Entropy and Cortical Activity: Information Theory and PET Findings

K. J. Friston,[1] C. D. Frith,[1] R. E. Passingham,[1,2] R. J. Dolan,[1] P. F. Liddle,[1] and R. S. J. Frackowiak[1]

[1] MRC Cyclotron Unit, Hammersmith Hospital, London W12 OHS, United Kingdom and
[2] Department of Experimental Psychology, Oxford University, Oxford OX1 3UD, United Kingdom

Functional segregation requires convergence and divergence of neuroanatomical connections. Furthermore, the nature of functional segregation suggests that (1) signals in convergent afferents are correlated and (2) signals in divergent efferents are uncorrelated. The aim of this article is to show that this arrangement can be predicted mathematically, using information theory and an idealized model of cortical processing.

In theory, the existence of bifurcating axons limits the number of independent output channels from any small cortical region, relative to the number of inputs. An information theoretic analysis of this special (high input:output ratio) constraint indicates that the maximal transfer of information between inputs, to a cortical region, and its outputs will occur when (1) extrinsic connectivity to the area is organized such that the entropy of neural activity in afferents is optimally low and (2) connectivity intrinsic to the region is arranged to maximize the entropy measured at the initial segments of projection neurons.

Under the constraints of the model, a low entropy is synonymous with high correlations between axonal firing rates (and vice versa). Consequently this antisymmetric arrangement of functional activity in convergent and divergent connections underlying functional segregation is exactly that predicted by the principle of maximum preservation of information, considered in the context of axonal bifurcation.

The hypothesis that firing in convergent afferents is correlated (has low entropy) and spatially coherent was tested using positron emission tomographic measurements of cortical synaptic function in man. This hypothesis was confirmed.

Certain patterns of cortical projections are so common that they could amount to rules of cortical connectivity. "These rules revolve around one, apparently, overriding strategy that the cerebral cortex uses—that of functional segregation" (Zeki, 1990). Functional segregation demands that cells with common functional properties be grouped together. This in turn necessitates both convergence and divergence of cortical connections. Anatomical convergence is required to assemble functionally distinct sets of signals, distributed over a functionally heterogeneous area, into a specialized area. Divergent efferents segregate and disseminate mixed signals to more specialized regions. Convergence is seen on many scales. For example, the connections between V1 and V5 are convergent in the sense that one V5 cell receives projections from many V1 cells, evidenced by the smaller areal extent of V5 compared to V1 and the larger receptive fields found in V5 (Zeki, 1971). Similarly, there is convergent input from V1 blobs (in which low spatial frequency and wavelength selectivity are represented) to the thin stripes of V2 in which cells have similar properties with larger receptive fields (Livingstone and Hubel, 1984) and from several thin stripes in V2 to V4 (Zeki and Shipp, 1989). As convergent projections assemble similar attributes of the visual field, it is inferred that firing rates in convergent afferents are correlated. Divergent connections, on the other hand, mediate redistribution of functionally distinct signals to different areas and subareas. Indeed, it was on the basis of the anatomical evidence for multiple and divergent projections from V1 to extrastriate areas that the role of V1 as a functional segregator was first proposed (Zeki, 1975). As divergent connections parcel out functionally different signals to various extrastriate regions, it is concluded that firing rates in divergent efferents are largely uncorrelated. The capacity to decorrelate outputs is considered by some to be a central component of feature detection (e.g., Foldiak, 1989; Oja, 1989; Hornik and Kuan, 1992).

Alternative arrangements are unlikely on the grounds that they would preclude categorization of events in the sensory field. In general, uncorrelated signals in convergent connections would confound independent features and disallow any subsequent separate categorization on the basis of those features. In most processes with a biological flavor (e.g., Pois-

son point processes and stationary Gaussian processes), conflating two dissimilar inputs (A and B) is irreversible, where categorizaton can proceed on the basis of A or B but not on the basis of A and B alone. Similarly, correlated signals in divergent efferents would lead to a complete failure in segregating a functionally mixed input and a potential failure to extract features necessary for categorization.

Functional segregation therefore suggests two antisymmetric features of cortical organization: (1) signals in convergent afferents are correlated, and (2) signals in divergent efferents are uncorrelated. The aim of this article is to show that this arrangement and its conceptual counterpart—functional segregation—are exactly consistent with a simple formulation of cortical processing in information theoretical terms.

By considering a particular constraint imposed by axonal bifurcation, we demonstrate that information transfer is optimized when inputs to a cortical region are substantially correlated (have an optimal and low entropy) and outputs are uncorrelated (have a high entropy). We describe the theory below and provide empirical evidence of autocorrelated, coherent afferent activity in the cortex using PET measurements of neurophysiology.

Theory

The behavior of any cortical region, of small arbitrary diameter, is characterized by its inputs (afferents), outputs (efferents), and the transformation of neural discharge patterns between the two. Using the terminology of Shepherd and Koch (1990), we define inputs as extrinsic axons (arising from distant cells) giving rise to arborizations that synapse on cell processes within the region. Outputs are the single output points (initial segments) of projection (principal or relay) cells giving rise to at least one extrinsic axon. Extrinsic connections connect distant cortical regions, as distinct from intrinsic connections (e.g., interneurons or recurrent axonal collaterals).

This definition of an output is strictly neuroanatomical, but functional independence is implicit. Two axons deriving from the same initial segment are considered to be part of the same output and exhibit the same pattern of firing. Only different outputs can fire independently.

The synaptic, parasynaptic, and ephaptic transformation of discharge patterns is effected by direct connections between extrinsic afferents and projection cells (e.g., direct axodendritic synapses on the apical dendrites of large pyramidal cells), by intrinsic connections (e.g., stellate interneurons), and by intrinsic recurrent collaterals (Mountcastle, 1978; Powell, 1981).

Axonal Bifurcation and Constraints on Extrinsic Connectivity

One special aspect of brain connectivity, central to the argument developed below, is that the number of inputs to a cortical region exceeds the number of outputs. This characteristic is the main constraint under which the optimization of information transfer is considered. It is self-evident that the number of inputs to a single neuron (multiple dendritic synapses) exceeds the number of outputs (single initial segment) from that neuron. It is also self-evident, but perhaps not so obvious, that the number of inputs to gray matter area will, on average, be substantially greater than the number of outputs. This is a consequence of bifurcating or nonrecurrent axonal collaterals.

Imagine a closed surface or boundary at all gray–white matter interfaces. The number of axonal fibers crossing that surface into white matter is less than or equal to the number crossing in the opposite direction—the gray matter inputs (the possible inequality results from axonal bifurcation in the white matter volume). Because one initial segment (output) can give rise to several extrinsic axons, the number of outputs is less than the number of fibers entering white matter, which in turn is less than or equal to inputs traversing the boundary in the other direction. Therefore, the number of outputs is less than the number of gray matter inputs. In general, if an efferent axon bifurcates on average n times there will be $n + 1$ axonal fibers for each output. Given that all extrinsic afferent fibers represent an input, the input:output ratio would be $(n + 1):1$. This argument assumes that the difference between effectors and receptors is small in comparison to the total number of extrinsic axons.

Double labeling experiments have demonstrated that the Meynert cells of layer 6 in V1 project through bifurcating axons to both V5 and the superior colliculus (Fries et al., 1985). Nonrecurrent axonal collaterals (e.g., bifurcating axons) can mediate backward connections. Studies of axonal bifurcation show that, in general, backward projections are less submodality specific than outward projections (Bullier and Kennedy, 1987; Shipp and Zeki, 1989a, b). Because backward projections are common, axonal bifurcation may be ubiquitous.

Examples of High Input:Output Ratios

V5 receives convergent afferents from V1, shown by the fact that V5 receptive fields are larger than the V1 receptive fields of which they are composed (Zeki, 1971). Consequently, a single V5 neuron receives axonal afferents from more than one V1 cell. The number of initial segments on V5 projection cells cannot exceed the number of V5 cells. Therefore, the number of inputs to V5 exceeds its outputs. Examples can be found where the reduction in outputs is high. Each of the A laminae of the cat's lateral geniculate nucleus (LGN) contains roughly 400,000 cells, of which about 300,000 are projection cells. The LGN receives slightly fewer than 100,000 retinogeniculate axons and more than 4,000,000 corticogeniculate axons, in addition to afferents from the brainstem reticular formation and the reticular nucleus of the thalamus (Sherman and Koch, 1990). The input:output ratio is, in this example, about 14:1. Note that this output reduction is largely due to backward projections from the cortex.

Entropy and Correlations

The application of information theory often takes the form of optimizing a particular aspect of performance under a series of constraints. In what follows, the transfer of information associated with the transformation of neural firing by a small region of cortex is the object of optimization. In the context of brainlike function, the principle of maximum information preservation has an intuitive, predictive, and construct validity (Linsker, 1988; Foldiak, 1990) and is related to the concept of redundancy reduction (Barlow, 1961; Atick and Redlich, 1990). Linsker has coined the term *infomax* in reference to this principle and has discussed its ramifications and precedents (Linsker, 1988).

The main constraint under which the principle of information preservation is developed is the reduction or constriction of outputs relative to inputs. A heuristic argument suggests that if information in an input space is redistributed over a substantially smaller number of outputs, information transfer will be enhanced by mutual predictability in the inputs. This implies that firing in one input is predictive of (and predicted by) activity in the remainder; that is inputs will be correlated. Conversely, efficient use of (noiseless) outputs prohibits mutual prediction and requires the outputs to be independent or uncorrelated.

Put another way, if there is a constriction in the number of available channels, transfer will be more efficient if less information tries to get through at once. These heuristic arguments can be illustrated more formally using information theory: information (I) is the improbability of an event (x) expressed as the logarithm of the inverse of its probability [$p(x)$], or

$$I(x) = -\ln(p(x)). \tag{1}$$

An event with low information is highly probable, and its occurrence could have been predicted. For a number of continuous events (neural activity across several axons, x) the average information is referred to as entropy [$H(x)$]. For an n-dimensional space where the probability density of axonal firing is Gaussian, the entropy is given by (Jones, 1979)

$$H(x) = \ln((2\pi e)^n \cdot |\rho x|)/2, \tag{2}$$

where ρx is the covariance matrix (and $|\rho x|$ its determinant) describing the covariance between neural firing. Following transformation, in accord with the principle of information preservation, the entropy (average information) of the outputs should be high. The model of cortical processing we use is defined by the following assumptions: (1) stationary multivariate Gaussian continuous input with zero mean and unit variance, (2) additive uncorrelated orthogonal Gaussian noise in, and only in, the inputs, and (3) linear transformations under the constraint that $\Sigma^i c_{ij}^2 = 1$, where c_{ij} is the coefficient scaling the contribution from input i to output j. The last constraint conserves total synaptic contacts a dendritic tree can express. For example, if synaptic efficacy is proportional to the radii of postsynaptic specializations, then the total areal extent of all specializations on one tree is unity. For this idealized model, the entropy of the outputs [$H(y)$] is arithmetically related to the mutual information between inputs and outputs [$I(x, y)$]. The mutual information reflects the information that y conveys about x (Jones, 1979):

$$I(x, y) = H(y) - H(z), \tag{3}$$

where $H(z)$ is the entropy of noise in the inputs. For unchanging noise characteristics, an increase in $H(y)$ is equivalent to an increase in the mutual information. Optimizing information transfer thus reduces to maximizing $H(y)$ under the constraints imposed (1) by the model assumptions and (2) by a high input:output ratio.

The linear transformation that maximizes output entropy (under the above assumptions) is a principal component transformation (Linsker, 1988; Foldiak, 1989; Oja, 1989; Hornik and Kuan, 1992). This transformation renders the outputs orthogonal or independent (not mutually predictive). For a more detailed analysis of the role of noise under less restrictive constraints than those assumed here, see Linsker (1988) and Atick and Redlich (1990). We use the principal component transformation to examine how the upper limit on $H(y) = H_{max}$ depends on the covariance structure of the inputs (Cx) and the input:output ratio (n:m).

This dependency is illustrated in Figure 1 by modeling the input covariance matrix as a Gaussian autocovariance matrix:

$$\rho x(i,j;\ i - j = b) = \exp(-b^2/(2\beta^2)). \tag{4}$$

As β gets bigger, the covariance between distant inputs increases and the inputs exhibit a greater degree of intercorrelation or coherence. β will be referred to as coherence. The upper limit on $H(y)$ is given by Equation 2 where $|\rho x| = \Pi\epsilon_i$ and ϵ_i are the m largest eigenvalues of ρx. It is immediately obvious, from Figure 1, that as soon as an initially incoherent input becomes more coherent, H_{max} increases, therein potentially optimizing information transfer. This effect is more sustained with higher input:output ratios. The upper solid line in Figure 1 corresponds to a ratio of 2:1 (on average extrinsic axons from projection cells bifurcate once). In this case, the optimal input coherence is realized fairly soon. Subsequent increases result in a progressive reduction in output entropy as the capacity to support high variances in the large number of orthogonal outputs fails. For higher ratios, the optimal coherence (largest H_{max}) is much greater. This behavior is a general feature of all monotonic decreasing autocovariance functions we have examined. Figure 2 illustrates the generality of this behavior by relating H_{max} to the entropy of the inputs for a number of autocovariance functions. Initially as input entropy falls (coherence increases), H_{max} increases to a maximum and then declines monotonically. The location of the maxima depends on the nature of the autocovariance function and input:output functions; however, all are associated with nontrivially low entropies (nonzero coherence). The three autocovari-

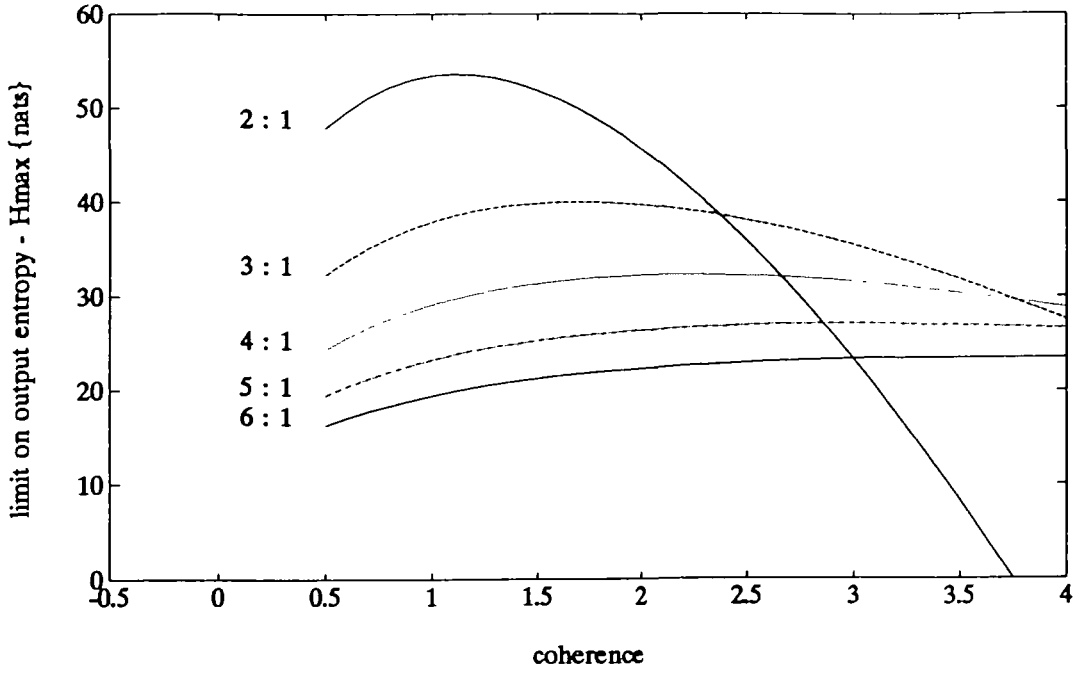

Figure 1. Dependency of upper limit on output entropy (H_{max}) on coherence (standard deviation of a Gaussian autocovariance function) of a stationary input. This relationship is shown for five cases of increasing input:output ($n:m$) ratio with $n = 64$ inputs. $H_{max} = \ln[(2\pi e)^m \cdot \Pi\epsilon_i]/2$, where ϵ_i are the m largest eigenvalues of $\rho x(i,j: i - j = h) = \exp(-h^2/(2\beta^2))$.

ance functions depicted in Figure 2 are a Gaussian (Eq. 4), an exponential $\rho x(h) = \exp(-h/\beta)|_{\beta>0}$, and a hyperbolic function $\rho x(h) = (h + 1)^{\beta}|_{\beta<0}$ (see Fig. 2 caption for the ranges of β used).

The use of autocovariance matrices assumes that the input covariance pattern [$\rho x(h)$] is stationary; cross-covariances are a function of distance between inputs (h), not specific locations. This is appropriate given that we are modeling an organizational principle that is invariant over the entire cortex.

In conclusion, there is an optimal coherence that maximizes the entropy of the outputs for any input: output ratio. In other words, given the above constraints, the most informative and balanced neural activity in a series of independent, uncorrelated outputs from an area is associated with low-optimal-entropy, correlated activity in the inputs. This is exactly the arrangement predicted by functional segregation: (1) correlated activity in convergent afferents and (2) uncorrelated activity in divergent efferents. Hebb's rule can be seen as satisfying a special case of this arrangement—where the input entropy is unspecified and the cortical area in question reduces to a single dendritic tree. Oja (1982) has analyzed a model with a single output unit using a local Hebbian connection strength modification rule and demonstrated that the unit extracts the principal component with the largest eigenvalue (Ω) from a stationary input. The entropy [$H(y)$] of a unidimensional Gaussian distribution is given by (Jones, 1979) $H(y) = \log(2\pi e\Omega)/2$. Consequently, output entropy is maximized. Note the antisymmetric nature of this conclusion: to comply with the principle of maximal information preservation—in this context maximization of output entropy—there is an almost paradoxical requirement that the entropy of the inputs be significantly less than chance expectation.

Examples of Correlated Inputs

"The EEG is both a consequence and a sign of correlated activity in the brain" (Cook, 1991). If neurons converging onto cortex all fired independently, then the effects on the electrical field outside the cranium would largely cancel. High- and low-frequency field potential changes imply a strong local correlation.

Example of Independent Outputs

There are divergent projections from V1 to V5 and V4. Independent activation of V5 and V4 has been demonstrated in human subjects using functional imaging (Zeki et al., 1991). Therefore, firing in V1 efferents projecting to V5 can be independent of firing in projections to V4.

Empirical Validation

We predicted that convergent afferentation in the cortex will have a low entropy over a range of spatial domains. This prediction can be reformulated in terms of the instantaneous measurement of neural activity, predicted to have nontrivial autocovariance, or to be spatially coherent.

There is a substantial amount of empirical evidence to suggest that regional cerebral blood flow (rCBF) is coupled to neural discharge rates in cortical afferents (e.g., Fox and Raichle, 1986). Compelling evidence that this coupling operates over small spatiotemporal domains is provided by high-resolution optical imaging of microcirculatory events in ocular dominance columns of monkey cortex during visual stimulation (Frostig et al., 1990). This and other ev-

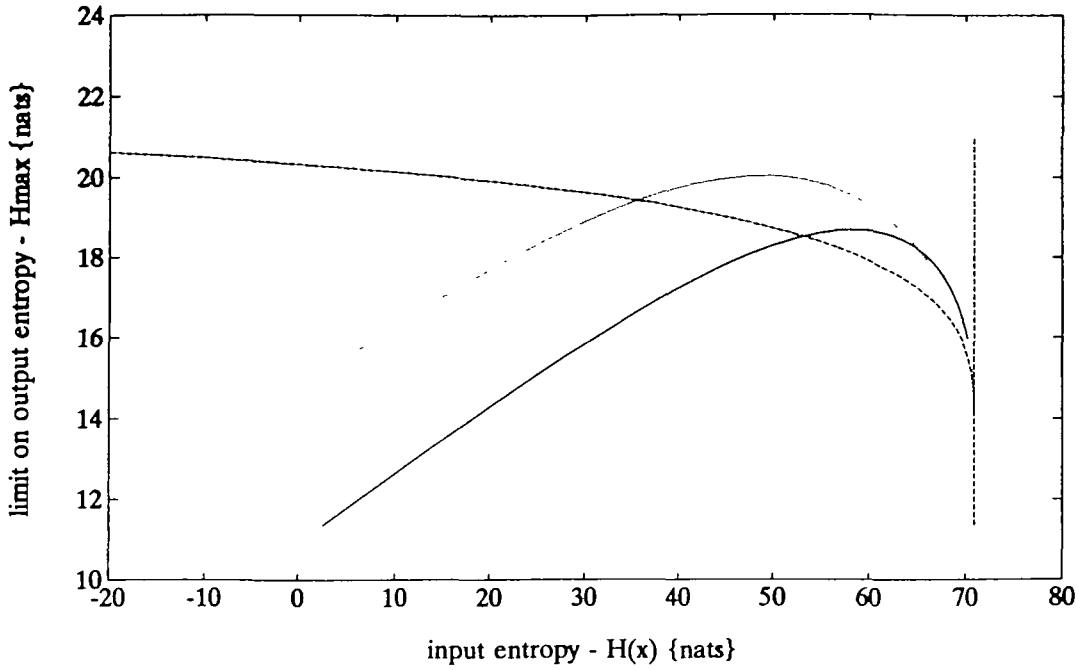

Figure 2. Similar to Figure 1, but the input coherence is expressed as entropy [$H(x)$] and the input:output ratio ($n:m$) is 5:1 ($n = 50$, $m = 10$). The *three lines* correspond to three autocovariance functions of β: (1) Gaussian [as in Fig 1, $\exp(-h^2/(2\beta^2))$, $\beta = 0$–2; *dashed line*], (2) exponential [$\exp(-h/\beta)$, $\beta = 1$–41; *dotted line*] and (3) hyperbolic [$(h + 1)^\beta$, $\beta = -2.5$–0; *solid line*] All three cases demonstrate that as input entropy falls from incoherence [$H(x) = 70.95$] output entropy increases (until a maximum is reached).

idence (Conrad and Klingelhofer, 1989) suggests that rCBF is coupled to afferent activity over a scale of less than 1 mm and less than 1 sec. rCBF is therefore a neurophysiological index of afferent activity that we predicted would show nontrivial autocovariance over many millimeters.

The spatial frequencies of multiple realizations of stationary processes remain unchanged when integrated. Consequently, the coherence or autocovariance of the stationary component of many instantaneous rCBF measurements integrated over time will be the same as any single measurement in isolation. We use this to advantage in the PET technique, which gives the integrated rCBF over several minutes.

Clearly, actual cortical activity may have many nonstationary components, reflecting regional specificity of function. These nonstationary components are not the subject of the present analysis. As a first step, we wished to demonstrate coherence as a ubiquitous and regionally invariant characteristic of cortical activity. Provisional work (K. J. Friston, C. D. Frith, R. E. Passingham, R. J. Dolan, P. F. Liddle, and R. S. J. Frackowiak, unpublished observations) using statistical tests of sphericity, suggests that it is possible to measure regional differences in entropy.

The spatial resolution of PET is poor, but in a well-behaved way (Glick et al., 1989). The confounding effect of poor resolution on estimating autocovariance (coherence) can be accounted for by deconvolution with the noise power spectrum (NPS).

Methods

To remove systematic and nonstationary neurophysiological components, due to regional variations in perfusion and anatomical configuration of gyri, we used the difference in activity between two measurements of cortical rCBF to estimate its autocovariance. Experimental control was exerted over the physiological differences by using two tasks repeated in a pairwise fashion. The choice of these tasks was arbitrary from the point of view of the present analysis, as regional differences were not an issue. The tasks were chosen because they activate extensive and widespread cortical regions (Frith et al., 1991).

Data Acquisition

Six normal male volunteers where scanned 12 times in the same session while performing one of two tasks in an alternating sequence (repeating a heard letter and responding with a word that began with a heard letter). Similarly, the standard (Hoffman) three-dimensional human brain phantom was scanned 12 times using ^{18}F at a concentration of 0.6 μCi/cc in the gray matter compartment. The total counts per image corresponded to the human studies. Permission to perform these studies was obtained from the local ethical committee and Advisory Committee for the Administration of Radioactive Substances of the UK. Scans were obtained with a CTI (model 953B; CTI, Knoxville, TN) PET camera as a fully three-dimensional acquisition. Reconstructed (Townsend et al., 1992) images had a resolution of 5.2 mm (T. J. Spinks, T. Jones, D. L. Bailey, D. W. Townsend, S. Grootnook, P M. Bloomfield, M. C. Galardi, M. E. Casey, B. Sipe, and J. Reed, unpublished observations). The volume images contained 128 × 128 × 31 voxels corresponding to 2 × 2 × 3.1 mm. ^{15}O was administered intravenously as radiolabeled water infused over 2 min.

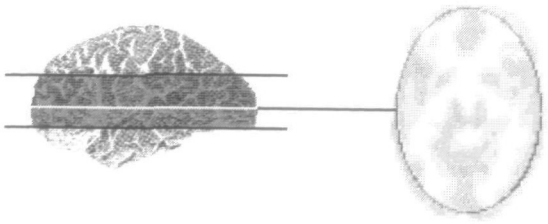

Figure 3. Picture of the brain showing the extent of cortical surface analyzed. An example of a fitted cortical ellipse is superimposed on a transverse section.

The total counts per voxel during the buildup phase of radioactivity served as an estimate of rCBF (Fox and Mintun, 1989). The tasks began 20 sec prior to delivery of radiolabeled water.

Data Analysis

The cortical rim was sampled from the 12 scans from each subject and the phantom. This sampling used an ellipse fitted to the length and width of 20 consecutive slices (see Fig. 3). Subtracted sequential pairs (Kijewski and Judy, 1987) were used to estimate the rCBF autocovariance function.

The objective of our analysis was to show that the human data, but not the phantom data, exhibited nontrivial autocovariance over many millimeters. Following normalization to zero mean and unit variance, the one dimensional subtracted rCBF data were subject to Fast Fourier transformation. The transformed rCBF data were averaged across all cortical ellipses from one subject and divided by the NPS in frequency space. This corresponds to deconvolution in Cartesian space. The resulting spectral density functions can be seen in Figure 4*a* for the six subjects and the phantom data, included for comparison. Inverse Fourier transformation of the spectral density functions yielded the corresponding autocovariance functions (Cox and Miller, 1980), which because of the initial normalization are also the autocorrelation functions (Fig. 4*b*).

We used an empirical estimate of the NPS (polynomial fit of the Fourier transform of subtracted sequential phantom pairs) to ensure a proper estimate of low spatial frequencies. This accounted for two-dimensional aliasing due to pixel sampling (Kijewski and Judy, 1987).

Results

Figure 4*a* shows the normalized spectral density functions for the phantom data (solid line circled at the beginning) and the six subjects. The phantom data receive equal contributions from all frequencies, consistent with uncorrelated white noise these data represent. In contrast, the physiological data have relatively large contributions from low spatial frequencies that result in monotonic declining autocorrelations at increasing distances. This coherence is seen in the corresponding autocorrelation functions in Figure 4*b*. Again, the phantom autocorrelation function is given for comparison and is rendered as a solid line. Autocorrelation is evident (in this conservative analysis; see below) at 5 mm and beyond. This coherence cannot be explained by intrinsic connectivity, which has a maximal spatial extent of 3 mm (Mountcastle, 1978).

Although the autocovariance is significantly greater than 0 over large distances [e.g., mean $\rho(8\text{ mm}) = 0.0644$; $t = 6.47$; $p < 0.001$; df = 5), the size of the autocorrelations is very small. This is because the rCBF process is embedded in large amounts of uncorrelated noise associated with the high-resolution PET technique used. Figure 4*b* illustrates this point in that the low-frequency structure in the rCBF data "sits on top of" a substantial amount of white noise distributed uniformly over all frequencies. We chose to use a high-resolution technique in order to lend a stereotactic validity to our sampling of the cortex. It is possible to extend the analysis and partition the observed process into an uncorrelated noise process and a residual correlated process corresponding to the underlying rCBF differences. However, the results presented above are sufficient to confirm the hypothesis that afferent cortical activity exhibits autocovariance.

Discussion

We have examined the implications that functional segregation holds for the arrangement of activity in convergent afferents and divergent efferents. The arrangement implied by functional segregation is precisely that predicted theoretically using ideas from information theory. This theoretical analysis led to a hypothesis about the spatial autocorrelations of activity in convergent afferents in the cortex. We designed an experiment to test and confirm this hypothesis.

In analogy with the application of information theory to continuous channels (e.g., Shannon's continuous channel theorem), we have taken an idealized model and reduced the problem to one of finding the largest $H(y)$ subject to certain constraints (e.g., when the input is *power limited*; see Jones, 1979). The constraint of particular interest in this formulation derives not from the role of noise (Atick and Redlich, 1990) or available power in the inputs but from the topographic arrangement of the connections, namely, a reduction in the number of independent channels at each cortical transformation. Axonal bifurcation has been used as empirical support for this constraint, which implies that the dimensionality of inputs to a cortical region exceeds that of the outputs. Inputs are defined as extrinsic axonal afferents to the region. Outputs are the initial segments of projection cells. This information theoretic approach suggests that a low optimal input entropy maximizes information transfer. Firing rates measured in afferent axons will be of low entropy when those firing rates are correlated or coherent. This would mean, at a symbolic level, that events captured by neural firing map close together if and only if those events were in some way mutually predictive. In other words, probabilistic regularities or invariances in extrapersonal space would be embodied in the spatial convergence of neural projections.

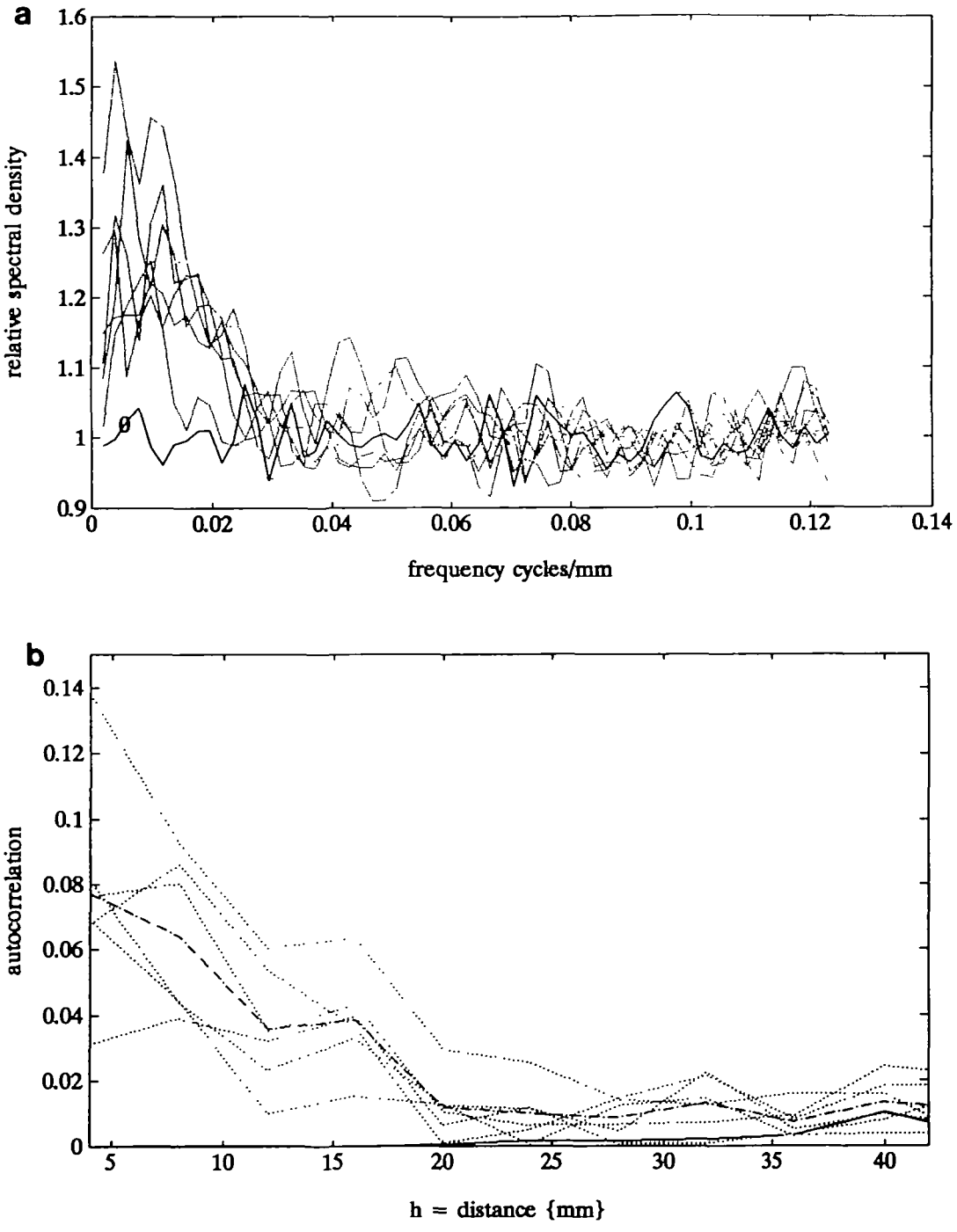

Figure 4. *a*, Spectral density functions for phantom data (*solid line circled at beginning*) and six subjects (*dotted lines*). The spectral density functions have been normalized to have a mean density of 1 in spatial frequency bins greater than 0.05 cycles/mm. These functions were derived from a series of cortical ellipses (20 from each subject) at least 128 pixels = 256 mm in length. The density functions have been resolution corrected by deconvolution with an empirical estimate of the NPS. The central finding is that only the physiological data have large amounts of slowly undulating low spatial frequencies. *b*, the same data but presented as autocorrelation functions following inverse Fourier transform of the spectral density functions. Autocorrelation for the phantom data (*solid line*) is near zero at all distances. In contrast, the rCBF data (*dotted lines*; *dashed line* = mean) show nonzero autocorrelations up to 20 mm.

Spatial coherence was measured in terms of the autocorrelation of a series of subtracted rCBF measurements in the cortex. The equivalence between cross-correlations over time and the autocorrelation of a single observation over space depends on the assumption of stationariness. As predicted, autocorrelation was evident over extensive (10–20 mm) domains.

Massive divergence and convergence are a necessary consequence of these entropic considerations. As extrinsic efferents (and their collaterals) are uncorrelated (high entropy), it is unlikely they will terminate in the same cortical area. This implies that efferents are divergent. Divergence implies convergence, and both are features of cortical organization (Mesulam, 1990; Zeki, 1990). The classification of

convergent versus divergent connections is purely a matter of where the connections being described are referred. A cortical focus can receive convergent projections from extensive and different areas and subareas but can only give rise to divergent connections. No single set of axons can be both convergent with reference to one point and divergent with respect to another. This is important given the assertion that these two classes have opposite entropic tendencies.

The transformation effected by intrinsic connectivity is assumed to result in a decorrelation of outputs (Foldiak 1989). This decorrelation results in a high entropy. A high entropy is characteristic of "functional segregators." Each millimeter of V1 contains all the visual information from a particular retinal point that is destined for the cortex. Since it is difficult to imagine that the same signals are relayed in the divergent and parallel projections, V1 was proposed to act as a functional segregator (Zeki, 1975). Both a convergence of correlated activity and a divergence of uncorrelated activity are implicit. Correlated visual motion information converges on V5 from widely distributed V1 efferents. Conversely, uncorrelated submodality information (e.g., motion, color, depth) is divergently redistributed to functionally specialized areas from V1. Within a grossly homogeneous functional area (e.g., V5), the divergent outputs should be uncorrelated and correspond to a finer but still orthogonal segregation of submodality information, for example, a segregation into the direction and magnitude components of the velocity vector. There is evidence for speed- and direction-invariant cells in V5 (Zeki, 1990).

Mechanisms

Principal component transformation has been used to estimate the theoretical limit on entropy over m outputs given n inputs and their covariance matrix. This does not imply that intrinsic connectivity performs a principal component transformation. However, there are specific proposals that finding the principal component space is an important theme in feature detection (Linsker, 1988; Oja, 1989; Foldiak, 1990; Rubner and Schulten, 1990). Foldiak has described a (anti-)Hebbian mechanism that effects a transformation of a high (n)-dimensional input into a lower (m)-dimensional output that spans the same subspace as the m-largest principal components of the input (Foldiak, 1989, 1990). In addition to Hebbian feedforward connections, Foldiak's model depends on anti-Hebbian feedback connections between the output "neuron-like units" to keep the outputs uncorrelated (orthogonal). Durbin and Mitchison (1990) have used cortical wiring length constraints to model connectivity in the primary visual cortex (V1). They present simulations that are remarkably reminiscent of empirically determined configurations. One of the basic tenets of their approach is the requirement that contiguous regions of an external parameter space (e.g., position in the visual field) should be represented close together in the cortical sheet The spatiotemporal contiguity of real events would confer coherence (minimize the entropy of discharge rates measured in afferent fibers impinging on contiguous regions in V1) in retinotopic maps if and only if there is a topographic preservation of real space–time contiguity relationships of the sort they suggest.

From the point of view of the theory of neuronal group selection (Edelman, 1978, 1987), the organizational tendencies discussed above may be relevant at the level of the selective expression of the secondary repertoire. The convergence of temporally correlated inputs onto the same dendritic tree can be envisioned in terms of synaptic consolidation, which depends on synaptic discharge and changes in transmembrane potential. This is because the behavior of each axonal input is accessible to the remaining inputs by (electrotonic) communication through the dendritic processes they all share. We have, however, discussed coherence in terms of spatial domains, which are far more extensive than a single dendritic tree. In this case, a different mechanism must be postulated that does not depend on convergence onto one neuron. Such a mechanism (based on nitric oxide) has been proposed (Gally et al., 1990; Montague et al., 1991). The effects of a short-lived, rapidly diffusible signal on local synaptic plasticity have been simulated and shown to be able to link the activity in a local volume of tissue, regardless of whether the neurons are directly connected by synapses.

Notes

K.J.F. was funded by the Wellcome Trust. We thank Peter Foldiak and the Fellows of the Neurosciences Institute for inspiring discussions, and Semir Zeki for invaluable guidance during the development of this work.

Correspondence should be addressed to Dr. Karl J. Friston, The Neurosciences Institute, 1230 York Avenue, New York, NY 10021.

References

Atick JJ, Redich AN (1990) Towards a theory of early visual processing. Neural Comput 2:308–320.

Barlow, HB (1961) Possible principles underlying the transformation of sensory messages. In: Sensory communication (Rosenblith WA, ed). Cambridge, MA: MIT Press.

Bullier J, Kennedy H (1987) Axonal bifurcation in the visual systems. Trends Neurosci 10:205–210.

Conrad R, Klingelhofer J (1989) Dynamics of regional cerebral blood flow for various visual stimuli. Exp Brain Res 77:437–441.

Cook JE (1991) Correlated activity in the CNS: a role on every timescale? Trends Neurosci 14:397–401.

Cox DR, Miller HD (1980) The theory of stochastic processes, pp 272–336. New York. Chapman and Hall.

Durbin R, Mitchison G (1990) A dimension reduction framework for understanding cortical maps. Nature 343: 644–647.

Edelman GM (1978) Group selection and phasic re-entrant signalling: a theory of higher brain function. In The mindful brain (Edelman GM, Mountcastle VB, eds), pp 55–100. Cambridge, MA: MIT Press

Edelman GM (1987) Neural Darwinism, the theory of neuronal group selection. New York: Basic.

Foldiak P (1989) Adaptive network for optimal linear feature extraction. In: Proceedings of the IEEE/INNS joint conference on neural networks, pp 401–405 New York: IEEE

Foldiak P (1990) Forming sparse representations by local anti-Hebbian learning. Biol Cybern 64:165–170.
Fox PT, Mintun MA (1989) Non-invasive functional brain mapping by change distribution analysis of averaged PET images of $H_2{}^{15}O$ tissue activity. J Nucl Med 30:141–149.
Fox PT, Raichle ME (1986) Focal physiological uncoupling of cerebral blood flow and oxidative metabolism during somatosensory stimulation in human subjects. Proc Natl Acad Sci USA 83:1140–1144.
Fries W, Keizer K, Kuyprs HGJM (1985) Large layer VI cells in macaque striate cortex (Meynert cells) project to both superior colliculus and prestriate visual area V5. Exp Brain Res 58:613–616
Frith CD, Friston KJ, Liddle PF, Frackowiak RSJ (1991) Willed action and the prefrontal cortex in man. Proc R Soc Lond [Biol] 244.241–246.
Frostig RD, Lieke EE, Ts'o DY, Grinvald A (1990) Cortical functional architecture and local coupling between neuronal activity and the microcirculation revealed by *in vivo high-resolution optical imaging of intrinsic signals*. Proc Natl Acad Sci USA 87:6082–6086
Gally JA, Montague PR, Reeke GN, Edelman GM (1990) The NO hypothesis: possible effects of a short lived, rapidly diffusible signal in the development and function of the nervous system Proc Natl Acad Sci USA 87:3547–3551.
Glick SJ, King MA, Penny BC (1989) Characterization of the modulation transfer function of discrete filtered backprojection. IEEE Trans Med Imaging 8:203–213.
Hornik K, Kuan CM (1992) Convergence analysis of local feature extraction algorithms. Neural Networks 5:229–240.
Jones DS (1979) Elementary information theory, p 152. Oxford: Clarendon.
Kijewski MF, Judy PF (1987) The noise power spectrum of CT images Phys Med Biol 32:565–575.
Linsker R (1988) Self organization in a perceptual network Computer 21:105–117
Livingstone MS, Hubel DH (1984) Anatomy and physiology of a color system in the primate visual cortex. J Neurosci 4:309–356.
Mesulam MM (1990) Large scale neurocognitive networks and distributed processing for attention language and memory. Ann Neurol 28:597–613.
Montague PR, Gally JA, Edelman GM (1991) Spatial signalling in the development and function of neural connections. Cereb Cortex 1:199–220
Mountcastle VB (1978) The mindful brain, pp 7–51 Cambridge: MIT Press.
Oja E (1982) A simplified neuron model as a principal component analyzer J Math Biol 15:267–273.
Oja E (1989) Neural networks, principal components, and subspaces. Int J Neural Systems 1:61–68.
Powell TPS (1981) Certain aspects of the intrinsic organization of the cerebral cortex. In. Brain metabolism and perceptual awareness (Pompeiano O, Aimone Marsan C, eds), pp 1–19. New York: Raven.
Rubner J, Schulten K (1990) Development of feature detectors by self organization: a network model. Biol Cybern 62:193–199.
Shepherd G, Koch C (1990) Introduction to synaptic circuits. In: The synaptic organization of the brain (Shepherd GM, ed), p 3. London: Oxford UP
Sherman, SM, Koch C (1990) Thalamus. In: The synaptic organization of the brain (Shepherd GM, ed), p 246 London: Oxford UP.
Shipp S, Zeki S (1989a) The organization of connections between areas V5 and V1 in the macaque monkey visual cortex Eur J Neurosci 1:309–332.
Shipp S, Zeki S (1989b) The organization of connections between areas V5 and V2 in the macaque monkey visual cortex. Eur J Neurosci 1:333–354
Townsend DW, Geissbuhler A, Defrise M, Hoffman EJ, Spinks TJ, Bailey D, Gilardi MC, Jones T (1992) Fully three-dimensional reconstruction for a PET camera with retractable septa. IEEE Trans Med Imaging, in press.
Zeki S (1971) Convergent input from the striate cortex (area 17) to the cortex of the superior temporal sulcus in the rhesus monkey. Brain Res 28:338–340.
Zeki S (1975) The functional organizations of projections from striate to prestriate visual cortex in the rhesus monkey. Cold Spring Harbor Symp Quant Biol 40:591–600.
Zeki S (1990) The motion pathways of the visual cortex. In: Vision: coding and efficiency (Blakemore C, ed), pp 321–345. Cambridge: Cambridge UP.
Zeki S, Shipp S (1989) Modular connections between areas V2 and V4 of the macaque monkey visual cortex. Eur J Neurosci 1:494–506.
Zeki S, Watson J, Lueck C, Friston KJ, Kennard C, Frackowiak RSJ (1991) A direct demonstration of functional specialization in human visual cortex. J Neurosci 11:641–649.