Prediction of subjective affective state from brain activations. — **Source link**

Edmund T. Rolls, Fabian Grabenhorst, Leonardo Franco

Institutions: University of Oxford

Related papers:

- Cognitive Modulation of Olfactory Processing

- How Cognition Modulates Affective Responses to Taste and Flavor: Top-down Influences on the Orbitofrontal and Pregenual Cingulate Cortices

- Activation of the Human Orbitofrontal Cortex to a Liquid Food Stimulus is Correlated with its Subjective Pleasantness

- Memory, Attention, and Decision-Making: A unifying computational neuroscience approach

- How Pleasant and Unpleasant Stimuli Combine in Different Brain Regions: Odor Mixtures

Prediction of Subjective Affective State From Brain Activations

Edmund T. Rolls,[1] Fabian Grabenhorst,[2] and Leonardo Franco[3]

[1]*Oxford Centre for Computational Neuroscience, Oxford, United Kingdom;* [2]*University of Oxford, Department of Experimental Psychology, Oxford, United Kingdom; and* [3]*Department of Lenguajes y Ciencias de la Computación, Universidad de Málaga, Malaga, Spain*

Rolls ET, Grabenhorst F, Franco L. Prediction of subjective affective state from brain activations15 . *J Neurophysiol* 101: 1294–1308, 2009. First published December 24, 2008; doi:10.1152/jn.91049.2008. Decoding and information theoretic techniques were used to analyze the predictions that can be made from functional magnetic resonance neuroimaging data on individual trials. 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 (with a maximum of 0.6) bits. The prediction was typically a little better with multiple voxels than with one voxel, and the information increased sublinearly with the number of voxels up to typically seven 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, 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, and the information encoded by small numbers of neurons is typically independent. In contrast, the activation of a $3 \times 3 \times 3$-mm voxel reflects the activity of ~0.8 million neurons or their synaptic inputs and is not part of the information encoding used by the brain, thus providing a relatively poor readout of information compared with that available from small populations of neurons.

INTRODUCTION

Predicting which stimulus has been shown, which stimulus is rewarding, or which decision will be taken on an individual trial from the activity of single neurons or populations of single neurons is a fundamental approach to understanding what is represented in a brain region, how it is represented, and how information is processed in the brain to reach a decision. The information available in a neural representation on a single trial is crucial for understanding how the brain performs its computations, and with what information, because the brain cannot average across large numbers of trials when it operates on a single occasion. Important questions that have been addressed include how good the prediction on a single trial is from a single neuron, whether different neurons contribute independently, and how much any stimulus-dependent cross-correlations between neurons contribute relative to that contributed by the firing rate response (Aggelopoulos et al. 2005; Gawne and Richmond 1993; Golomb et al. 1997; Richmond and Opti-

can 1990; Rolls 2008; Rolls and Treves 1998; Rolls et al. 1997a,b; Singer 1999). Analogous questions are now being asked with data from functional neuroimaging of the brain, including 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 (Eger et al. 2008; Hampton and O'Doherty 2007; Haynes and Rees 2005a,b, 2006; Haynes et al. 2007; Kriegeskorte et al. 2006, 2007; Pessoa and Padmala 2005, 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, it was possible to decode or predict whether addition or subtraction would be performed from a set of medial prefrontal voxels within a radius of three voxels with a linear support vector classifier with accuracies in the order of 70%, where chance was 50% (Haynes et al. 2007).

In this study, we developed an information theoretic approach to measure the information from the activations in sets of voxels, basing this on previous information theoretic approaches used for neuronal activity (Aggelopoulos et al. 2005; Franco et al. 2004; Rolls 2008; Rolls et al. 1997a). This enabled us to measure the amount of information provided by any one voxel, whether each voxel carried independent information or whether there was redundancy, how the information obtained scaled with the number of voxels considered, whether combining voxels from different brain areas yielded more information than taking the same number of voxels from one brain area, and whether there was 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, that could potentially encode information about which event was present. This evidence from trial by trial correlations between voxels that depends on the stimulus presented is referred to as stimulus-dependent noise (or trial by trial) correlation information. The "noise" in this case refers to trial by trial variation, and is distinguished from effects related to how similar two stimuli or signals are, averaged over many trials, which is referred to as a signal correlation (Averbeck and Lee 2004; Gawne and Richmond

1993; Oram et al. 1998; Rolls 2008; Shadlen and Newsome 1994).

This information theoretic approach was used to measure how well the activations of a set of voxels could predict the hidden affective state present in an individual before the affective state was reported. The stimuli used were a warm (41°C) pleasant stimulus, a cold (12°C) unpleasant stimulus, and combinations of warm and cold stimuli, applied to the hand. On each trial, the subject received the stimulus but only reported the subjective state it produced after an 8-s delay, by reporting after the delay using rating scales how pleasant and intense the stimulus had been. Measurement of activations produced during the delivery of the stimuli were used to make predictions about the subjective pleasantness and intensity ratings that would be given later in the trial. The use of ratings of both the pleasantness and the intensity of the stimuli on each trial enabled us to test whether there was relatively more information about affective value in some brain regions and about intensity in other brain regions (Rolls and Grabenhorst 2008). The activations produced in different brain regions with these thermal stimuli have been described elsewhere (Rolls et al. 2008), and here we focus on the information theoretic analysis of these data, to assess how well it is possible to predict the subjective state from the brain activations on a single trial.

METHODS

Design

In the experiment described here, we compared brain responses to a warm pleasant stimulus (41°C) applied to the hand (warm2), a cool unpleasant stimulus (12°C) applied to the hand (cold), a combined warm and cold stimulus (warm2+cold), and a second combination designed to be less pleasant (39 + 12°C) (warm1+cold). The stimuli were delivered in random permuted sequence, and on every trial, the participant rated the subjective pleasantness and subjective intensity of the stimulus. Two ratings of pleasantness were taken, one for values in the range 0 (neutral) to +2 (very pleasant) and a second for values in the range 0 to −2 (very unpleasant), to study whether the activations in similar brain areas were correlated with the pleasantness of stimuli both when they were pleasant (≥0) and when they were unpleasant (≤0) or whether different brain areas code for thermal stimuli that are pleasant or unpleasant. For this study, the average of these two pleasantness ratings was used. The participants were instructed to rate the subjective affective experience in terms of pleasantness/unpleasantness, and with the combined thermal stimuli, the participants reported that they did offset each other in terms of the overall subjective pleasantness, which they found easy and natural to rate.

In a previous analysis of this data set (Rolls et al. 2008), we studied how the thermal component stimuli and the mixtures were represented in brain areas identified by prior hypotheses such as the orbitofrontal and anterior cingulate cortex and ventral striatum where the pleasantness and unpleasantness of touch and oral temperature are represented (Guest et al. 2007; Rolls et al. 2003c) and in the insula and somatosensory cortex where thermal stimuli are represented (Brooks et al. 2005; Craig et al. 1996, 2000; Tracey et al. 2000). Given the aims of the study, we used both Statistical Parametric Mapping (SPM) (Wellcome Institute of Cognitive Neurology) correlation analyses between the subjective ratings and the activations in these brain areas and SPM contrasts between the activations produced to the different thermal stimuli, in these brain areas, to study the effects of the thermal stimuli.

Participants

Twelve healthy volunteers (6 male and 6 female; mean age, 26 yr) participated in a study of how affectively pleasant and unpleasant thermal stimuli are represented in the brain (Rolls et al. 2008) and how decisions about these stimuli are made (Grabenhorst et al. 2008b). The analyses described in this study were focused at the single subject level, because we wished to study how well one could predict the hidden affective state in a delay period from brain activations on a single trial in an individual subject and how much information was represented. The main analyses presented were performed on four separate participants and were confirmed as typical by further analyses in the other participants. Ethical approval (Central Oxford Research Ethics Committee) and written informed consent from all subjects were obtained before the experiment.

Stimuli

Controlled cool thermal stimuli were applied using an adapted commercially available Peltier thermode (MEDOC, Haifa, Israel; 30 × 30-mm thermo-conducting surface) strapped to the dorsum of the left hand. The thermode produces a trapezoid-like stimulus, with a time to reach the target temperature of 12°C of 5 s, with a similar period to return to baseline temperature. The plateau temperature was held for 4 s, and subsequent data analyses focused on brain activation during the time of this maintained (plateau) temperature. The warm stimulus was applied using a 20 × 15-mm thermal resistor strapped to the palm of the left hand. The thermal resistor device was designed and built at the Oxford Centre for Functional Magnetic Resonance Imaging of the Brain (FMRIB) and ramped the temperature to 41 (for the warm2 stimulus) or 39°C (for the warm1 stimulus) in <2 s (Bantick et al. 2002). The placement of the stimuli on the dorsum and palm of the hand was designed to minimize thermal interaction between the stimuli in the short delivery period of 4 s and was designed so that even with any topologically mapped representation of the body surface that might be present in the activated brain regions, the regions of activation would be close in the brain. The method of stimulus delivery ensured that the devices were continually in place during the experiment and that only temperature changes were occurring in the stimulation periods. In preliminary testing, the exact temperatures used for each subject were tailored ±2°C, so that warm2 was rated as very pleasant; cold as unpleasant but not painful or very unpleasant; when it was combined with warm2, the combination was at least sometimes more pleasant than neutral, and warm1 was adjusted so that it was less pleasant than warm2 and more pleasant than neutral.

Experimental protocol

During the functional MRI (fMRI) experiment, the subjects gave psychophysical ratings of pleasantness and intensity on every trial, so that correlation analyses between the ratings and the brain activations could be performed. The experimental protocol consisted of an event-related interleaved design presenting in random permuted sequence the four experimental conditions described above. Each trial started at *time 0* with a small 1-s visual stimulus to indicate the start of the trial, and at the same time, the thermal stimulus was switched on to allow it to reach plateau. The plateau was reached by time = 5 s, and a 1-s stimulus appeared on the visual display stating "Rate" to indicate that subjective ratings were needed on this trial. There was a 4-s period in which the temperature stimuli were held constant, and a green cross was shown indicating to the subject that this was the relevant period for which ratings were required. It was made clear to the subjects in the instructions that this was the steady-state period within which the evaluation of the pleasantness and intensity of the stimuli was to be determined by them. The actual ratings were made later, as described next, so that no aspect of making the ratings would occur in the

steady-state period in which the stimuli were being evaluated. After the 4-s plateau period, the thermal stimuli were switched off. The subjective ratings were then made. The first rating was for the pleasantness of the stimulus in the plateau period for values of 0 (neutral) to +2 (very pleasant). The second rating was for the pleasantness of the stimulus in the plateau period for values of 0 (neutral) to −2 (very unpleasant). In this study, the mean of these two ratings was used, producing a single pleasantness value in the range +2 to −2. The instructions to the participants were to rate the overall pleasantness of the stimulus being applied and not its components. The third rating was for the intensity of the stimulus in the plateau period on a scale from 0 (very weak) to 4 (very intense). The ratings were made with a visual analog rating scale in which the subject moved the bar to the appropriate point on the scale using a button box. Subjects were pretrained outside the scanner in the whole procedure and use of the rating scales. Each of the four trial types was presented in random permuted sequence 15 times. This general protocol and design has been used successfully in previous studies to investigate activations and their relation to subjective ratings in cortical areas (de Araujo et al. 2005; Grabenhorst et al. 2007, 2008a; Rolls et al. 2003b,c). On some other trials, instead of "Rate," the word "Decide" appeared, and the subjects had to decide whether they would choose to repeat the particular stimulus that had just been delivered if the opportunity was available after the experiment (Grabenhorst et al. 2008b).

fMRI data acquisition

Images were acquired with a 3.0-T VARIAN/SIEMENS whole body scanner at the FMRIB, where 27 T2*-weighted EPI coronal slices with in-plane resolution of 3 × 3 mm and between-plane spacing of 4 mm were acquired every 2 s (TR = 2). We used the techniques that we have developed over a number of years (de Araujo et al. 2003; O'Doherty et al. 2001) and as described in detail by Wilson et al. (2002) and carefully selected the imaging parameters to minimize susceptibility and distortion artifact in the orbitofrontal cortex. The relevant factors include imaging in the coronal plane, minimizing voxel size in the plane of the imaging, as high a gradient switching frequency as possible (960 Hz), a short echo time of 28 ms, and local shimming for the inferior frontal area. The matrix size was 64 × 64 and the field of view was 192 × 192 mm. Continuous coverage was obtained from +62 (A/P) to −46 (A/P). A whole brain T2*-weighted EPI volume of the above dimensions and an anatomical T1 volume with coronal plane slice thickness 3 mm and in-plane resolution of 1 × 1 mm were also acquired.

fMRI data analysis

The imaging data were analyzed using SPM5 (Wellcome Institute of Cognitive Neurology). Preprocessing of the data used SPM5 realignment, reslicing with sinc interpolation, and normalization to the MNI coordinate system (Montreal Neurological Institute) (Collins et al. 1994). Spatial smoothing with a 6-mm full-width at half-maximum isotropic Gaussian kernel was used only for the conventional single event contrast and correlation analyses with SPM, the results of which are described elsewhere (Rolls et al. 2008), and were used to identify regions for this study of how well the subjective state could be predicted from single trials. The time series at each voxel were low-pass filtered with a hemodynamic response kernel. Time series nonsphericity at each voxel was estimated and corrected for (Friston et al. 2002), and a high-pass filter with a cut-off period of 128 s was applied for the conventional analyses.

For the information theoretic and prediction analyses described here, no spatial or temporal smoothing was used (except for temporal detrending described below), and the raw activation values were extracted from the normalized and realigned volumes (the wr* files in SPM), as described below. Voxels were selected for the prediction

and information theoretic analyses based on statistically significant results in a priori–defined regions for a contrast or correlation in the conventional SPM analyses, the results of which are reported elsewhere (Grabenhorst et al. 2008b; Rolls et al. 2008). The 3 × 3 × 3-mm voxels within a sphere of 3-voxel radius providing 33 voxels were used in the analysis, as were, for comparison, the central voxel alone, and the 7 voxels within the same sphere with the most significant difference in the mean activations between the different conditions being compared. The study of this number of voxels (33) in the analyses is justified by the post hoc finding described in the results that most of the information was encoded in the first seven voxels of a set or fewer.

Data analysis

Techniques have been developed to enable the information provided by populations of simultaneously recorded neurons to be analyzed (Aggelopoulos et al. 2005; Franco et al. 2004; Rolls et al. 1997a), and in this section, we extend these techniques to the analysis of functional imaging data. These techniques enable fundamental questions to be addressed. One is whether each neuron conveys independent information, which is an extremely powerful form of representation if present. In this case, the information increases linearly with the number of neurons, and the number of stimuli or events that can be encoded increases exponentially with the number of neurons (because information is a log measure) (Cover and Thomas 1991; Rolls 2008; Rolls and Deco 2002; Rolls and Treves 1998; Rolls et al. 1997a). If the information increases less than linearly, this indicates the existence of some redundancy in the information conveyed by the neurons, and the information theoretic approach enables this to be measured precisely. A second type of question that can be answered is about the extent to which a pair of neurons that may have correlated activity for some but not other stimuli, by virtue of this stimulus-dependent cross-correlation, encodes information about the stimulus or event. Information theory allows not only the measure of such stimulus-dependent cross-correlation information, but very importantly, how much contribution it makes relative to any change of firing rates that the neurons may show to the stimuli. Indeed, information theory provides the only way that such contributions of different types of encoding, in this case from rates versus correlations, can be compared on the same scale, and indeed assessed to determine whether they are uncorrelated with each other (Aggelopoulos et al. 2005; Franco et al. 2004; Rolls 2008). Information theory can also be applied to different types of data and can show for example on the same measurement scale how much information is available from a single neuron and how this compares to the amount of information available to the whole observer. In the present context, this allows comparison of the information encoded by neurons with the information available from voxels obtained with functional neuroimaging, which is one of the issues we address in this paper.

Techniques for measuring information in this way have been developed for neurophysiology, where the firing rates of neurons are measured, together with the extent to which the neurons have pairwise correlations for some but not other stimuli or events (Aggelopoulos et al. 2005; Franco et al. 2004). Very similar questions arise in functional imaging. To what extent do voxels in the same brain area convey independent information, which might be used to for example predict behavior or an affective state? If the voxels come from different brain areas (both activated in a task), is the information more likely to be independent (as it might be if the brain areas make for example different contributions to a decision)? Furthermore, to what extent do voxels show pairwise behavior that might convey information, for example, predicting outcome in a way that depends on whether the two voxels are both activated at the same time or not? Because these are fundamental questions when predicting outcomes such as behavior, emotional state, etc., from functional neuroimaging data, we developed ways of applying information theoretic approaches to these

particular issues in functional neuroimaging, as described next. The methods are based on those developed for neurophysiology, and further details are provided elsewhere (Aggelopoulos et al. 2005; Franco et al. 2004; Rolls 2008; Rolls et al. 1997a).

Information measurement algorithm

The direct approach to compute the information about a set of stimuli conveyed by the responses of a set of neurons, or in this case, voxels, is to apply the Shannon mutual information measure (Shannon 1948; Cover and Thomas 1991)

$$I(s,\vec{r}) = \sum_{s \in S} \sum_{\vec{r}} P(s,\vec{r}) \log_2 \frac{P(s,\vec{r})}{P(s)P(\vec{r})} \qquad (1)$$

where $P(s,\vec{r})$ is a probability table embodying a relationship between the variable s (here, the stimulus) and $\vec{r}$ (a vector of responses on a single trial, where each element r_i is the activation of a voxel (indexed by i). The activation of a voxel r_i is measured for example by the signal intensity or activation of a voxel or set of voxels on an individual trial from the scanner, as in this study and in related studies (Haynes et al. 2007). It is crucial that the set or vector of the responses, in this case the activation or intensity, is measured on a single trial, because the aim is to study how much information is available on an individual trial from the activations about the behavior or state that occurs on that trial.

However, because the probability table of the relation between the responses and the stimuli, $P(s,\vec{r})$ is so large (given that there may be many stimuli and that the response space is very large, growing exponentially with the number of voxels; Panzeri et al. 1999; Treves and Panzeri 1995), in practice, it is difficult to obtain a sufficient number of trials for every stimulus to generate the probability table accurately. To circumvent this undersampling problem, Rolls et al. (1997a) developed a decoding procedure, in which an estimate (or guess) of which stimulus (called s') was shown on a given trial is made from a comparison of the responses on that trial with the responses made to the whole set of stimuli on other trials. One obtains a conjoint probability table $P(s,s')$, and the mutual information I_p based on probability estimation (PE) decoding between the estimated stimulus s' and the actual stimulus s that was shown can be measured

$$I_p = \sum_{s \in S} \sum_{s' \in S} P(s,s') \log_2 \frac{P(s,s')}{P(s)P(s')} \qquad (2)$$

$$= \sum_{s \in S} P(s) \sum_{s' \in S} P(s'|s) \log_2 \frac{P(s'|s)}{P(s')} \qquad (3)$$

These measurements are in the low dimensional space of the number of stimuli, and therefore the number of trials of data needed for each stimulus is of the order of the number of stimuli, which is feasible in experiments. In practice, it is found that, for accurate information estimates of neurophysiological data with the decoding approach, the number of trials for each stimulus should be at least twice the number of stimuli (with a minimum of 16 trials for each stimulus) (Franco et al. 2004). The advantage of the decoding method (Franco et al. 2004) used here over earlier methods that directly compute the Shannon information (Hatsopoulos et al. 1998; Oram et al. 2001; Rolls et al. 2003a, 2004) is that the decoding method works successfully with large numbers of simultaneously measured responses (Franco et al. 2004; Rolls et al. 1997a).

The decoding procedure essentially compares the vector of responses on a single (test) trial with the average (or distribution of the) response vectors obtained previously on other (training) trials to each stimulus in a cross-validation procedure (Rolls et al. 1997a). This decoding can be as simple as measuring the correlation, or dot (inner) product, between the test trial vector of responses and the response

vectors to each of the stimuli. The result of the decoding might be a best guess or prediction from the responses about which stimulus or condition was present on a trial, and this is shown in Fig. 1 and is referred to as maximum likelihood decoding (Rolls 2008; Rolls et al. 1997a). When the responses are just the magnitudes or activation values of the fMRI signals, just the left part of the table shown in Fig. 1 is used. In this study, we used a Bayesian procedure based on a Gaussian assumption of the activation probability distributions as described in detail by Rolls et al. (1997a, 2003a). This has the advantage that the decoding provides the probability that it was each stimulus in the set of stimuli on one trial and is referred to as PE decoding.

A new step introduced by Franco et al. (2004) and used in this study is to introduce into the table data $(s,\vec{r})$ new columns (shown on the right of Fig. 1) containing a measure of the single trial cross-correlation for some pairs of cells, or, in this case, voxels. The decoding procedure can 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 these cross-correlations are stimulus dependent, their positive contribution to the information encoded can be measured. We note that the information measured with any decoding procedure provides a lower bound on the true information that might be measured directly but that the decoding procedure has been validated and shown to be efficient by Franco et al. (2004).

Further details of the decoding procedures (which have been validated by Franco et al. (2004)) are as follows. The full probability table estimator (PE) algorithm uses a Bayesian approach to extract $P(s'|\vec{r})$ for every single trial from an estimate of the probability $P(\vec{r}|s')$ of a stimulus–response pair made from all the other trials (as shown in Bayes' rule shown in *Eq. 4* in a cross-validation procedure)
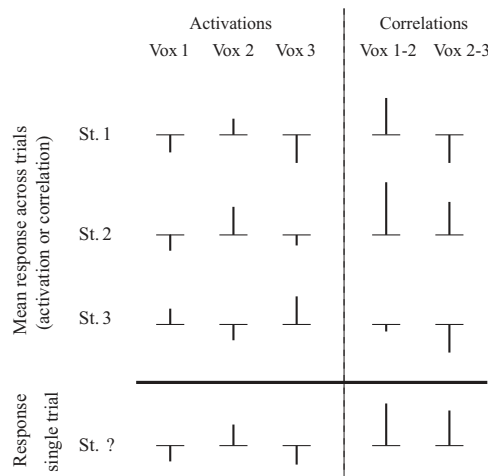

FIG. 1. The *left part* of the diagram shows the average response of each of 3 cells or voxels (labeled as activations for voxels 1, 2, and 3) to a set of 3 stimuli. The *right 2 columns* show a measure (averaged across trials) of the cross-correlation measured on each trial for some pairs of cells or voxels (labeled as correlations voxels 1–2 and 2–3). The *bottom row* (labeled 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 (and their distribution) across trials shown in the top part of the table. From the responses on the single trial, the most probable decoded stimulus in this example is stimulus 2, based on the values of both the rates (or voxel activations) and the cross-correlations between pairs of voxels (Franco et al. 2004).

$$P(s'|\vec{r}) = \frac{P(\vec{r}|s')P(s')}{P(\vec{r})} \qquad (4)$$

where $P(\vec{r})$ (the probability for the vector $\vec{r}$ containing the firing rate of each neuron or the activation of a voxel) is obtained as

$$P(\vec{r}) = \sum_{s'} P(\vec{r}|s')P(s') \qquad (5)$$

This requires knowledge of the response probabilities $P(\vec{r}|s')$ which can be estimated for this purpose from $P(\vec{r},s')$ which is equal to $P(s')\prod_c P(r_c|s')$ where r_c is the response of voxel c. We note that $P(r_c|s')$ is derived from the responses of voxel c from all of the trials except for the current trial for which the probability estimate is being made. The probabilities $P(\vec{r},s')$ are fitted with a Gaussian distribution whose amplitude at r_c gives $P(r_c|s')$. By summing over different test trial responses to the same stimulus s, we can extract the probability that by presenting stimulus s, the response is interpreted as having been elicited by stimulus s'

$$P(s'|s) = \sum_{\vec{r} \epsilon \text{ test}} P(s'|\vec{r})P(\vec{r}|s) \qquad (6)$$

After the decoding procedure, the estimated relative probabilities (normalized to 1) were averaged over all "test" trials for all stimuli to generate a (regularized) table $P_N^R(s,s')$ describing the relative probability of each pair of actual stimulus s and posited stimulus s' (computed with N trials). From this probability table, the mutual information measure I_p was calculated as described above in *Eq. 3*. We note that any decoding procedure can be used in conjunction with information estimates both from the full probability table (to produce I_p) and from the most likely estimated stimulus for each trial in a frequency table $P_N^F(s,s^P)$ (to produce I_{ml}) (referred to as maximum likelihood decoding). With maximum likelihood decoding, the single stimulus that was most likely or predicted (i.e., s^P) by the decoding (Bayesian in this study) to have been presented on that trial was estimated and was used to calculate the percentage correct predictions (Rolls et al. 1997a).

Because the probability tables from which the information is calculated may be unregularized with a small number of trials, a bias correction procedure to correct for the undersampling is applied (Panzeri and Treves 1996; Rolls et al. 1997). The correction term, C_1, to be used takes the form

$$C_1 \approx \frac{1}{2N \log(2)} \sum_s P(s) \sum_{s'} \left[\frac{Q_N^R(s,s')}{P_N^R(s,s')} - \frac{P_N^R(s,s')}{P(s)} \right]$$
$$- \frac{1}{2N \log(2)} \sum_{s'} \left[\frac{Q_N^R(s')}{P_N^R(s')} - P_N^R(s') \right] \quad (7)$$

where $Q_N^R(s,s')$ is the table obtained analogously to $P_N^R(s,s')$ but averaging over all test trials $P^2(s'|r)$ instead of $P(s'|r)$, and where care has to be taken in performing the sums over s', to avoid including stimuli posited to have zero probability. For a derivation of this and other correction terms and for that required to correct I_{ml}, we refer to Panzeri and Treves (1996). In practice, the bias correction that is needed with information estimates using the decoding procedures described here and by Rolls et al. (1997a) is small, typically <10% of the uncorrected estimate of the information, provided that the number of trials for each stimulus is in the order of twice the number of stimuli (with a minimum of 16 trials for each stimulus).

The data from the signals in the voxels used to compute the joint probability distribution $P_N^R(s,s')$ was the signal extracted from the volumes realigned and normalized to MNI space and without spatial smoothing. (In SPM, these are the wr* files.) For each time point for which a signal (i.e., activation value) was needed, one per trial, the signal was the average of that in the volumes that occurred 4 and 6 s after the delivery of the stimulus, which, given the typical delays in activations in fMRI experiments, provides a useful single trial estimate of the signal. The average value of the signal in the preceding 36 volumes was subtracted to subtract temporal variations over the course of the experiment. (High-pass temporal filtering with a duration of 72 s was used. An alternative to averaging 2 poststimulus signal values at the appropriate time is to use a preceding step involving convolution of the signal values for a voxel with the hemodynamic response function, and this produced similar results.) The time point in each trial selected for the analyses of predictions about pleasantness was at $t = 6$ s, which is when the green light indicated to a participant that the thermal sensation at that time should be evaluated for a rating to be made at some time >4 s later. Evidence that the analyses could distinguish the activations about pleasantness at $t = 6$ s from effects related to using the rating scales is that activations in this dataset at $t = 6$ s related to pleasantness were found in the orbitofrontal and pregenual cingulate cortex, whereas activations related to movements involved in making the ratings at times after $t = 10$ s were found in the supplementary and primary motor cortex (Grabenhorst et al. 2008b).

The measure of the cross-correlation c_i between two voxels x and y that was introduced into the data table $(s,\vec{r})$ on each trial i was

$$c_i = [(x_i - x_m)/x_m] \times [(y_i - y_m)/y_m] \qquad (8)$$

where x_i is the activation of voxel x on trial i, and x_m is its mean across trials, and where y_i is the activation of voxel y on trial i, and y_m is its mean across trials. Before this, the mean value of all the voxels was subtracted from each value. This measure of the cross-correlation was used because it can provide a measure on a single trial. These values were scaled to be in the same range as the voxel activation values used in the information theoretic analyses. To not overload the decoding process, only the six voxel pairs from the four voxels with the largest difference in activations between the conditions was used. (This ensured that the voxels were being influenced by the stimulus conditions. These voxels were selected from those in the sphere of radius 3 voxels from the peak voxel.)

If the activations of all the voxels vary together between trials and in a stimulus-independent way, this will reduce the information that can be extracted from a single trial. This is a stimulus-independent noise (i.e., trial by trial) correlation term, and we estimated this by shuffling the order of the trials within a stimulus and comparing the measured information without and with shuffling. This term captures the extent to which the activations of different voxels covary within a trial (and interact with the similarity of the average across trials of the activations of the voxels to each of the set of stimuli (see Franco et al. 2004; Oram et al. 1998; and Rolls et al. 2003a, 2004 for further discussion of the underlying concepts). Part of the concept here is that if stimulus-independent noise has reduced the activations of all voxels on a trial, this noise effect could seriously impair the decoding of which stimulus had been present on that trial. However, if shuffling across trials but within a stimulus has been performed to make a pseudotrial, at least some of the voxels with have more typical activations in the pseudotrial. This allows the magnitude of effects that reflect noise to produce trial by trial variation of the voxel activations (and that does not depend on which stimulus was present) to be estimated, as shown later. This shuffling was performed when measuring how much the information available from voxel activations, i.e., the data shown in the left of Fig. 1, was affected by trial-by-trial variation, which might be produced for example by noise in the measurement process.

The maximum likelihood decoding method described above predicts the particular stimulus that was shown on a trial. Other methods of prediction using the same data were also used, the linear support vector classifier and a backpropagation of error classifier, both to compare with our maximum likelihood method but particularly to

allow comparison with predictions made with these other methods in different studies (Haynes et al. 2007; Ku et al. 2008). The vector support machine and backpropagation of error algorithms used were those implemented in the weka package (Witten and Frank 2005) (http://www.cs.waikato.ac.nz/ml/weka) and were used with cross-validation (i.e., with number of folds = number of trials).

RESULTS

Predictions about pleasantness ratings

First, we show the results of the information theoretic analyses by taking data from *participant 1* in a region with a significant correlation with the pleasantness ratings in the conventional SPM analysis in the medial prefrontal cortex area 10 centered at [−4, 66, 2] ($z = 4.39$, $P < 0.004$; corrected for false discovery rate). Figure 2 (*left*) shows the information available about whether the two stimuli (41 and 12°C) were later rated as pleasant (>0 on a scale from −2 to +2) or unpleasant (≤0) based on different numbers of voxels. We emphasize that, for the information theoretic analysis, the data were divided according to the pleasantness rating given on each trial by the participants and not by the stimulus that had been applied, so that we could test how well activations could be used to predict the hidden affective state in the delay period and not the stimulus that had been delivered. The average amount of information provided by any 1 of the 13 voxels analyzed at these coordinates was 0.20 bits. Taking the average of any two voxels yielded 0.32 bits, of three voxels yielded

~0.37 bits, and of 13 voxels yielded 0.61 bits. The information thus increases as the number of voxels is increased but does not increase linearly. Thus the information provided by the different voxels is not independent, and there is some redundancy. [The asymptotic behavior shown in Fig. 2 is not just because the information ceiling is 1 bit for this binary classification, because the expected shape based on independent information of the voxels and an asymptotic approach to the information ceiling of 1 bit is shown by the dashed line in Fig. 2 (*left*) (Rolls et al. 1997a).]

We performed the type of analysis shown in Fig. 2 for larger numbers of voxels centered at the same coordinate but found that the average value for any one voxel was lower (e.g., for 32 voxels, 0.1 bits), and the asymptote was at 0.43 bits. The fact that the average value for any one voxel was lower than for the 13 voxels shown in Fig. 2 indicates that some of the 32 voxels did not have high information values. The fact that the asymptote is lower for 32 voxels indicates that noise is actually introduced into the decoding by including voxels with low information values. We note that the 13 voxels used for the analysis shown in Fig. 2 were those with the highest t values for a test of the difference in the activations between the two categories within a sphere of 3-voxel radius centered at the coordinates given.

The percentage correct of the predictions for the same dataset as a function of the number of voxels is shown in Fig. 2 (*middle*). It can be seen that the asymptotic value for the 13 voxels is 90% correct (with chance being 50% correct and
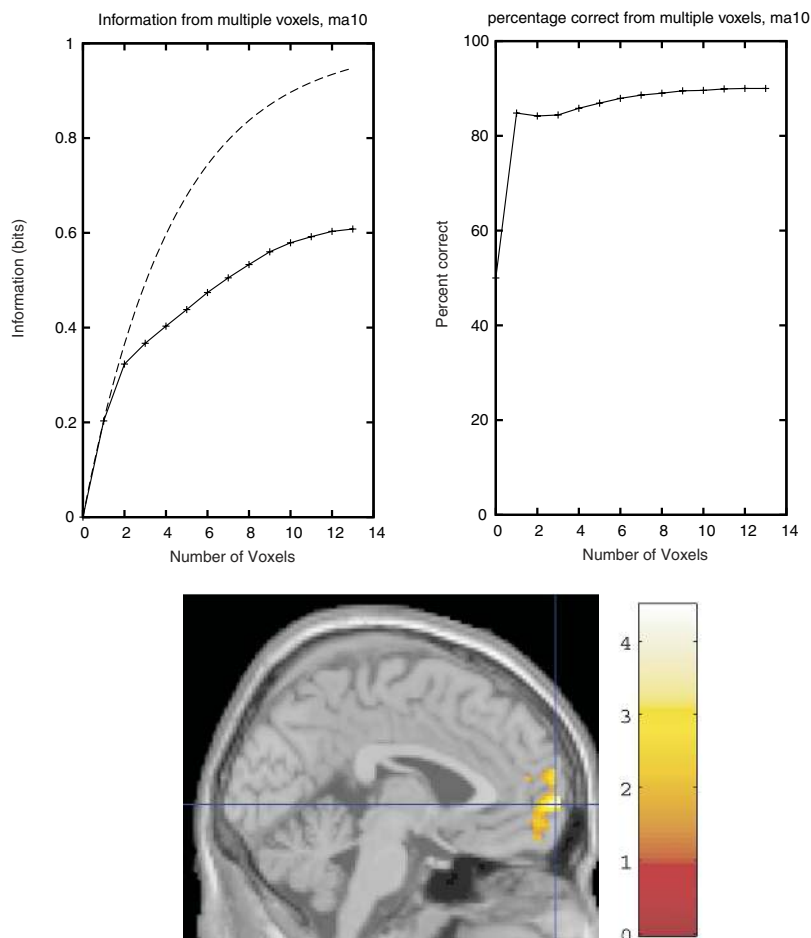

FIG. 2. *Top*: the information available about whether the stimuli were pleasant (>0 on a scale from −2 to +2) or unpleasant (≤0) (*left*), together with the curve that would be produced if the voxels provided independent information (dashed line), and the percentage correct predictions (*right*) based on the activations in different numbers of voxels from the medial prefrontal cortex area 10 centered at [−4, 66, 2]. For the percentage correct, in this and subsequent figures, the chance value is shown as the value when the number of voxels is 0 and is close to 50% but not exactly 50% if there were different numbers of trials for the 2 stimuli. 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 centered at [−4, 66, 2] were obtained.

indicated by the prediction with 0 voxels). With one voxel, the prediction is on average 85% correct, and after this, there was in general an increase in the prediction, with 89% correct possible with on average eight voxels. The shape of the function is different from the information function, because the percentage correct is based just on the most likely single stimulus for a trial, whereas the information measure shown in Fig. 2 (*left*) reflects a probability estimate for each of the stimuli as shown in *Eqs. 2* and *3*. The way in which the prediction or information changes with the number of voxels has not been brought out in previous analyses (Haynes et al. 2007). To check that our maximum likelihood algorithm used to obtain the percentage correct was reasonably efficient, we compared it to the predictions made with the linear support vector method (SVM in Table 1, which has this dataset near the top), which gave 87% correct for 13 voxels, and with the backpropagation of error [multilayer perceptron (MLP) in Table 1] algorithm, which gave 87% correct. Thus the maximum likelihood algorithm used in our program was powerful and efficient with this type of fMRI data.

We also measured how much information was present from this set of voxels (in *participant 1* at $[-4, 66, 2]$) about the intensity of the thermal stimuli. The result was 0.02 bits, and the percentage correct was 60% (as shown in Table 1). Thus the information theoretic approach can provide a quantitative comparison of what can be decoded from a brain region about one property of the hidden internal subjective state (e.g., pleasantness) versus another (e.g., intensity). In this case, much more information was provided about pleasantness than intensity.

Thus far, we considered binary predictions of whether the rating will be pleasant (>0) or unpleasant (≤ 0) from two stimuli: warm (41°C) and cold (12°C). If we make the same binary predictions for the same dataset in *participant 1*, but now based on four stimuli, two of which were mixtures, 0.23 bits and 82% correct were obtained with 13 voxels, 0.18 bits and 80% correct were obtained with 32 voxels, and (as shown in Table 1) 0.20 bits and 82% correct were obtained with 7 voxels. The less good performance is because some of the mixtures were close to the decision border of 0. It was also

TABLE 1. *Information values and predictions for different datasets*

Prediction, n stim	Brain Region	Coordinates	z Value	Number of Voxels	PCC, %	PE Inform Bits	MLP, %	SVM, %
Participant 1								
Pleas 2	Area 10	−4, 66, 2	4.39	32	88	0.43	73	77
Pleas 2	Area 10	−4, 66, 2		13	90	0.61	87	87
Pleas 2	Area 10	−4, 66, 2		7	87	0.48	94	85
Pleas 4	Area 10	−4, 66, 2		7	82	0.2	82	80
Intens 2	Area 10	−4, 66, 2	2.71	33	78	0.2	91	82
Intens 2	Insula	−36, −24, 2	3.81	33	62	0.04	45	41
Pleas 2	Insula	−36, −24, 2	1.78	33	58	0.02	47	66
Pleas 4	Lat OFC	52, 44, −10	4.45	33	80	0.29	87	83
Decide vs. rate	Lat OFC	52, 44, −10	1.80	33	54	0.01	50	57
Decide vs. rate	Premotor	−38, 2, 54	6.30	33	63	0.04	61	62
Pleas 4	PGC	−2, 40, 6	3.68	7	73	0.15	70	76
	Lat OFC	−40, 28, −12	4.82	7	77	0.17	72	82
	Medial OFC	−14, 38, −30	3.72	7	77	0.09	63	66
	Mid OFC	26, 26, −16	3.12	7	63	0.04	62	67
Decide vs. rate	Med 10	8, 60, 10	4.00	7	61	0.03	54	64
Decide vs. rate	Vent premotor	−32, 0, 64	>7.0	7	77	0.21	75	76
Participant 2								
Pleas 4	PGC	−16, 42, 4	3.61	7	76	0.20	87	64
Pleas 4	dACC	−8, 12, 16	3.57	7	71	0.10	63	73
Pleas 4	Mid OFC	−16, 24, −10	4.10	7	68	0.07	55	67
Pleas 4	Lat OFC	40, 44, −2	3.15	7	68	0.05	62	65
Pleas 4	Mid OFC	20, 40, −20	3.30	7	75	0.12	65	69
Decide vs rate	Med 10	−10, 66, 10	6.08	15	69	0.11	62	71
Participant 3								
Pleas 4	vSTR	4, 6, −14	4.88	7	67	0.04	58	65
Pleas 4	Lat OFC	−54, 32, −2	3.84	7	78	0.21	74	82
Pleas 4	PGC	10, 62, 2	6.85	7	70	0.13	77	84
Pleas 4	mOFC	12, 54, −24	5.81	7	72	0.11	66	72
Pleas 4	Mid OFC	−14, 46, −26	5.33	7	62	0.05	75	79
Pleas 4	Lat OFC	42, 46, −8	3.98	7	53	0.07	87	75
Participant 4								
Pleas 4	PGC	0, 42, 0	3.34	7	78	0.03	78	81
Pleas 4	mOFC	−10, 46, −12	3.78	7	76	0.11	75	76
Pleas 4	dACC	2, 26, 32	3.94	7	81	0.19	77	79
Pleas 4	Lat OFC	38, 50, −6	3.24	7	64	0.01	64	71
Decide vs rato	Med 10	12, 60, −8	6.45	7	66	0.06	64	66

PCC %, prediction as percent correct from the decoding; PE inform, information from probability estimation decoding; MLP %, prediction as percent correct from a multilayer perceptron; SVM %, prediction as percent correct from a support vector method; Pleas, binary prediction of pleasantness from number of stimuli indicated. Pleas 2 refers to the warm and cold stimuli applied separately; Decide vs. rate, binary prediction of whether this will be a choice decision or rating trial; Intens, binary prediction of whether the intensity rating was greater or less than the median for that participant; *z*, *z* value from the conventional SPM analysis for the peak voxel. dACC, dorsal anterior cingulate cortex; Lat OFC, lateral orbitofrontal cortex; Med 10, medial prefrontal cortex area 10; OFC, orbitofrontal cortex; PGC, pregenual cingulate cortex; Premotor, premotor cortex; vSTR, ventral striatum.
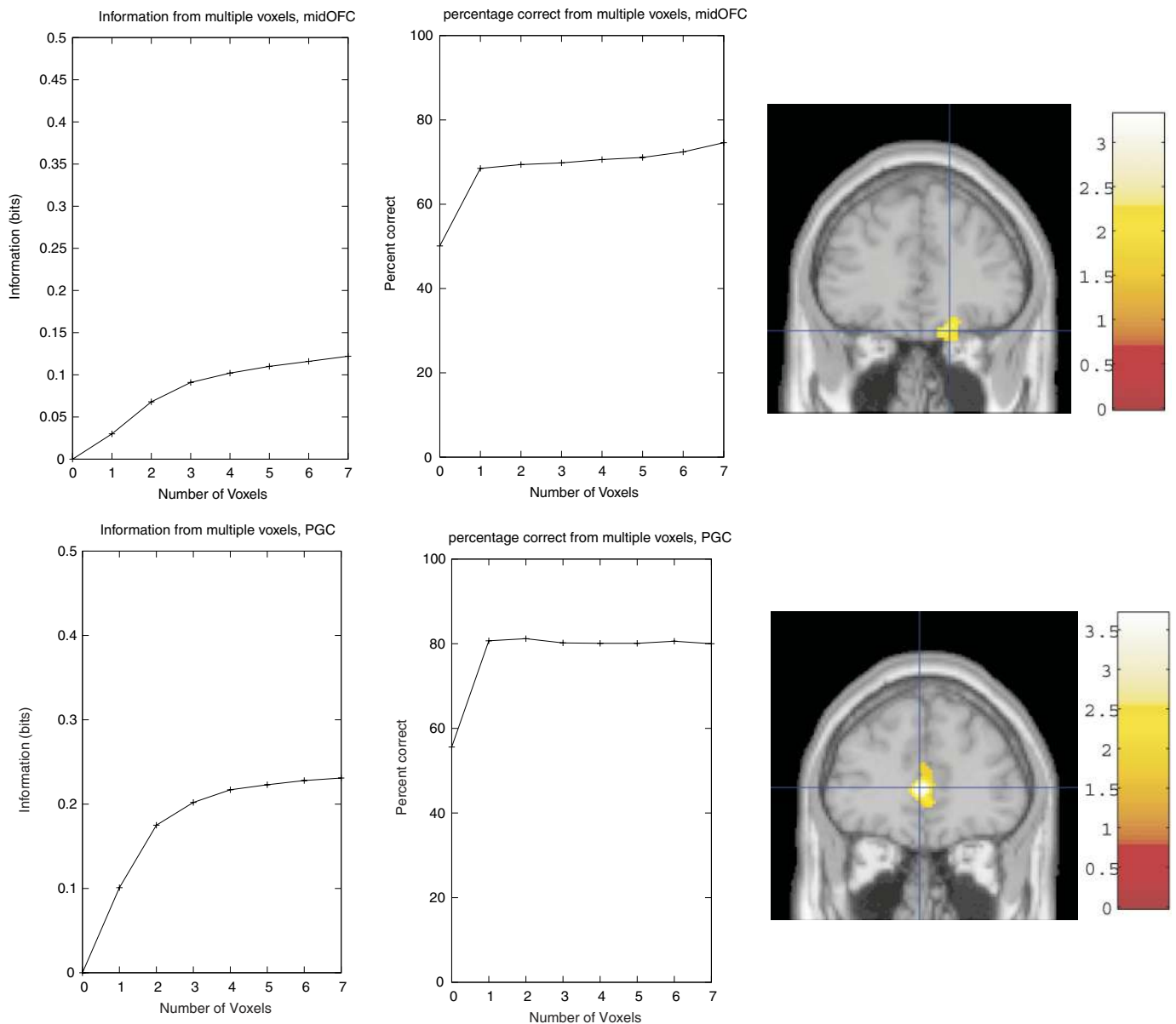
FIG. 3. *Top*: the information available about whether the stimuli were pleasant (>0 on a scale from −2 to +2) or unpleasant (≤0) (*left*) and the percentage correct predictions (*middle*) based on the activations in different numbers of voxels from the mid/medial orbitofrontal cortex centered at [20, 40, −20] (*right*). The prediction was for the ratings that would be made by *participant 2*. *Bottom*: the information available about whether the stimuli were pleasant (>0 on a scale from −2 to +2) or unpleasant (≤0) (*left*) and the percentage correct predictions (*middle*) based on the activations in different numbers of voxels from the pregenual cingulate cortex centered at [−2, 40, 6] (*right*). The prediction was for the ratings that would be made by *participant 1*.

possible to make predictions about larger numbers of affective states than two. For example, taking the three stimuli, warm, cold, and a mixture of the warm (42°C) and cold stimuli, it was possible to predict the stimulus and the pleasantness state it produced at 58% correct (where chance is 33% correct), and 0.25 bits of information were encoded about the three stimuli.

In Fig. 3, we provide examples (with data from *participant 2*) of the predictions and information encoded by different numbers of voxels about the pleasantness ratings that would be given later in the trial from two further brain regions, the mid/medial orbitofrontal cortex (above) and the pregenual cingulate cortex (below), in both of which there are correlations across trials and subjects of the activations with the pleasantness ratings (Rolls et al. 2008). For the orbitofrontal cortex, the prediction was 75% correct, with 0.12 bits of

information (from 7 voxels), and for the pregenual cingulate cortex, the prediction was 73% correct, with 0.15 bits of information (from 7 voxels). In both cases, the prediction was almost as good from one voxel, and the information increased over three to seven voxels. In both these brain areas, the SPM analyses showed a correlation with the pleasantness ratings (Rolls et al. 2008).

Some details of the analyses shown in Fig. 3 (*bottom*) are now considered. The information increase as a function of the number of voxels shown in Fig. 3 (*bottom left*) is sublinear, indicating some redundancy of the information provided by the different voxels. The fact that the graph of percentage correct (Fig. 3, *bottom right*) shows a small decline of its values as the number of voxels increases is also a consequence of the redundancy between voxels that happens to be highlighted

because maximum likelihood (ML) decoding was used to calculate the percent correct, whereas PE decoding was used to calculate the information. The ML estimation method used for the computation of the percentage correct decoding uses a single stimulus (that found most likely to elicit the observed response) rather than the probabilities estimated for each stimulus, and thus is more strongly affected by the redundancy of the information conveyed by the different voxels and chance effects because of the selection of different voxels when there are limited numbers of trials and more trials are added that add noise but no further information. We note that with the PE method used to calculate the information, the high regularization tends to produce a smoothed gradually increasing information estimate as the number of voxels is increased (as

illustrated in Figs. 2–4). We were able to confirm that if the ML decoding is used to calculate the information, then the shape of the curve becomes somewhat more similar to that of the percent correct prediction as the number of voxels is increased. Because the predictions typically did not improve with more than seven voxels, and sometimes became worse as more voxels were added that introduced noise but no further useful information, the data shown in Table 1 and elsewhere are for seven voxels except where stated.

Table 1 summarizes data from many such analyses about the predictions of the pleasantness ratings that will be made later in trial. [In Table 1, Pleas 2 refers to binary predictions of pleasant vs. unpleasant using 2 thermal stimuli (41 and 12°C) and Pleas 4 to binary predictions using the 4 thermal stimuli.]
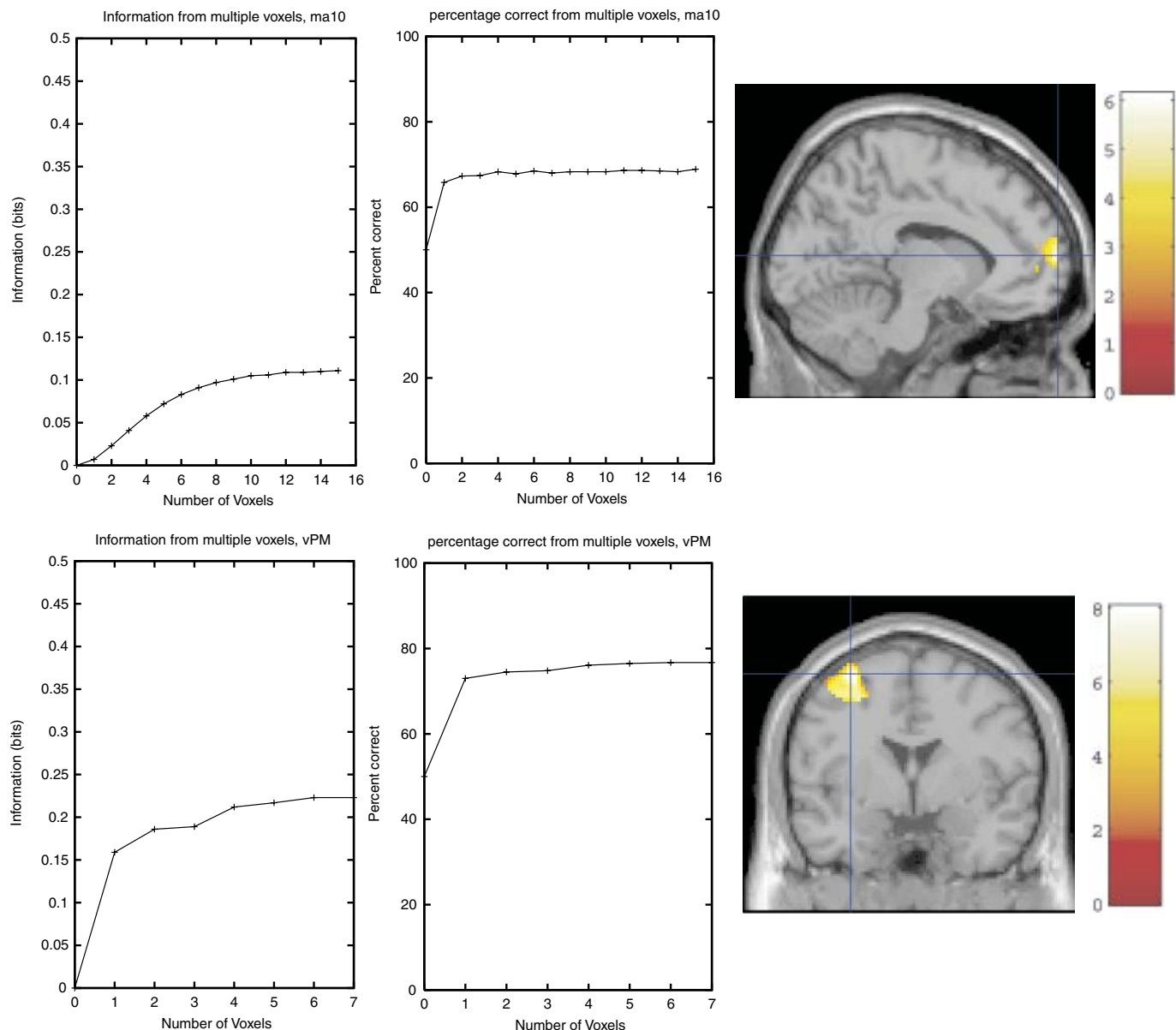

FIG. 4. *Top*: prediction (*middle*) and information encoded (*left*) as a function of the number of voxels in the medial prefrontal cortex area 10 [−10, 66, 10] about whether the trial was one on which a decision about the thermal stimulus (whether it would be accepted in future) was being made or whether it was a trial on which ratings on continuous scales of pleasantness and intensity were to be made. The brain region in *participant 2* from which the activations were measured is shown on the *right*. *Bottom*: a similar analysis for activations in the ventral premotor cortex (vPM) [−32, 0, 64] in *participant 1*.

In terms of brain regions, it was possible to predict pleasantness ratings (pleasant vs. unpleasant) from the orbitofrontal cortex with a mean percent correct of 71% (SD = 8%, $n = 13$ sites in 4 subjects, best 3 regions 80, 78, and 77% correct), and the average information available was 0.11 bits (SD = 0.07 bits, best 3 regions 0.29, 0.27, and 0.21 bits). For the pregenal cingulate cortex, it was possible to predict pleasantness ratings (pleasant vs. unpleasant) with a mean percent correct of 74% (SD = 4%, $n = 4$ sites, best 2 regions 78 and 76% correct), and the average information available was 0.13 bits (SD = 0.07 bits, best 2 regions 0.20 and 0.15 bits). From medial area 10, one site yielded prediction of pleasantness ratings of 90%, with the information available being 0.61 bits. To place these results in the context of the statistics in the SPM analyses, the z values for the peak voxels in the related contrast analyses (and correlation analyses with the rating as a regressor) were typically >4 as shown in Table 1, and the z values in the group random effects analyses were typically in the range 3–4 as shown elsewhere (Grabenhorst et al. 2008b; Rolls et al. 2008).

As noted in METHODS, these information theoretic and prediction analyses are primarily at the single subject level, and we showed data for four individual participants in Table 1. To check that these results were representative, we performed further analyses on other participants scanned in the original experiment (Rolls et al. 2008). Analogous results were found in these further analyses. For example, when testing for predictions of pleasant versus unpleasant ratings from four stimuli for voxels in the orbitofrontal cortex, the mean percent correct prediction (across 7 further participants) was 69%, and the mean information was 0.05 bits. Over all these 11 participants, the mean prediction from the orbitofrontal cortex activations of whether the affective state would later be rated as pleasant or unpleasant was 71 ± 2.5% (SE) correct, and the ability to make a prediction from the activations that was better than chance was highly significant ($t = 8.64$, df = 10, $P < 0.00001$).

The results across all the datasets show that what was shown in Figs. 2 and 3 is the general pattern of results. That is, in all cases, the information increases sublinearly with the number of voxels; the information maximum was obtained for a set of voxels that was typically in the order of 7–20, with 33 voxels either yielding no more information, or in some cases, less because of the introduction of noisy measures to the decoding algorithm as the number of voxels was increased. In terms of predictions, the prediction that could be made from any one voxel in a region was typically good and improved typically by <7% as more voxels, up to typically eight, were added.

Predictions about pleasantness based on data from two brain regions versus one brain region

We next consider how the information from voxels from different brain areas adds compared with voxels in the same brain area. We consider predictions about pleasantness and the information encoded about pleasantness across the two brain regions shown in Fig. 3: the mid/medial orbitofrontal cortex and the pregenual cingulate cortex in *participant 2*. In this case, three voxels (selected repeatedly at random from the best 7) in the pregenual cingulate cortex and four voxels (from the best 7) in the orbitofrontal cortex gave 0.20 bits and 81.4% correct, whereas the seven voxels from the pregenual cingulate cortex gave 0.13 bits and 78% correct, and the seven voxels from the

orbitofrontal cortex gave 0.16 bits and 78% correct. Thus there was little difference in whether the voxels came from the same or different brain regions, implying that, in this case, the evidence available from both regions was similar, at least for this pleasantness prediction and encoding.

The overall results for the information and predictions by considering activations from two versus one brain area were as follows. We consider predictions about pleasantness and the information encoded about pleasantness in 11 different tests performed in four subjects involving combinations of four or three voxels from two brain regions that included the medial/mid orbitofrontal cortex, the pregenual cingulate cortex, the dorsal part of the anterior cingulate cortex, and the lateral orbitofrontal cortex. Across the 11 tests, the mean ratio of the information obtained from two sites compared with the activations taken from the better of the sites of each pair was 1.06 ± 0.21 (SD). (The relevant comparison is the better site, as taking any 3–4 voxels from the seven best voxels at a site provides most of the information.) This ratio was not significantly different from 1.00 ($t = 0.99$, df = 10, $P = 0.34$). Thus overall, there was no evidence that, for this binary prediction, of whether the rating made later would be pleasant versus unpleasant, taking voxels at random from the sets of voxels at two sites provided more information than when the voxels came from the better of the two sites. Similarly, the prediction of whether the stimulus was pleasant versus unpleasant was not improved by taking voxels from two areas versus the same number of voxels from the better of the two areas (mean percent correct from the better of 2 areas calculated over 7 voxels = 79%, mean percent correct from 7 voxels taken from two areas = 82%, ratio = 1.05, SD = 0.13, $t = 1.39$, $P = 0.19$, df = 10).

Predictions about intensity

We tested for brain areas from which intensity can be predicted and for which affective value cannot. An example was found in the somatosensory insula [38, 0, 14], where from 33 voxels, the prediction of intensity was 66.7% correct with 0.02 bits, whereas the prediction of pleasantness was 55.0% correct with 0.00 bits of information. Dissociations of this type based on the information provided in different brain areas by representations about different properties of stimuli or events can provide a quantitative approach to the different functionality of different brain areas. Further examples are shown in Table 1, in which some brain areas provide information about for example affective value but not about choice decision making, supporting what was found by SPM analyses (Grabenhorst et al. 2008b).

Predictions about mental operations involved in decision making versus subjective ratings

Figure 4 compares information theoretic analyses for a brain area from which the task being performed by the subject, decision making versus rating, produces different activations, with more activation when decisions were being taken (Grabenhorst et al. 2008b). The activation value for each voxel was the fMRI signal when the thermal stimulus was on and the subject had been instructed 1 s earlier that the trial was either one on which a binary decision was required (of whether or not

they would choose the stimulus) or was a trial on which ratings of pleasantness and intensity were to be made, in both cases after a delay period. Figure 4 (*right*) shows that it is possible to predict with 69% correct on a single trial by the activations in medial prefrontal cortex area 10 whether it is a decision trial or a rating trial. This level of prediction was possible from the 15 voxels, with 66% correct based on 1 voxel. Figure 4 (*left*) shows that 0.11 bits of information were present from 15 voxels, with 7 voxels providing most of the information. The amount of information from any one voxel is quite low (0.015 bits), and this is associated with an approximately linear increase of information over the first seven voxels.

As shown in Table 1, this was a typical result across participants, with similar predictions of decision making versus rating shown in three subjects from the activation on a single trial in medial prefrontal cortex area 10. It was also possible to predict that it was a decision-making trial from activations in the ventral premotor cortex, as shown in Table 1, and this is of interest, because this region is implicated in decision making by single neuron recording studies (Romo et al. 2004).

Information in the correlations between voxel activations

We used the information theoretic method to measure how much information was present in stimulus-dependent cross-correlations between the voxels. This was performed by using the decoding based only on the correlations between voxels on each trial indicated in the *right columns* of Fig. 1. Six correlation values between pairs of voxels were used, and these were from the four voxels in a dataset that had the largest difference in activation to the two thermal stimuli, warm and cold, to ensure that these are voxels influenced by the stimuli and that would contribute to significant effects in contrast and correlation analyses with SPM. Ten datasets from four subjects were analyzed in this way. The average information available from the stimulus-dependent noise cross-correlations in these 10 datasets was 0.043 ± 0.070 bits. This was not significantly different ($P = 0.24$, t-test) from the information measured when the data were randomly permuted between trials within a stimulus, to break any trial-by-trial noise cross-correlation (0.021 ± 0.037 bits). Thus there was no evidence for information in stimulus-dependent noise (trial-by-trial) correlations. Indeed, if we take the difference of the measured and shuffled values, obtaining 0.022 bits, we find that this is very small compared with the information measured from the activation values of the voxels, which was 0.149 ± 0.035 bits, that is, 6.8 times larger, and significantly different ($P < 0.004$, t-test).

With the approach shown in Fig. 1, we were also able to measure from just the activation values on the *left* of Fig. 1, the effect of trial-by-trial or "noise" effects that were stimulus independent. This was implemented by randomly permuting the activation values within a voxel and within a stimulus across trials. For the same 10 datasets, the measured information after the random shuffling was 0.408 ± 0.077 bits, which is much higher than the true 0.149 bits measured with the data not shuffled between trials. The reason for this is that, on some trials, the values for all the voxels may be lower than usual, and on other trials, they may all be higher than usual, with this occurring independently of which stimulus was present. The effect of this type of stimulus independent noise (i.e., trial-by-trial) variation is to make the decoding of the data from any one

trial difficult, because all the voxel activations may randomly be higher or lower than usual on a given trial. (In this situation, the shuffling between trials increases the information measured, because at least some of the voxels on the pseudotrials will have activations that are more representative of what occurs usually.) Put quantitatively, the loss of information produced by stimulus-independent noise or trial by trial correlation of the voxel activation values was $0.408 - 0.149$ bits $= 0.259$ bits. Put another way, the stimulus-independent noise correlations resulted in a loss of 63.5% of the information (0.259/0.408). The source of this noise is probably largely caused by noise in the fMRI BOLD signal measurement process itself, and it is interesting to see it quantified.

Consistent with these analyses, the average correlations across stimuli and trials between the voxel pairs were quite high, with a mean Pearson correlation of 0.83 ± 0.09 (SD). For comparison, the representations of different stimuli provided by a population of inferior temporal cortex neurons are relatively 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 (SE)] (Franco et al. 2007).

Perhaps the most important point from these correlation analyses is that no significant information was available in the stimulus-dependent cross-correlations between voxels.

DISCUSSION

The application of the methods described here enabled us to predict hidden affective states on a single trial produced by warm versus cold stimuli with quite high levels of accuracy, typically 60–80% correct (with a mean of 71% correct for predictions of pleasantness from the orbitofrontal and cingulate cortices) as shown for the four participants in Table 1. Furthermore, over all 11 participants, the mean prediction from the orbitofrontal cortex activations of whether the affective state would later be rated as pleasant or unpleasant was $71 \pm 2.5\%$ (SE) correct, and the ability to make a prediction from the activations that was better than chance was highly statistically significant ($P < 0.00001$). The percentage correct for the predictions is comparable to some other studies in which predictions of hidden states have been reported. For example, in a study in which the prediction was about whether a subject would add or subtract, the average prediction accuracy across subjects from the activation of multiple voxels was 70% (Haynes et al. 2007). However, the information theoretic approach used here enables much more than simple predictions from brain states to be analyzed.

First, we analyzed how the prediction, and the maximum likelihood information that corresponds to this, varies with the number of voxels. For most sites, the results were similar to those shown in Figs. 2–4, in that the predictions (percent correct) were not much improved by adding more than one voxel. In fact, what Figs. 2–4 show are the average predictions from any one voxel in the set, from any two voxels, etc. Of course, if a particular voxel with little information is selected, and a second is added with more information, the second voxel will add to the first. Therefore the results in Figs. 2–4 must be understood as showing what happens on average with any one voxel in the set analyzed, any two voxels, etc. Provided that there is a small number of voxels in the set, the average across

voxels is likely to be close to the peak of the prediction that can be made from a voxel, but with a large number of voxels in a set, the best voxel may perform better than the average for one voxel. We used the average for one voxel here so that we can compare this value with the values for combinations of two or more voxels from the same set. We ensured that the values reported for one voxel were close to the maximum from any one voxel by checking the data with small datasets of seven or fewer voxels.

Second, we analyzed how the probability estimation information increases with the number of voxels. Here we found, as shown in Figs. 2–4, that the information typically increases as more voxels are added. However, the information did not increase linearly with the number of voxels, indicating that the voxels were not providing independent information and that instead there was redundancy because of correlated profiles across the set of stimuli of the different voxels: these are referred to as signal correlations (Rolls 2008; Rolls et al. 2003a, 2004.

These findings can be very interestingly compared with the information encoding provided by single neurons and by populations of single neurons. We are able to make this comparison directly because we used the same information analysis routines to measure the information from neurons and from voxel activations. Let us consider some of the main findings from single neurons (Rolls 2008). If we consider an analogous task analyzed in monkeys performing a visual discrimination task in which one visual stimulus, a triangle, predicted fruit juice reward (a pleasant stimulus), and the other visual stimulus, a square, predicted a saline taste (an unpleasant stimulus), a typical orbitofrontal cortex single neuron such as the one shown in Fig. 5 can predict the affective choice with 90% correct and 0.5 bits of (PE) information on a single trial (data from Rolls et al. 1996b; new information theoretic analysis performed for this paper). This analysis is supported by further data for neurons in the macaque orbitofrontal cortex, in that new analyses of the information about a set of six tastants (glucose 1.0 M, NaCl 0.1 M, HCl 0.01 M, quinine-HCl 0.001 M, monosodium glutamate 0.1 M, and distilled water) provided by orbitofrontal cortex neurons about which taste stimulus had been presented was 0.45 bits for each neuron, averaged across 135 gustatory neurons recorded in previous studies (Critchley and Rolls 1996; Rolls et al. 1996a, 1999). Further evidence that these single neuron information values are representative is that the average (probability estimation) values were 0.3–0.4 bits per neuron for populations of inferior temporal cortex neurons encoding which visual stimulus was shown (Rolls et al. 1997a). Thus the information available and the prediction from a single neuron is typically better than that achieved by the activations from a single voxel containing hundreds of thousands of neurons, as shown in Table 1, with consistent fMRI results obtained in other studies (Eger et al. 2008; Hampton and O'Doherty 2007; Haynes and Rees 2005a,b 2006; Haynes et al. 2007; Kriegeskorte et al. 2006, 2007; Pessoa and Padmala 2005, 2007). Indeed, as shown in Table 1, the average information for sets of seven or more voxels in the orbitofrontal cortex coding for pleasant versus unpleasant was 0.11 bits. Thus much more information is available from a single neuron in the orbitofrontal cortex (or inferior temporal visual cortex) than is available from seven voxels in the human orbitofrontal cortex containing very large

Orbitofrontal cortex neuron: visual discrimination task
Predicts choice of rewarding visual stimulus with 90% correct
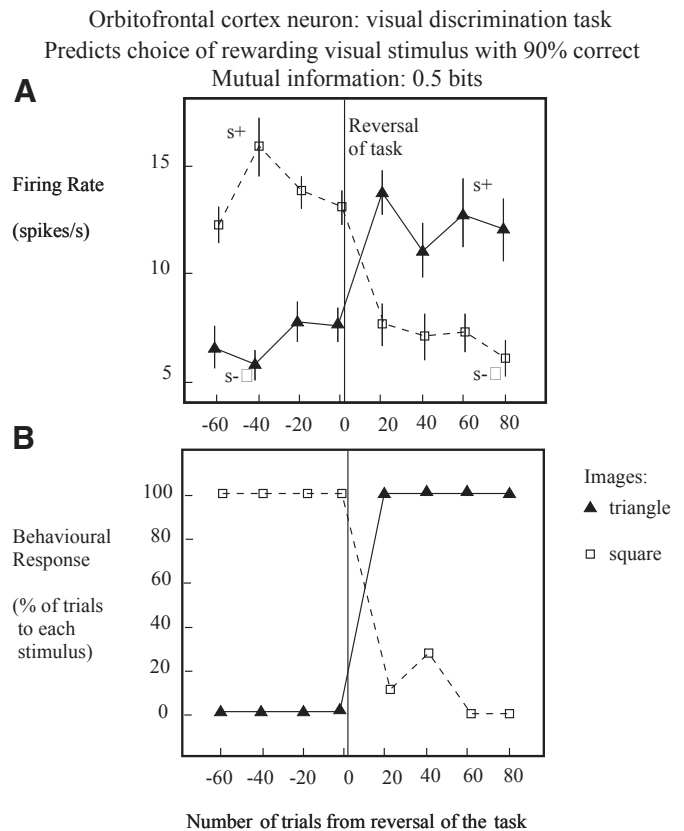Mutual information: 0.5 bits

FIG. 5. Orbitofrontal cortex: visual discrimination reversal. The activity of an orbitofrontal cortex visual neuron during performance of a visual discrimination task and its reversal. The stimuli were a triangle and a square presented on a video monitor. *A*: each point represents the mean poststimulus activity in a 500-ms period of the neuron to ~10 trials of the different visual stimuli. The SE of these responses is shown. After 60 trials of the task, the reward associations of the visual stimuli were reversed (+, lick response to that visual stimulus produces fruit juice reward; −, lick response to that visual stimulus results in a small drop of aversive tasting saline). This neuron reversed its responses to the visual stimuli following the task reversal. *B*: The behavioral response of the monkey to the task. It is shown that the monkey performs well, in that he rapidly learns to lick only to the visual stimulus associated with fruit juice reward. The information about which decision would be taken on each trial was calculated from the neuronal responses in the prereversal set of trials, using the number of spikes from the neuron in a 500-ms period starting 100 ms after stimulus onset (Rolls et al. 1996).

numbers of neurons. Part of the difference is that the fMRI BOLD signal is inherently noisy with variation from trial to trial, and this stimulus-independent noise correlation, quantified above to result in a loss of 63.5% of the information that might be available without this trial by trial variation, accounts in part for the difference between the information that can be read from single neurons and from fMRI voxel activations. However, there are more fundamental differences, as follows.

Another difference is that the information from single neurons typically increases linearly with the number of neurons (at least up to the order of tens of neurons) (Rolls et al. 1997a), indicating a very powerful encoding principle: that each neuron carries information that is independent from that of other neurons, at least in high-order visual areas where many possible stimuli are encoded. This is factorial encoding. This is not a property of the information available from different numbers of voxels, as shown in Figs. 2–4.

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, because 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, 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. Therefore, at the neuron level, because 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 with the activation of another voxel, because 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 30,000 neurons/mm^3 (Abeles 1991; Rolls 2008), a $3 \times 3 \times 3$-mm 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 that tends to place neurons with similar responses close together in the map and neurons with different responses further apart in the map (Rolls 2008). Therefore some information will be available about which stimulus was shown by measuring the average activation in different parts of the map. However, 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. Because the code provided by neurons is independent, the code can never be read adequately by any process that averages across many neurons (and synaptic currents) (Logothetis 2008), such as fMRI.

Third, we 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).

Fourth, the comparison of information from multiple voxels within a brain area compared with the information from the same number of voxels but from different brain areas showed that there was no advantage to taking the evidence from more than one brain area. This was found in a situation in which the two brain areas each had activations related to the same binary prediction. It might have been the case that voxels from different brain areas were less correlated and thus provided more information, but this was not found. In a task in a higher-dimensional space (i.e., with more alternatives), and where the evidence had to incorporate evidence from different sources, such as whether the stimulus is both warm and blue, combining evidence from different brain areas would be expected to be advantageous.

The predictions from and the information encoded by voxels as described here are related to what can be performed based on a single trial of data. The reason for this is that to understand information encoding and transmission in the brain, and how the brain produces a state, decision, or action, what is relevant is what happens on a single trial (Rolls 2008). On the other hand, if one wishes to know whether there is a significant difference between the activations in two conditions, one performs a statistical analysis to test whether the mean activations are significantly different based on all the trials of data available, as in a standard contrast analysis with fMRI. In relation to this, we found in this study that only quite significant statistical values (greater than $z = 3.5$) for a voxel in a conventional contrast analysis with 15 trials per condition are likely to contain much information (>0.1 bits) or to be useful for good prediction (better than 75%), on a single trial, as shown in Table 1.

We note that information analyses of neuronal activity are performed within a subject, so that one can compare the encoding by different neurons perhaps in different brain areas and address what carries the information (e.g., the number of spikes vs. stimulus-dependent neuronal synchronization), how the information scales with the number of neurons, and how the information encoded by single or populations of neurons compares with that being used by the subject to perform the task. This is how we have analyzed the information in this study, aimed at understanding some of the principles of the information encoded by voxels in functional neuroimaging activations. One can make predictions from the voxel activity, and compare them to the subject's performance. If one wishes to make a prediction from the activation of particular voxels in any subject in the population of subjects, it is of course possible to perform a random effects analysis in which the data from a set of subjects is combined (Haynes et al. 2007), but this is not the aim here. It would also be possible to predict how pleasant a stimulus was on average for a subject by averaging across trials within a subject, but again that does not address the issue of information encoding and transmission in the brain when a particular decision is reached or value is described on each trial.

Information theory goes beyond making predictions of percentage correct performance when applied to neuronal and functional neuroimaging data because independent contributions sum linearly when the information transmitted is mea-

sured, and this is not the case for percentage correct (Cover and Thomas 1991; Rolls 2008). It is this property of information theory that allows one, as shown here, to address questions such as whether neurons (or voxels) convey independent information or whether there is redundancy; how much one can learn from neuronal firing rates (or voxel activations) compared with how much one can learn from stimulus-dependent cross-correlations between neurons (or voxels), and whether these two contributions are uncorrelated with each other (i.e., independent); how much one can learn by combining evidence from nearby neurons (or voxels) compared with more distant neurons (or voxels), which addresses whether there is local redundancy, and whether it is useful to measure from more than the single most strongly activated voxel; to what extent evidence is lost because of signal correlations (i.e., correlations between responses that are related to the similarity of the input) versus noise correlations (stimulus-independent trial-by-trial variation, caused, for example, by measurement noise). The use of information theory also allows direct comparisons on the same absolute scale (bits) between different types of measure, for example, what evidence is provided by neuronal firing rates versus voxel activations versus behavioral reports. Although information theory is the only way to address these issues quantitatively, it is more complicated than measuring the percent correct, and care is needed in its use. For example, a decoding step may be needed, because many trials of data as possible are desirable, and it may be necessary to correct the information estimates for the limited number of trials of data that are usually available. These issues are covered in depth by Rolls (2008).

REFERENCES

Abeles M. *Corticonics: Neural Circuits of the Cerebral Cortex.* New York: Cambridge, 1991.

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

Averbeck BB, Lee D. Coding and transmission of information by neural ensembles. *Trends Neurosci* 27: 225–230, 2004.

Bantick SJ, Wise RG, Ploghaus A, Clare S, Smith SM, Tracey I. Imaging how attention modulates pain in humans using functional MRI. *Brain* 125: 310–319, 2002.

Brooks JC, Zambreanu L, Godinez A, Craig AD, Tracey I. Somatotopic organisation of the human insula to painful heat studied with high resolution functional imaging. *NeuroImage* 27: 201–209, 2005.

Collins DL, Neelin P, Peters TM, Evans AC. Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *J Comput Assist Tomogr* 18: 192–205, 1994.

Cover TM, Thomas JA. *Elements of Information Theory.* New York: Wiley, 1991.

Craig AD, Chen K, Bandy D, Reiman EM. Thermosensory activation of insular cortex. *Nat Neurosci* 3: 184–190, 2000.

Craig AD, Reiman EM, Evans A, Bushnell MC. Functional imaging of an illusion of pain. *Nature* 384: 258–260, 1996.

Critchley HD, Rolls ET. Responses of primate taste cortex neurons to the astringent tastant tannic acid. *Chem Senses* 21: 135–145, 1996.

de Araujo IET, Kringelbach ML, Rolls ET, Hobden P. The representation of umami taste in the human brain. *J Neurophysiol* 90: 313–319, 2003.

de Araujo IET, Rolls ET, Velazco MI, Margot C, Cayeux I. Cognitive modulation of olfactory processing. *Neuron* 46: 671–679, 2005.

Eger E, Ashburner J, Haynes JD, Dolan RJ, Rees G. fMRI activity patterns in human LOC carry information about object exemplars within category. *J Cogn Neurosci* 20: 356–370, 2008.

Franco L, Rolls ET, Aggelopoulos NC, Jerez JM. Neuronal selectivity, population sparseness, and ergodicity in the inferior temporal visual cortex. *Biol Cybern* 96: 547–560, 2007.

Franco L, Rolls ET, Aggelopoulos NC, Treves A. The use of decoding to analyze the contribution to the information of the correlations between the firing of simultaneously recorded neurons. *Exp Brain Res* 155: 370–384, 2004.

Friston KJ, Glaser DE, Henson RN, Kiebel S, Phillips C, Ashburner J. Classical and Bayesian inference in neuroimaging: applications. *NeuroImage* 16: 484–512, 2002.

Gawne TJ, Richmond BJ. How independent are the messages carried by adjacent inferior temporal cortical neurons? *J Neurosci* 13: 2758–2771, 1993.

Golomb D, Hertz J, Panzeri S, Treves A, Richmond B. How well can we estimate the information carried in neuronal responses from limited samples? *Neural Comput* 9: 649–665, 1997.

Grabenhorst F, Rolls ET, Bilderbeck A. How cognition modulates affective responses to taste and flavor: top down influences on the orbitofrontal and pregenual cingulate cortices. *Cerebral Cortex* 18: 1549–1559, 2008a.

Grabenhorst F, Rolls ET, Margot C, da Silva MAAP, Velazco MI. How pleasant and unpleasant stimuli combine in different brain regions: odor mixtures. *J Neurosci* 27: 13532–13540, 2007.

Grabenhorst F, Rolls ET, Parris BA. From affective value to decision-making in the prefrontal cortex. *Eur J Neurosci* 28: 1930–1939, 2008b.

Guest S, Grabenhorst F, Essick G, Chen Y, Young M, McGlone F, de Araujo I, Rolls ET. Human cortical representation of oral temperature. *Physiol Behav* 92: 975–984, 2007.

Hampton AN, O'Doherty J, P. Decoding the neural substrates of reward-related decision making with functional MRI. *Proc Natl Acad Sci USA* 104: 1377–1382, 2007.

Hatsopoulos NG, Ojakangas CL, Paninski L, Donoghue JP. Information about movement direction obtained by synchronous activity of motor cortical neurons. *Proc Natl Acad Sci USA* 95: 15706–15711, 1998.

Haynes JD, Rees G. Decoding mental states from brain activity in humans. *Nat Rev* 7: 523–534, 2006.

Haynes JD, Rees G. Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nat Neurosci* 8: 686–691, 2005a.

Haynes JD, Rees G. Predicting the stream of consciousness from activity in human visual cortex. *Curr Biol* 15: 1301–1307, 2005b.

Haynes JD, Sakai K, Rees G, Gilbert S, Frith C, Passingham RE. Reading hidden intentions in the human brain. *Curr Biol* 17: 323–328, 2007.

Kriegeskorte N, Formisano E, Sorger B, Goebel R. Individual faces elicit distinct response patterns in human anterior temporal cortex. *Proc Natl Acad Sci USA* 104: 20600–20605, 2007.

Kriegeskorte N, Goebel R, Bandettini P. Information-based functional brain mapping. *Proc Natl Acad Sci USA* 103: 3863–3868, 2006.

Ku SP, Gretton A, Macke J, Logothetis NK. Comparison of pattern recognition methods in classifying high-resolution BOLD signals obtained at high magnetic field in monkeys. *Magn Reson Imag* 26: 1007–1014, 2008.

Logothetis NK. What we can do and what we cannot do with fMRI. *Nature* 453: 869–878, 2008.

O'Doherty J, Rolls ET, Francis S, Bowtell R, McGlone F. The representation of pleasant and aversive taste in the human brain. *J Neurophysiol* 85: 1315–1321, 2001.

Oram MW, Foldiak P, Perrett DI, Sengpiel F. The 'Ideal Homunculus': decoding neural population signals. *Trends Neurosci* 21: 259–265, 1998.

Oram MW, Hatsopoulos NG, Richmond BJ, Donoghue JP. Excess synchrony in motor cortical neurons provides direction information that is redundant with the information from coarse temporal response measures. *J Neurophysiol* 86: 1700–1716, 2001.

Panzeri S, Treves A. Analytical estimates of limited sampling biases in different information measures. *Network* 7: 87–107, 1996.

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

Pessoa L, Padmala S. Decoding near-threshold perception of fear from distributed single-trial brain activation. *Cereb Cortex* 17: 691–701, 2007.

Pessoa L, Padmala S. Quantitative prediction of perceptual decisions during near-threshold fear detection. *Proc Natl Acad Sci USA* 102: 5612–5617, 2005.

Richmond BJ, Optican LM. Temporal encoding of two-dimensional patterns by single units in primate primary visual cortex. II. Information transmission. *J Neurophysiol* 64: 370–380, 1990.

Rolls ET. *Memory, Attention, and Decision-Making: A Unifying Computational Neuroscience Approach.* Oxford: Oxford University Press, 2008.

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

Rolls ET, Critchley H, Wakeman EA, Mason R. Responses of neurons in the primate taste cortex to the glutamate ion and to inosine 5′-monophosphate. *Physiol Behav* 59: 991–1000, 1996a.

Rolls ET, Critchley HD, Browning AS, Hernadi A, Lenard L. Responses to the sensory properties of fat of neurons in the primate orbitofrontal cortex. *J Neurosci* 19: 1532–1540, 1999.

Rolls ET, Critchley HD, Mason R, Wakeman EA. Orbitofrontal cortex neurons: role in olfactory and visual association learning. *J Neurophysiol* 75: 1970–1981, 1996b.

Rolls ET, Deco G. *Computational Neuroscience of Vision.* Oxford: Oxford University Press, 2002.

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

Rolls ET, Grabenhorst F. The orbitofrontal cortex and beyond: from affect to decision-making. *Progress Neurobiol* 86: 216–244, 2008.

Rolls ET, Grabenhorst F, Parris BA. Warm pleasant feelings in the brain. *NeuroImage* 41: 1504–1513, 2008.

Rolls ET, Kringelbach ML, de Araujo IET. Different representations of pleasant and unpleasant odors in the human brain. *Eur J Neurosci* 18: 695–703, 2003b.

Rolls ET, O'Doherty J, Kringelbach ML, Francis S, Bowtell R, McGlone F. Representations of pleasant and painful touch in the human orbitofrontal and cingulate cortices. *Cereb Cortex* 13: 308–317, 2003c.

Rolls ET, Treves A. *Neural Networks and Brain Function.* Oxford: Oxford University Press, 1998.

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

Rolls ET, Treves A, Tovee MJ, Panzeri S. Information in the neuronal representation of individual stimuli in the primate temporal visual cortex. *J Comput Neurosci* 4: 309–333, 1997b.

Romo R, Hernandez A, Zainos A. Neuronal correlates of a perceptual decision in ventral premotor cortex. *Neuron* 41: 165–173, 2004.

Shadlen MN, Newsome WT. Noise, neural codes and cortical organization. *Curr Opin Neurobiol* 4: 569–579, 1994.

Shannon CE. A mathematical theory of communication. AT&T Bell Laboratories Technical Journal 27: 379–423, 1948.

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

Tracey I, Becerra L, Chang I, Breiter H, Jenkins L, Borsook D, Gonzalez RG. Noxious hot and cold stimulation produce common patterns of brain activation in humans: a functional magnetic resonance imaging study. *Neurosci Lett* 288: 159–162, 2000.

Treves A, Panzeri S. The upward bias in measures of information derived from limited data samples. *Neural Comput* 7: 399–407, 1995.

Wilson JL, Jenkinson M, Araujo IET, Kringelbach ML, Rolls ET, Jezzard P. Fast, fully automated global and local magnetic field optimisation for fMRI of the human brain. *NeuroImage* 17: 967–976, 2002.

Witten IH, Frank E. *Data Mining: Practical Machine Learning Tools and Techniques.* San Francisco: Morgan Kaufmann, 2005.