed by population analyses, which indicate that across a population on neurons, much information is available in short time periods.
4. More information is available using this rate code in a short period (of, e.g. 20 ms) than from just the first spike.
5. Little information is available by time variations within the spike train of individual neurons for static visual stimuli (in periods of several hundred milliseconds), apart from a small amount of information from the onset latency of the neuronal response. A static stimulus encompasses what might be seen in a single visual fixation, what might be tasted with a stimulus in the mouth, what might be smelled in a single breath, etc. For a time-varying stimulus, clearly the firing rate will vary as a function of time, with the firing rate coding system analyzed here capable of encoding a stimulus shown for as little as 20 ms, and responding to changes in the stimuli on that timescale.
6. Across a population of neurons, the firing rate information provided by each neuron tends to be independent; that is, the information increases approximately linearly with the number of neurons. This applies of course only when there is a large amount of information to be encoded, that is with a large number of stimuli. The outcome is that the number of stimuli that can be encoded rises exponentially in the number of neurons in the ensemble. (For a small stimulus set, the information saturates gradually as the amount of information available from the neuronal population approaches that required to code for the stimulus set.) This applies up to the number of neurons tested and the stimulus set sizes used, but as the number of neurons becomes very large, this is likely to hold less well. An implication of the independence is that the response profiles to a set of stimuli of different neurons are uncorrelated.
7. The information in the firing rate across a population of neurons can be read moderately efficiently by a decoding procedure as simple as a dot product. This is the simplest type of processing that might be performed by a neuron, as it involves taking a dot product of the incoming firing rates with the receiving synaptic weights to obtain the activation (e.g. depolarization) of the neuron. This type of information encoding ensures that the simple emergent properties of associative neuronal networks such as generalization, completion, and graceful degradation (Rolls, 2008) can be realized very naturally and simply.
8. There is little additional information to the great deal available in the firing rates from any stimulus-dependent cross-correlations or synchronization that may be present. Stimulus-dependent synchronization might in any case only be useful for grouping different neuronal populations, and would not easily provide a solution to the binding problem in vision. Instead, the binding problem in vision may be solved by the presence of neurons that respond to combinations of features in a given spatial position with respect to each other.
9. There is little information available in the order of the spike arrival times of different neurons for different stimuli that is separate or additional to that provided by a rate code. The presence of spontaneous activity in cortical neurons facilitates rapid neuronal responses, because some neurons are close to threshold at any given time, but this also would make a spike order code difficult to implement.
10. Analysis of the responses of single neurons to measure the sparseness of the representation indicates that the representation is distributed, and not grandmother cell like (or local). Moreover, the nature of the distributed representation, that it can be read by dot product decoding, allows simple emergent properties of associative neuronal networks such as generalization, completion, and graceful degradation (Rolls, 2008) to be realized very naturally and simply.
11. The representation is not very sparse in the perceptual systems studied (as shown for example by the values of the single cell sparseness a^s), and this may allow much information to be represented. At the same time, the responses of different neurons to a set of stimuli are decorrelated, in the sense that the correlations between the response profiles of different neurons to a set of stimuli are low. Consistent with this, the neurons convey independent information, at least up to reasonable numbers of neurons. The representation may be more sparse in memory systems such as the hippocampus, and this may help to maximize the number of memories that can be stored in associative networks.
12. The nature of the distributed representation can be understood further by the firing rate probability distribution, which has a long tail with low probabilities of high firing rates. The firing rate probability distributions for some neurons fit an exponential distribution, and for others there are too few very low rates for a good fit to the exponential distribution. An implication of an exponential distribution is that this maximizes the entropy of the neuronal responses for a given mean firing rate under some conditions. It is of interest that in the inferior temporal visual cortex, the firing rate probability distribution is very close to exponential if a large number of neurons are included without scaling of the firing rates of each neuron. An implication is that a receiving neuron would see an exponential firing rate probability distribution.
13. The population sparseness a^p, that is the sparseness of the firing of a population of neurons to a given stimulus (or at one time), is the important measure for setting the capacity of associative neuronal networks. In populations of neurons studied in the inferior temporal cortex, hippocampus, and orbitofrontal cortex, it takes the same value as the single cell sparseness a^s, and this is a situation of weak ergodicity that occurs if the response profiles of the different neurons to a set of stimuli are uncorrelated.
14. Although oscillations *per se* do not code information, they can influence the transmission of information between cortical areas if the oscillations are coherent and in phase, and can increase the speed of processing within a cortical area by a mechanism like stochastic resonance (Fries, 2005, 2009; Deco and Rolls, in preparation; Smerieri et al., 2010; Buehlmann and Deco, 2010; Wang, 2010).

Understanding the neuronal code, the subject of this paper, is fundamental for understanding how memory and related perceptual systems in the brain operate, as follows:

Understanding the neuronal code helps to clarify what neuronal operations would be useful in memory and in fact in most mammalian brain systems (e.g. dot product decoding, that is taking a sum in a short time of the incoming firing rates weighted by the synaptic weights).

It clarifies how rapidly memory and perceptual systems in the brain could operate, in terms of how long it takes a receiving neuron to read the code.

It helps to confirm how the properties of those memory systems in terms of generalization, completion, and graceful degradation occur, in that the representation is in the correct form for these properties to be realized (Rolls, 2008).

Understanding the neuronal code also provides evidence essential for understanding the storage capacity of memory

systems, and the representational capacity of perceptual systems (Rolls, 2008).

Understanding the neuronal code is also important for interpreting functional neuroimaging, for it shows that functional imaging that reflects incoming firing rates and thus currents injected into neurons, and probably not stimulus-dependent synchronization, is likely to lead to useful interpretations of the underlying neuronal activity and processing (Rolls et al., 2010b,c). Of course, functional neuroimaging cannot address the details of the representation of information in the brain (Rolls et al., 2009) in the way that is essential for understanding how neuronal networks in the brain could operate, for this level of understanding (in terms of all the properties and working hypotheses described above) comes only from an understanding of how single neurons and populations of neurons encode information.

Finally, we remark that the neuronal encoding scheme used by the brain is one reason for the tractability of the brain (Rolls, 2012), in that the code can be read substantially from the firing rates of individual neurons or relatively small populations of neurons (Section 3 and Rolls, 2008, Appendix C). The deep computational reason for this appears to be that neurons decode the information by dot product decoding (Fig. 18), and the consequence is that each of the independent inputs to a neuron adds information to what can be categorized by the neuron (Rolls, 2008). The brain would have been much less tractable if binary encoding of the type used in a computer was used, as this is a combinatorial code, and any single bit in the computer word, or any subset of bits, yields little evidence on its own about the particular item being represented. Further reasons why the brain is relatively tractable, and why rapid progress in understanding many aspects of its functions is now being made, are described by Rolls (2012).

5. Information theory terms – a short glossary

1. The **amount of information**, or **surprise**, in the occurrence of an event (or symbol) s_i of probability $P(s_i)$ is

$$I(s_i) = \frac{\log_2 1}{P(s_i)} = -\log_2 P(s_i). \tag{35}$$

   (The measure is in bits if logs to the base 2 are used.) This is also the amount of **uncertainty** removed by the occurrence of the event.
2. The average amount of information per source symbol over the whole alphabet (S) of symbols s_i is the **entropy**,

$$H(S) = -\sum_i P(s_i)\log_2 P(s_i) \tag{36}$$

   (or *a priori* entropy).
3. The probability of the pair of symbols s and s' is denoted $P(s, s')$, and is $P(s)P(s')$ only when the two symbols are **independent**.
4. Bayes theorem (given the output s', what was the input s ?) states that

$$P(s|s') = \frac{P(s'|s)P(s)}{P(s')} \tag{37}$$

   where $P(s'|s)$ is the **forward** conditional probability (given the input s, what will be the output s' ?), and $P(s|s')$ is the **backward** (or posterior) conditional probability (given the output s', what was the input s ?). The prior probability is $P(s)$.
5. **Mutual information**. Prior to reception of s', the probability of the input symbol s was $P(s)$. This is the *a priori* probability of s. After reception of s', the probability that the input symbol was s becomes $P(s|s')$, the conditional probability that s was sent given that s' was received. This is the *a posteriori* probability of s. The difference between the *a priori* and *a posteriori* uncertainties measures the gain of information due to the reception of s'. Once averaged across the values of both symbols s and s', this is the **mutual information**, or **transinformation**

$$I(S,S') = \sum_{s,s'} P(s,s')\{\log_2\left[\frac{1}{P(s)}\right] - \log_2\left[\frac{1}{P(s|s')}\right]\} \tag{38}$$

$$= \sum_{s,s'} P(s,s')\log_2\left[\frac{P(s|s')}{P(s)}\right].$$

   Alternatively,

$$I(S,S') = H(S) - H(S|S'). \tag{39}$$

   $H(S|S')$ is sometimes called the **equivocation** (of S with respect to S').

Acknowledgements

ETR is grateful to Larry Abbott, Nicholas Aggelopoulos, Roland Baddeley, Francesco Battaglia, Michael Booth, Hugo Critchley, Gustavo Deco, Leonardo Franco, Pierre Georges-Francois, Fabian Grabenhorst, Miki Kadohisa, Stefano Panzeri, Robert Robertson, Martin Tovee, and Justus Verhagen for contributing to many of the collaborative studies described here. Support from the Medical Research Council, the Wellcome Trust, and the Oxford McDonnell Centre in Cognitive Neuroscience is acknowledged.

AT is grateful to Erika Cerasti, Elisa Ciaramelli, Marianne Fyhn, Torkell Hafting, Karel Jezek, Rosapia Lauro-Grotto, Edvard I. Moser, May-Britt Moser, and Carolina Piccini, for contributing to many of the collaborative studies described here.

References

Abbott, L.F., Rolls, E.T., Tovee, M.J., 1996. Representational capacity of face coding in monkeys. Cerebral Cortex 6, 498–505.
Abbott, L.F., Varela, J.A., Sen, K., Nelson, S.B., 1997. Synaptic depression and cortical gain control. Science 275, 220–224.
Abeles, A., 1991. Corticonics. Cambridge University Press, New York.
Aertsen, A.M.H.J., Gerstein, G.L., Habib, M.K., Palm, G., 1989. Dynamics of neuronal firing correlation: modulation of 'effective connectivity'. Journal of Neurophysiology 61, 900–917.
Aggelopoulos, N.C., Rolls, E.T., 2005. Natural scene perception: inferior temporal cortex neurons encode the positions of different objects in the scene. European Journal of Neuroscience 22, 2903–2916.
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.
Akam, T., Kullmann, D.M., 2010. Oscillations and filtering networks support flexible routing of information. Neuron 67, 308–320.
Akam, T.E., Kullmann, D.M. Selective communication through oscillatory coherence: can it work?, in press.
Attneave, F., 1954. Informational aspects of visual perception. Psychological Review 61, 183–193.
Bacon-Mace, N., Mace, M.J., Fabre-Thorpe, M., Thorpe, S.J., 2005. The time course of visual processing: backward masking and natural scene categorisation. Vision Research 45, 1459–1469.
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. Proceedings of the Royal Society B 264, 1775–1783.
Barlow, H., 1972. Single units and sensation: a neuron doctrine for perceptual psychology? Perception 1, 371–394.
Barlow, H., 1995. The neuron doctrine in perception. In: Gazzaniga, M.S. (Ed.), The Cognitive Neurosciences. MIT Press, Cambridge, MA, (Chapter 26), pp. 415–435.
Barlow, H.B., 1961. Possible principles underlying the transformation of sensory messages. In: Rosenblith, W. (Ed.), Sensory Communication. MIT Press, Cambridge, MA.
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 temporal lobe neocortex. Journal of Neuroscience 7, 330–342.
Belitski, A., Gretton, A., Magri, C., Murayama, Y., Montemurro, M.A., Logothetis, N.K., Panzeri, S., 2008. Low-frequency local field potentials and spikes in primary visual cortex convey independent visual information. Journal of Neuroscience 28, 5696–5709.

Belitski, A., Panzeri, S., Magri, C., Logothetis, N.K., Kayser, C., 2010. Sensory information in local field potentials and spikes from visual and auditory cortices: time scales and frequency bands. Journal of Computational Neuroscience 29, 533–545.
Bezzi, M., Diamond, M.E., Treves, A., 2002. Redundancy and synergy arising from pairwise correlations in neuronal ensembles. Journal of Computational Neuroscience 12, 165–174.
Bialek, W., Rieke, F., de Ruyter van Steveninck, R.R., Warland, D., 1991. Reading a neural code. Science 252, 1854–1857.
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.
Brunel, N., Wang, X.J., 2003. What determines the frequency of fast network oscillations with irregular neural discharges?, I. Synaptic dynamics and excitation-inhibition balance. Journal of Neurophysiology 90, 415–430.
Buehlmann, A., Deco, G., 2008. The neuronal basis of attention: rate versus synchronization modulation. Journal of Neuroscience 28, 7679–7686.
Buehlmann, A., Deco, G., 2010. Optimal information transfer in the cortex through synchronization. PLoS Computational Biology 6, e1000934.
Cerasti, E., Treves, A., 2010. How informative are spatial CA3 representations established by the dentate gyrus? PLoS Computational Biology 6, e1000759.
Ciaramelli, E., Lauro-Grotto, R., Treves, A., 2006. Dissociating episodic from semantic access mode by mutual information measures: evidence from aging and Alzheimer's disease. Journal de Physiologie Paris 100, 142–153.
Cover, T.M., Thomas, J.A., 1991. Elements of Information Theory. Wiley, New York.
deCharms, R.C., Merzenich, M.M., 1996. Primary cortical representation of sounds by the coordination of action-potential timing. Nature 381, 610–613.
Deco, G., Rolls, E.T., 2006. A neurophysiological model of decision-making and Weber's law. European Journal of Neuroscience 24, 901–916.
Deco, G., Rolls, E.T. Reconciling oscillations and firing rates, in preparation.
Delorme, A., Thorpe, S.J., 2001. Face identification using one spike per neuron: resistance to image degradations. Neural Networks 14, 795–803.
Desimone, R., Albright, T.D., Gross, C.G., Bruce, C., 1984. Stimulus-selctive responses of inferior temporal neurons in the macaque. Journal of Neuroscience 4, 2051–2062.
DeWeese, M.R., Meister, M., 1999. How to measure the information gained from one symbol. Network 10, 325–340.
Eichenbaum, H., 2006. Remembering: functional organization of the declarative memory system. Current Biology 16, R643–R645.
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.
Eskandar, E.N., Richmond, B.J., Optican, L.M., 1992. Role of inferior temporal neurons in visual memory. I. Temporal encoding of information about visual images, recalled images, and behavioural context. Journal of Neurophysiology 68, 1277–1295.
Field, D.J., 1994. What is the goal of sensory coding? Neural Computation 6, 559–601.
Földiák, P., 2003. Sparse coding in the primate cortex. In: Arbib, M.A. (Ed.), Handbook of Brain Theory and Neural Networks. 2nd edn. MIT Press, Cambridge, MA, pp. 1064–1608.
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.
Franco, L., Rolls, E.T., Aggelopoulos, N.C., Jerez, J.M., 2007. Neuronal selectivity, population sparseness, and ergodicity in the inferior temporal visual cortex. Biological Cybernetics 96, 547–560.
Fries, P., 2005. A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. Trends in Cognitive Sciences 9, 474–480.
Fries, P., 2009. Neuronal gamma-band synchronization as a fundamental process in cortical computation. Annual Reviews of Neuroscience 32, 209–224.
Fyhn, M., Hafting, T., Treves, A., Moser, M.B., Moser, E.I., 2007. Hippocampal remapping and grid realignment in entorhinal cortex. Nature 446, 190–194.
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.
Gieselmann, M.A., Thiele, A., 2008. Comparison of spatial integration and surround suppression characteristics in spiking activity and the local field potential in macaque V1. European Journal of Neuroscience 28, 447–459.
Gilboa, A., Ramirez, J., Kohler, S., Westmacott, R., Black, S.E., Moscovitch, M., 2005. Retrieval of autobiographical memory in Alzheimer's disease: relation to volumes of medial temporal lobe and other structures. Hippocampus 15, 535–550.
Gochin, P.M., Colombo, M., Dorfman, G.A., Gerstein, G.L., Gross, C.G., 1994. Neural ensemble encoding in inferior temporal cortex. Journal of Neurophysiology 71, 2325–2337.
Golledge, H.D., Panzeri, S., Zheng, F., Pola, G., Scannell, J.W., Giannikopoulos, D.V., Mason, R.J., Tovee, M.J., Young, M.P., 2003. Correlations, feature-binding and population coding in primary visual cortex. Neuroreport 14, 1045–1050.
Golomb, D., Kleinfeld, D., Reid, R.C., Shapley, R.M., Shraiman, B., 1994. On temporal codes and the spatiotemporal response of neurons in the lateral geniculate nucleus. Journal of Neurophysiology 72, 2990–3003.
Golomb, D., Hertz, J.A., Panzeri, S., Treves, A., Richmond, B.J., 1997. How well can we estimate the information carried in neuronal responses from limited samples? Neural Computation 9, 649–665.
Graham, K.S., Hodges, J.R., 1997. Differentiating the roles of the hippocampal complex and the neocortex in long-term memory storage: evidence from the study of semantic dementia and Alzheimer's disease. Neuropsychology 11, 77–89.
Hamming, R.W., 1990. Coding and Information Theory, 2nd edn. Prentice-Hall, Englewood Cliffs, NJ.
Hampton, A.N., O'Doherty, J.P., 2007. Decoding the neural substrates of reward-related decision making with functional MRI. Proceedings of the National Academy of Sciences USA 104, 1377–1382.
Haynes, J.D., Rees, G., 2005a. Predicting the orientation of invisible stimuli from activity in human primary visual cortex. Nature Neuroscience 8, 686–691.
Haynes, J.D., Rees, G., 2005b. Predicting the stream of consciousness from activity in human visual cortex. Current Biology 15, 1301–1307.
Haynes, J.D., Rees, G., 2006. Decoding mental states from brain activity in humans. Nature Reviews Neuroscience 7, 523–534.
Haynes, J.D., Sakai, K., Rees, G., Gilbert, S., Frith, C., Passingham, R.E., 2007. Reading hidden intentions in the human brain. Current Biology 17, 323–328.
Heller, J., Hertz, J.A., Kjaer, T.W., Richmond, B.J., 1995. Information flow and temporal coding in primate pattern vision. Journal of Comparative Neuroscience 2, 175–193.
Hertz, J.A., Kjaer, T.W., Eskander, E.N., Richmond, B.J., 1992. Measuring natural neural processing with artificial neural networks. International Journal of Neural Systems 3 (Suppl.), 91–103.
Hopfield, J.J., 1982. Neural networks and physical systems with emergent collective computational abilities. Proceedings of the National Academy of Sciences USA 79, 2554–2558.
Hummel, J.E., Biederman, I., 1992. Dynamic binding in a neural network for shape recognition. Psychological Review 99, 480–517.
Huxter, J., Burgess, N., O'Keefe, J., 2003. Independent rate and temporal coding in hippocampal pyramidal cells. Nature 425, 828–832.
Huxter, J.R., Senior, T.J., Allen, K., Csicsvari, J., 2008. Theta phase-specific codes for two-dimensional position, trajectory and heading in the hippocampus. Nature Neuroscience 11, 587–594.
Ince, R.A., Mazzoni, A., Petersen, R.S., Panzeri, S., 2010a. Open source tools for the information theoretic analysis of neural data. Frontiers in Neuroscience 4, 62–70.
Ince, R.A., Senatore, R., Arabzadeh, E., Montani, F., Diamond, M.E., Panzeri, S., 2010b. Information-theoretic methods for studying population codes. Neural Networks 23, 713–727.
Jensen, O., Lisman, J.E., 2000. Position reconstruction from an ensemble of hippocampal place cells: contribution of theta phase coding. Journal of Neurophysiology 83, 2602–2609.
Jezek, K., Henriksen, E.J, Treves, A., Moser, E.I., Moser, M.B., 2011. Theta-paced flickering between place-cell maps in the hippocampus. Nature, 478,246–249.
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. Neuronal representations of stimuli in the mouth: the primate insular taste cortex, orbitofrontal cortex, and amygdala. Chemical Senses 30, 401–419.
Kadohisa, M., Rolls, E.T., Verhagen, J.V., 2005b. The primate amygdala: neuronal representations of the viscosity, fat texture, grittiness and taste of foods. Neuroscience 132, 33–48.
Kayser, C., Montemurro, M.A., Logothetis, N.K., Panzeri, S., 2009. Spike-phase coding boosts and stabilizes information carried by spatial and temporal spike patterns. Neuron 61, 597–608.
Keysers, C., Perrett, D.I., 2002. Visual masking and RSVP reveal neural competition. Trends in Cognitive Sciences 6, 120–125.
Kiani, R., Esteky, H., Mirpour, K., Tanaka, K., 2007. Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. Journal of Neurophysiology 97, 4296–4309.
Kjaer, T.W., Hertz, J.A., Richmond, B.J., 1994. Decoding cortical neuronal signals: networks models, information estimation and spatial tuning. Journal of Computational Neuroscience 1, 109–139.
Koepsell, K., Wang, X., Hirsch, J.A., Sommer, F.T., 2010. Exploring the function of neural oscillations in early sensory systems. Frontiers in Neuroscience 4, 53.
Kraskov, A., Quiroga, R.Q., Reddy, L., Fried, I., Koch, C., 2007. Local field potentials and spikes in the human medial temporal lobe are selective to image category. Journal of Cognitive Neuroscience 19, 479–492.
Lau, H.C., Rogers, R.D., Passingham, R.E., 2006. On measuring the perceived onsets of spontaneous actions. Journal of Neuroscience 26, 7265–7271.
Lauro-Grotto, R., Borgo, F., Piccini, C., Treves, A., 1997a. What remains of memories lost in Alzheimer's Disease and Herpes Simplex encephalitis. Society for Neuroscience Abstracts 23, 734.
Lauro-Grotto, R., Piccini, C., Shallice, T., 1997b. Modality-specific operations in semantic dementia. Cortex 33, 593–622.
Lauro-Grotto, R., Ciaramelli, E., Piccini, C., Treves, A., 2007. Differential impact of brain damage on the access mode to memory representations: an information theoretic approach. European Journal of Neuroscience 26, 2702–2712.
Lee, H., Simpson, G.V., Logothetis, N.K., Rainer, G., 2005. Phase locking of single neuron activity to theta oscillations during working memory in monkey extrastriate visual cortex. Neuron 45, 147–156.
Lehky, S.R., Sejnowski, T.J., Desimone, R., 2005. Selectivity and sparseness in the responses of striate complex cells. Vision Research 45, 57–73.
Leutgeb, S., Leutgeb, J.K., Treves, A., Moser, M.B., Moser, E.I., 2004. Distinct ensemble codes in hippocampal areas CA3 and CA1. Science 305, 1295–1298.
Levy, W.B., Baxter, R.A., 1996. Energy efficient neural codes. Neural Computation 8, 531–543.

Logothetis, N.K., 2008. What we can do and what we cannot do with fMRI. Nature 453, 869–878.
Loh, M., Rolls, E.T., Deco, G., 2007. A dynamical systems hypothesis of schizophrenia. PLoS Computational Biology 3, e228 doi:10.1371/journal.pcbi.0030228.
Magri, C., Whittingstall, K., Singh, V., Logothetis, N.K., Panzeri, S., 2009. A toolbox for the fast information analysis of multiple-site LFP, EEG and spike train recordings. BMC Neuroscience 10, 81.
Malsburg, C.V.D., 1990. A neural architecture for the representation of scenes. In: McGaugh, J.L., Weinberger, N.M., Lynch, G. (Eds.), Brain Organization and Memory: Cells, Systems and Circuits. Oxford University Press, New York, (Chapter 18), pp. 356–372.
Markram, H., Tsodyks, M., 1996. Redistribution of synaptic efficacy between neocortical pyramidal neurons. Nature 382, 807–810.
Masuda, N., Aihara, K., 2003. Ergodicity of spike trains: when does trial averaging make sense? Neural Computation 15, 1341–1372.
Mazzoni, A., Panzeri, S., Logothetis, N.K., Brunel, N., 2008. Encoding of naturalistic stimuli by local field potential spectra in networks of excitatory and inhibitory neurons. PLoS Computational Biology 4, e1000239.
Miller, G.A., 1955. Note on the bias of information estimates. Information Theory in Psychology; Problems and Methods II-B 95–100.
Montemurro, M.A., Rasch, M.J., Murayama, Y., Logothetis, N.K., Panzeri, S., 2008. Phase-of-firing coding of natural visual stimuli in primary visual cortex. Current Biology 18, 375–380.
Moscovitch, M., Nadel, L., Winocur, G., Gilboa, A., Rosenbaum, R.S., 2006. The cognitive neuroscience of remote episodic, semantic and spatial memory. Current Opinion in Neurobiology 16, 179–190.
Nadasdy, Z., 2010. Binding by asynchrony: the neuronal phase code. Frontiers in Neuroscience.
Nelken, I., Prut, Y., Vaadia, E., Abeles, M., 1994. Population responses to multifrequency sounds in the cat auditory cortex: one- and two-parameter families of sounds. Hearing Research 72, 206–222.
Nowak, L., Bullier, J., 1997. The timing of information transfer in the visual system. In: Rockland, K., Kaas, J., Peters, A. (Eds.), Cerebral Cortex: Extrastriate Cortex in Primate. Plenum, New York, p. 870.
O'Keefe, J., Burgess, N., 2005. Dual phase and rate coding in hippocampal place cells: theoretical significance and relationship to entorhinal grid cells. Hippocampus 15, 853–866.
O'Keefe, J., Dostrovsky, J., 1971. The hippocampus as a spatial map: preliminary evidence from unit activity in the freely moving rat. Brain Research 34, 171–175.
Olshausen, B.A., Field, D.J., 1997. Sparse coding with an incomplete basis set: a strategy employed by V1. Vision Research 37, 3311–3325.
Olshausen, B.A., Field, D.J., 2004. Sparse coding of sensory inputs. Current Opinion in Neurobiology 14, 481–487.
Optican, L.M., Richmond, B.J., 1987. Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex: III. Information theoretic analysis. Journal of Neurophysiology 57, 162–178.
Optican, L.M., Gawne, T.J., Richmond, B.J., Joseph, P.J., 1991. Unbiased measures of transmitted information and channel capacity from multivariate neuronal data. Biological Cybernetics 65, 305–310.
Oram, M.W., Foldiak, P., Perrett, D.I., Sengpiel, F., 1998. The 'ideal homunculus': decoding neural population signals. Trends in Neuroscience 21, 259–265.
Palanca, B.J., DeAngelis, G.C., 2005. Does neuronal synchrony underlie visual feature grouping? Neuron 46, 333–346.
Panzeri, S., Treves, A., 1996. Analytical estimates of limited sampling biases in different information measures. Network 7, 87–107.
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., Schultz, S.R., Treves, A., Rolls, E.T., 1999a. Correlations and the encoding of information in the nervous system. Proceedings of the Royal Society 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 epochs. Neural Computation 11, 1553–1577.
Panzeri, S., Petersen, R.S., Schultz, S.R., Lebedev, M., Diamond, M.E., 2001a. The role of spike timing in the coding of stimulus location in rat somatosensory cortex. Neuron 29, 769–777.
Panzeri, S., Rolls, E.T., Battaglia, F., Lavis, R., 2001b. Speed of feedforward and recurrent processing in multilayer networks of integrate-and-fire neurons. Network: Computation in Neural Systems 12, 423–440.
Panzeri, S., Brunel, N., Logothetis, N.K., Kayser, C., 2010. Sensory neural codes using multiplexed temporal scales. Trends in Neuroscience 33, 111–120.
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.
Pessoa, L., Padmala, S., 2005. Quantitative prediction of perceptual decisions during near-threshold fear detection. Proceedings of the National Academy of Sciences USA 102, 5612–5617.
Piolino, P., Desgranges, B., Belliard, S., Matuszewski, V., Lalevee, C., De la Sayette, V., Eustache, F., 2003. Autobiographical memory and autonoetic consciousness: triple dissociation in neurodegenerative diseases. Brain 126, 2203–2219.
Quian Quiroga, R., Panzeri, S., 2009. Extracting information from neuronal populations: information theory and decoding approaches. Nature Reviews Neuroscience 10, 173–185.
Richmond, B.J., 2009. Stochasticity, spikes and decoding: sufficiency and utility of order statistics. Biological Cybernetics 100, 447–457.
Richmond, B.J., Gawne, T.J., Jin, G.X., 1997. Neuronal codes: reading them and learning how their structure influences network organization. Biosystems 40, 149–157.
Rieke, F., Warland, D., Bialek, W., 1993. Coding efficiency and information rates in sensory neurons. Europhysics Letters 22, 151–156.
Rieke, F., Warland, D., de Ruyter van Steveninck, R.R., Bialek, W., 1997. Spikes: Exploring the Neural Code. MIT Press, Cambridge, MA.
Robertson, R.G., Rolls, E.T., Georges-François, P., Panzeri, S., 1999. Head direction cells in the primate pre-subiculum. Hippocampus 9, 206–219.
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, CA, (Chapter 13), pp. 240–265.
Rolls, E.T., 2000. Functions of the primate temporal lobe cortical visual areas in invariant visual object and face recognition. Neuron 27, 205–218.
Rolls, E.T., 2003. Consciousness absent and present: a neurophysiological exploration. Progress in Brain Research 144, 95–106.
Rolls, E.T., 2006. Consciousness absent and present: a neurophysiological exploration of masking. In: Ogmen, H., Breitmeyer, B.G. (Eds.), The First Half Second. MIT Press, Cambridge, MA, (Chapter 6), pp. 89–108.
Rolls, E.T., 2007. The representation of information about faces in the temporal and frontal lobes of primates including humans. Neuropsychologia 45, 124–143.
Rolls, E.T., 2008. Memory, Attention, and Decision-Making, A Unifying Computational Neuroscience Approach. Oxford University Press, Oxford.
Rolls, E.T., 2009. Functional neuroimaging of umami taste: what makes umami pleasant. American Journal of Clinical Nutrition 90, 803S–814S.
Rolls, E.T., 2010. A computational theory of episodic memory formation in the hippocampus. Behavioural Brain Research 215, 180–196.
Rolls, E.T., 2011a. Consciousness, decision-making, and neural computation. In: Cutsuridis, V., Hussain, A., Taylor, J.G. (Eds.), Perception-Action Cycle: Models, architecture, and hardware. Springer, Berlin, (Chapter 9), pp. 287–333.
Rolls, E.T., 2011b. Face neurons. In: Calder, A.J., Rhodes, G., Johnson, M.H., Haxby, J.V. (Eds.), The Oxford Handbook of Face Perception. Oxford University Press, Oxford, (Chapter 4), pp. 51–75.
Rolls, E.T. (2011c). Glutamate, obsessive–compulsive disorder, schizophrenia, and the stability of cortical attractor neuronal networks. Pharmacology, Biochemistry and Behavior. Epub ahead of print, 23 June.
Rolls, E.T., 2012. Neuroculture. Oxford University Press, Oxford.
Rolls, E.T., Deco, G., 2002. Computational Neuroscience of Vision. Oxford University Press, Oxford.
Rolls, E.T., Deco, G., 2010. The Noisy Brain: Stochastic Dynamics as a Principle of Brain Function. Oxford University Press, Oxford.
Rolls, E.T., Deco, G., 2011. A computational neuroscience approach to schizophrenia and its onset. Neuroscience and Biobehavioral Reviews 35, 1644–1653.
Rolls, E.T., Stringer, S.M., 2006. Invariant visual object recognition: a model, with lighting invariance. Journal of Physiology – Paris 100, 43–62.
Rolls, E.T., Tovee, M.J., 1994. Processing speed in the cerebral cortex and the neurophysiology of visual masking. Proceedings of the Royal Society B 257, 9–15.
Rolls, E.T., Tovee, M.J., 1995. Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. Journal of Neurophysiology 73, 713– 726.
Rolls, E.T., 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 University Press, Oxford.
Rolls, E.T., Webb, T.J., 2011. Cortical attractor network dynamics with diluted connectivity. Brain Research, Epub 7 August.
Rolls, E.T., Xiang, J.-Z., 2005. Reward-spatial view representations and learning in the primate hippocampus. Journal of Neuroscience 25, 6167–6174.
Rolls, E.T., Xiang, J.-Z., 2006. Spatial view cells in the primate hippocampus, and memory recall. Reviews in the Neurosciences 17, 175–200.
Rolls, E.T., Yaxley, S., Sienkiewicz, Z.J., 1990. Gustatory responses of single neurons in the orbitofrontal cortex of the macaque monkey. Journal of Neurophysiology 64, 1055–1066.
Rolls, E.T., Tovee, M.J., Purcell, D.G., Stewart, A.L., Azzopardi, P., 1994. The responses of neurons in the temporal cortex of primates, and face identification and detection. Experimental Brain Research 101, 474–484.
Rolls, E.T., 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., Robertson, R.G., Georges-François, P., 1997a. Spatial view cells in the primate hippocampus. European Journal of Neuroscience 9, 1789–1794.
Rolls, E.T., Treves, A., Tovee, M.J., 1997b. The representational capacity of the distributed encoding of information provided by populations of neurons in the primate temporal visual cortex. Experimental Brain Research 114, 149–162.
Rolls, E.T., Treves, A., Tovee, M., Panzeri, S., 1997c. Information in the neuronal representation of individual stimuli in the primate temporal visual cortex. Journal of Computational Neuroscience 4, 309–333.
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., Tovee, M.J., Panzeri, S., 1999. The neurophysiology of backward visual masking: information analysis. Journal of Cognitive Neuroscience 11, 335–346.
Rolls, E.T., Aggelopoulos, N.C., Zheng, F., 2003a. The receptive fields of inferior temporal cortex neurons in natural scenes. Journal of Neuroscience 23, 339–348.

Rolls, E.T., Franco, L., Aggelopoulos, N.C., Reece, S., 2003b. An information theoretic approach to the contributions of the firing rates and the correlations between the firing of neurons. Journal of Neurophysiology 89, 2810–2822.
Rolls, E.T., Verhagen, J.V., Kadohisa, M., 2003c. Representations of the texture of food in the primate orbitofrontal cortex: neurons responding to viscosity, grittiness, and capsaicin. Journal of Neurophysiology 90, 3711–3724.
Rolls, E.T., Aggelopoulos, N.C., Franco, L., Treves, A., 2004. Information encoding in the inferior temporal visual cortex: contributions of the firing rates and the correlations between the firing of neurons. Biological Cybernetics 90, 19–32.
Rolls, E.T., Xiang, J.-Z., Franco, L., 2005. Object, space and object-space representations in the primate hippocampus. Journal of Neurophysiology 94, 833–844.
Rolls, E.T., Franco, L., Aggelopoulos, N.C., Jerez, J.M., 2006. Information in the first spike, the order of spikes, and the number of spikes provided by neurons in the inferior temporal visual cortex. Vision Research 46, 4193–4205.
Rolls, E.T., Loh, M., Deco, G., 2008a. An attractor hypothesis of obsessive-compulsive disorder. European Journal of Neuroscience 28, 782–793.
Rolls, E.T., Loh, M., Deco, G., Winterer, G., 2008b. Computational models of schizophrenia and dopamine modulation in the prefrontal cortex. Nature Reviews Neuroscience 9, 696–709.
Rolls, E.T., Tromans, J.M., Stringer, S.M., 2008c. Spatial scene representations formed by self-organizing learning in a hippocampal extension of the ventral visual system. European Journal of Neuroscience 28, 2116–2127.
Rolls, E.T., Grabenhorst, F., Franco, L., 2009. Prediction of subjective affective state from brain activations. Journal of Neurophysiology 101, 1294–1308.
Rolls, E.T., Critchley, H., Verhagen, J.V., Kadohisa, M., 2010a. The representation of information about taste and odor in the primate orbitofrontal cortex. Chemosensory Perception 3, 16–33.
Rolls, E.T., Grabenhorst, F., Deco, G., 2010b. Choice, difficulty, and confidence in the brain. Neuroimage 53, 694–706.
Rolls, E.T., Grabenhorst, F., Deco, G., 2010c. Decision-making, errors, and confidence in the brain. Journal of Neurophysiology 104, 2359–2374.
Samengo, I., Treves, A., 2000. Representational capacity of a set of independent neurons. Physical Review E 63, 011910.
Shadlen, M., Newsome, W., 1995. Is there a signal in the noise? Current Opinion in Neurobiology 5, 248–250.
Shadlen, M., Newsome, W., 1998. The variable discharge of cortical neurons: implications for connectivity, computation and coding. Journal of Neuroscience 18, 3870–3896.
Shannon, C.E., 1948. A mathematical theory of communication. AT&T Bell Laboratories Technical Journal 27, 379–423.
Siegel, M., Warden, M.R., Miller, E.K., 2009. Phase-dependent neuronal coding of objects in short-term memory. Proceedings of the National Academy of Sciences USA 106, 21341–21346.
Singer, W., 1999. Neuronal synchrony: a versatile code for the definition of relations? Neuron 24, 49–65.
Singer, W., 2000. Response synchronisation: a universal coding strategy for the definition of relations. In: Gazzaniga, M. (Ed.), The New Cognitive Neurosciences. 2nd edn. MIT Press, Cambridge, MA, (Chapter 23), pp. 325–338.
Singer, W., Gray, C.M., 1995. Visual feature integration and the temporal correlation hypothesis. Annual Review of Neuroscience 18, 555–586.
Singer, W., Gray, C., Engel, A., Konig, P., Artola, A., Brocher, S., 1990. Formation of cortical cell assemblies. Cold Spring Harbor Symposium on Quantitative Biology 55, 939–952.
Skaggs, W.E., McNaughton, B.L., 1992. Quantification of what it is that hippocampal cell firing encodes. Society for Neuroscience Abstracts 18, 1216.
Skaggs, W.E., McNaughton, B.L., Gothard, K., Markus, E., 1993. An information theoretic approach to deciphering the hippocampal code. In: Hansen, S., Cowan, J., Giles, C. (Eds.), Advances in Neural Information Processing Systems, vol. 5. MIT Press, Cambridge, MA, pp. 1030–1037.
Smerieri, A., Rolls, E.T., Feng, J., 2010. Decision time, slow inhibition, and theta rhythm. Journal of Neuroscience 30, 14173–14181.
Tanaka, K., 1996. Inferotemporal cortex and object vision. Annual Review of Neuroscience 19, 109–139.
Thiele, A., Stoner, G., 2003. Neuronal synchrony does not correlate with motion coherence in cortical area MT. Nature 421, 366–370.
Thorpe, S.J., Delorme, A., Van Rullen, R., 2001. Spike-based strategies for rapid processing. Neural Networks 14, 715–725.
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., 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.
Tovee, M.J., Rolls, E.T., Azzopardi, P., 1994. Translation invariance and the responses of neurons in the temporal visual cortical areas of primates. Journal of Neurophysiology 72, 1049–1060.
Treves, A., 1990. Graded-response neurons and information encodings in autoassociative memories. Physical Review A 42, 2418–2430.
Treves, A., 1997. On the perceptual structure of face space. Biosystems 40, 189–196.
Treves, A., Panzeri, S., 1995. The upward bias in measures of information derived from limited data samples. Neural Computation 7, 399–407.
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., 1992. Computational constraints suggest the need for two distinct input systems to the hippocampal CA3 network. Hippocampus 2, 189–199.
Treves, A., Georges-François, P., Panzeri, S., Robertson, R.G., Rolls, E.T., 1999a. The metric content of spatial views as represented in the primate hippocampus. In: Torre, V., Nicholls, J. (Eds.), Neural Circuits and Neural Networks. Springer, Berlin, pp. 239–247.
Treves, A., Panzeri, S., Rolls, E.T., Booth, M., Wakeman, E.A., 1999b. Firing rate distributions and efficiency of information transmission of inferior temporal cortex neurons to natural visual stimuli. Neural Computation 11, 601–631.
Tulving, E., 1985. Memory and consciousness. Canadian Psychology 26, 1–12.
VanRullen, R., Guyonneau, R., Thorpe, S.J., 2005. Spike times make sense. Trends in Neuroscience 28, 1–4.
Verhagen, J.V., Rolls, E.T., Kadohisa, M., 2003. Neurons in the primate orbitofrontal cortex respond to fat texture independently of viscosity. Journal of Neurophysiology 90, 1514–1525.
Verhagen, J.V., Kadohisa, M., Rolls, E.T., 2004. The primate insular taste cortex: neuronal representations of the viscosity, fat texture, grittiness, and the taste of foods in the mouth. Journal of Neurophysiology 92, 1685–1699.
Victor, J.D., 2000. How the brain uses time to represent and process visual information. Brain Research 886, 33–46.
Vinck, M., Lima, B., Womelsdorf, T., Oostenveld, R., Singer, W., Neuenschwander, S., Fries, P., 2010. Gamma-phase shifting in awake monkey visual cortex. Journal of Neuroscience 30, 1250–1257.
Vinje, W.E., Gallant, J.L., 2000. Sparse coding and decorrelation in primary visual cortex during natural vision. Science 287, 1273–1276.
Wang, X.-J., 2008. Decision making in recurrent neuronal circuits. Neuron 60, 215–234.
Wang, X.J., 2010. Neurophysiological and computational principles of cortical rhythms in cognition. Physiological Reviews 90, 1195–1268.
Webb, T.J., Rolls, E.T, Deco, G., Feng, J., 2011. Noise in attractor networks in the brain produced by graded firing rate representations. PLoS One 6, e23630.
Westmacott, R., Leach, L., Freedman, M., Moscovitch, M., 2001. Different patterns of autobiographical memory loss in semantic dementia and medial temporal lobe amnesia: a challenge to consolidation theory. Neurocase 7, 37–55.
Womelsdorf, T., Schoffelen, J.M., Oostenveld, R., Singer, W., Desimone, R., Engel, A.K., Fries, P., 2007. Modulation of neuronal interactions through neuronal synchronization. Science 316, 1609–1612.
Zohary, E., Shadlen, M.N., Newsome, W.T., 1994. Correlated neuronal discharge rate and its implications for psychophysical performance. Nature 370, 140–143.