

Functional and Effective Connectivity in Neuroimaging: A Synthesis

Karl J. Friston

The MRC Cyclotron Unit, Hammersmith Hospital, London, England

Abstract: The brain appears to adhere to two principles of functional organization; *functional segregation* and *functional integration*. The integration within and between functionally specialized areas is mediated by *functional* or *effective connectivity*. The characterization of this sort of connectivity is an important theme in many areas of neuroscience. This article presents one approach that has been used in functional imaging.

This article reviews the basic distinction between functional and effective connectivity (as the terms are used in neuroimaging) and their role in addressing several aspects of functional organization (e.g. the topography of distributed systems, integration between cortical areas, time-dependent changes in connectivity and nonlinear interactions). Emphasis is placed on the points of contact between the apparently diverse applications of these concepts and in particular the central role of eigenimages or spatial modes. Although the framework that has been developed is inherently linear, it has been extended to assess nonlinear interactions among cortical areas. © 1994 Wiley-Liss, Inc.

Key words: functional connectivity, effective connectivity, PET, fMRI, eigenimages, spatial modes, multidimensional scaling, word generation, visual, modulation

INTRODUCTION

In the past decade functional neuroimaging has been extremely successful in establishing functional segregation as a principal of organization in the human brain. Functional segregation is usually inferred by the presence of activation foci in change score or statistical parametric maps. The nature of the functional specialization is then attributed to the sensorimotor or cognitive process manipulated experimentally. Newer approaches have addressed the integration of functionally specialized areas through characterizing neurophysiological activations in terms of distributed changes [e.g., Friston et al., 1993a,b; La-

greze et al., 1993; McIntosh et al., 1993; McIntosh and Gonzalez-Lima, 1991; Horwitz et al., 1991; Horwitz, 1990; Moeller et al., 1987]. We have published a series of papers along these lines that have introduced a number of concepts (e.g., functional and effective connectivity, eigenimages, spatial modes, information theory, multidimensional scaling) and their application to issues in imaging neuroscience (e.g., functional systems, cortical integration, associative plasticity, and nonlinear cortical interactions) [Friston et al., 1993a,b,c, 1994a,b,c,d]. The aim of this paper is to describe a framework that highlights the (usually simple) relationships among these concepts and their application to neuroimaging data.

In what follows, intuitive ideas about the integration of brain activity are formulated in terms of simple mathematical expressions. Many of the techniques presented have their origins and counterparts in the analysis of multiunit electrode recording data and

Received for publication May 5, 1994; revision accepted September 26, 1994.

Address reprint requests to Dr. K.J. Friston, The MRC Cyclotron Unit, Hammersmith Hospital, Du Cane Road, London W12 0HS, UK.

multichannel EEG (and MEG) data. An effort has been made to clarify these points of contact by referring to strategies adopted in these other modalities. The paper begins with a brief review of two data sets that are used to illustrate the techniques. The PET data were obtained during an activation study of intrinsic work generation. The fMRI data came from a single-subject photic stimulation study. The second section deals with the fundamental distinction between functional connectivity (*temporal correlations between remote neurophysiological events*) and effective connectivity (*the influence one neural system exerts over another*) and their relationship to similar concepts in electrophysiology. The remainder of the paper is divided into two parts which focus on 1) functional connectivity and 2) effective connectivity.

The discussion of functional connectivity starts with a description of how patterns of activity in the brain can be measured and how this formulation points to eigenimages (spatial modes) as a powerful characterization of distributed systems. Eigenimages are usually identified using singular value decomposition (SVD) or related techniques. SVD is introduced and explained. Eigenimages represent a mapping of *function into anatomical space*. These eigenvector solutions of the functional connectivity matrix can also be used to map *anatomy into a functional space*. This complementary transformation is equivalent to metric or classical (multidimensional) scaling. These techniques are demonstrated using the PET data set. The next section extends the concept of functional connectivity to the interactions between two anatomically distributed systems. The identification of systems responsible for inter-hemispheric integration during word generation is presented as an example. These sections on functional connectivity conclude by relating functional connectivity and information theory, in terms of the mutual information between two distributed systems.

The sections on effective connectivity begin by contrasting functional and effective connectivity and addressing some of the validation issues that ensue. The more powerful applications of effective connectivity are concerned with *changes* in effective connectivity. This is illustrated with *time-dependent* changes in effective connectivity in the PET data and characterizing nonlinear interactions between striate and extrastriate cortices with fMRI data. This interaction is an example of *activity-dependent* changes in effective connectivity. On the surface these two examples appear a world apart; however, they use identical techniques. These sections include a discussion of some plausible biological mechanisms that may medi-

ate the sort of effective connectivity measured with neuroimaging.

THE DATA

A PET activation study of word generation

The PET data were obtained from six subjects scanned 12 times (every 8 min) whilst performing one of two verbal tasks. Scans were obtained with a CTI PET camera (model 953B CTI, Knoxville, TN, USA). Reconstructed images had a resolution of 5.2 mm [Townsend et al., 1992; Spinks et al., 1992]. ^{15}O was administered intravenously as radiolabelled water infused over 2 min. Total counts per voxel during the buildup phase of radioactivity served as an estimate of rCBF [Fox and Mintun, 1989]. Subjects performed two tasks in alternation. The first task was *repeating a letter* presented aurally at one per two seconds. The second was a *paced verbal fluency* task, where the subjects responded with a word that began with the letter presented. This design ensured that the sensorimotor components were the same for both tasks. This data is the same as that used in Friston et al. [1993a]. To facilitate intersubject pooling, data were stereotactically normalized [Friston et al., 1989, 1991a; Talairach and Tournoux, 1988].

Whole brain differences were removed using ANCOVA [Friston et al., 1990] and the resulting mean activity (over subjects) for each voxel was used in subsequent analyses. A subset of voxels was selected in which a significant amount of variance, due to the 12 conditions, was observed (ANCOVA $F > 3.9$ $P < 0.05$ df 11,54). The time-series from each of these voxels formed a data matrix \mathbf{M} with 12 rows (one for each condition) and 6,477 columns (one for each voxel). Each column of \mathbf{M} was normalized to a mean of zero.

The fMRI data

The data were a time-series of 64 gradient-echo EPI coronal slices (5 mm thick, with 64×64 voxels $2.5 \times 2.5 \times 5$ mm) through the calcarine sulcus and extrastriate areas. Images were obtained every 3 seconds from a normal male subject using a 4.0T whole body system, fitted with a small (27 cm diameter) z-gradient coil (TE 25 ms, acquisition time 41 ms). Photic stimulation (at 16 Hz) was provided by goggles fitted with 16 light emitting diodes. The stimulation was off for the first 10 scans (30 s), on for the second 10, off for the third, and so on. Images were reconstructed without phase correction. The data were interpolated

to 128×128 voxels. Each interpolated voxel thus represented $1.25 \times 1.25 \times 5$ mm of cerebral tissue. The first four scans were removed to eliminate magnetic saturation effects.

To correct for (slight) subject movement during the scanning session the images were translated and rotated to minimize the sum of squares between each image and their average (both scaled to the same mean intensity) using the Levenberg-Marquardt method [More, 1977]. The rotation and translations were effected using a computationally efficient cubic convolution interpolation method [Keys, 1981]. After this correction each voxel time-series was normalized to a mean of zero. Only the 36×60 voxel subpartitions (of the original images) containing the brain were subject to further analysis. The result was a data matrix \mathbf{M} with 60 rows (one for each condition) and 2,160 columns (one for each voxel).

All image manipulations and calculations were performed using Matlab (MathWorks Inc, Sherborn, MA, USA).

FUNCTIONAL AND EFFECTIVE CONNECTIVITY

Origins and definitions

In the analysis of neuroimaging time-series functional connectivity is defined as the *temporal correlations between spatially remote neurophysiological events* [Friston et al., 1993a]. This definition is operational and provides a simple characterization of functional interactions. The alternative is to refer explicitly to effective connectivity (i.e., *the influence one neuronal system exerts over another*) [Friston et al., 1993b]. These sorts of concepts were originated in the analysis of separable spike trains obtained from multiunit electrode recordings [e.g., Gerstein and Perkel, 1969; Gerstein et al., 1989; Gochin, 1991; Aertsen and Preissl, 1991]. Functional connectivity is simply a statement about the observed correlations; it does not provide any direct insight into how these correlations are mediated. For example, at the level of multiunit microelectrode recordings, correlations can result from *stimulus-locked transients*, evoked by a common afferent input, or reflect *stimulus-induced oscillations*, phasic coupling of neural assemblies, mediated by synaptic connections [Gerstein et al., 1989]. To examine the integration within a distributed system, defined by functional connectivity, one turns to *effective connectivity*.

Effective connectivity is closer to the intuitive notion of a connection and can be defined as *the influence on neural system exerts over another*, either at a synaptic

(cf. synaptic efficacy) or cortical level. In electrophysiology there is a close relationship between effective connectivity and synaptic efficacy; "It is useful to describe the effective connectivity with a connectivity matrix of effective synaptic weights. Matrix elements $[C_{ij}]$ would represent the effective influence by neuron j on neuron i " [Gerstein et al., 1989]. It has also been proposed that "the [electrophysiological] notion of effective connectivity should be understood as the experiment and time-dependent, simplest possible circuit diagram that would replicate the observed timing relationships between the recorded neurons" [Aertsen and Preissl, 1991].

Although functional and effective connectivity can be invoked at a conceptual level in both neuroimaging and electrophysiology they differ fundamentally at a practical level. This is because the time-scales and nature of the neurophysiological measurements are very different (seconds vs. milliseconds and hemodynamic vs. spike trains).

In electrophysiology it is often necessary to remove the confounding effects of stimulus-locked transients (which introduce correlations that are *not* causally mediated by direct neuronal interactions) in order to reveal the underlying effective connectivity. The confounding effect of stimulus-evoked transients is less problematic in neuroimaging because the promulgation of dynamics from primary sensory areas onwards is mediated by neuronal connections (usually reciprocal and interconnecting). However it should be remembered that functional connectivity is not necessarily due to effective connectivity (e.g., common neuromodulatory input from ascending aminergic neurotransmitter systems or thalamo-cortical afferents) and, where it is, effective influences may be indirect (e.g., polysynaptic relays through multiple areas).

All the examples presented in this paper rely on descriptive, as opposed to inferential, statistics. In other words the objective is to characterize interactions among brain areas, not to demonstrate that any particular characterization is significant. In some applications distributional approximations do exist to compare one characterization with another, but they are not discussed here.

FUNCTIONAL CONNECTIVITY

Measuring a pattern of correlated activity

In this section we introduce a simple way of measuring the amount a pattern of activity (representing a connected brain system) contributes to the functional connectivity or variance-covariances observed.

Functional connectivity is defined in terms of correlations or covariance (correlations are normalized covariances). The point to point functional connectivity between one voxel and another is not usually of great interest. The important aspects of the covariance structure are the patterns of correlated activity, subtended by (the enormous number of) pairwise covariances. In measuring these patterns it is useful to introduce the concept of a *norm*. Vector and matrix norms serve the same purpose as absolute values for scalar quantities. In other words they furnish a measure of distance. One frequently used norm is the 2-norm, which is simply the length of the vector. The vector 2-norm can be used to measure the degree to which a particular pattern of brain activity contributes to the covariance structure: If a pattern is described by a column vector (\mathbf{p}), with an element for each voxel, then the contribution of that pattern to the covariance structure can be measured by the increase (or decrease) in length it experiences when multiplied by the data matrix \mathbf{M} . This measure is provided by the 2-norm of $\mathbf{M}\mathbf{p} = \|\mathbf{M}\mathbf{p}\|_2$. For mathematical expediency we will work with the square of the 2-norm (T will denote transposition).

$$\|\mathbf{M}\mathbf{p}\|_2^2 = \mathbf{p}^T \mathbf{M}^T \mathbf{M} \mathbf{p}. \quad (1)$$

Put simply the 2-norm is a number which reflects the amount of variance-covariance or functional connectivity that can be accounted for by a particular distributed pattern: If the time-dependent changes occur predominantly in the regions described by the pattern (\mathbf{p}) then the correlation between the pattern of activity and \mathbf{p} at any one time will itself vary substantially with each successive scan. The 2-norm measures this variance in the correlation. The pattern \mathbf{p} embodies the functional connectivity one is interested in. For example if one were interested in the functional connectivity between the left dorsolateral prefrontal cortex (DLPFC) and the left superior temporal region one could test for this interaction using the 2-norm in Equation (1) where \mathbf{p} had large values in the frontal and temporal regions of interest.

The 2-norm has been used to good effect in demonstrating abnormal prefrontotemporal integration in schizophrenia [Friston et al., 1994a,b]. The notion that schizophrenia represents a disintegration or fractionation of the psyche is as old as its name, which was introduced by Bleuler [1913] to convey a “splitting” of mental faculties. Many of Bleuler’s primary processes, such as “loosening of associations,” emphasize a fragmentation and loss of coherent integration. In

order to investigate the neurobiological basis of this disintegration we tested for differences in functional connectivity between the left DLPFC and superior temporal cortex between normal subjects and schizophrenic patients. This involved performing a word generation activation study in one normal and three schizophrenic groups and testing for the integrity of prefronto-temporal integration with a prefronto-temporal pattern (\mathbf{p}). This pattern had high values in the left DLPFC and negative values in the superior temporal regions (and smaller values elsewhere). Using this approach we were able to show that profound negative prefronto-temporal functional connectivity prevalent in normal subjects was absent in three groups of schizophrenic subjects (Fig. 1). In this instance \mathbf{p} was actually chosen to maximize the 2-norm differences between normal subjects and schizophrenic patients by applying SVD to the differences in covariance or functional connectivity matrices (see Friston et al. [1994a] for a fuller discussion). This sort of analysis identifies differences in functional connectivity or covariance structure (*a second order effect*) as opposed to regionally specific differences in physiology (*a first order effect*). In other words the 2-norm can be used to characterize the difference between one subject (or subjects) and another in terms of interactions (functional connectivity) among regions. The 2-norm is an example of a statistic which (in some instances) has a simple distributional approximation (not discussed). It should be noted, of course, that the 2-norm only measures the pattern one is interested in. There may be many other important differences in functional connectivity even if the 2-norms from two time-series are similar for a particular pattern.

Eigenimages and spatial modes

In this section the concept of eigenimages or spatial modes is introduced in terms of the patterns of activity (\mathbf{p}) in the previous section. In the previous section \mathbf{p} was chosen to investigate an interaction of interest. In this section we show that the spatial modes are simply those patterns which account for the *most* variance-covariance (i.e., have the largest 2-norm).

Consider the ubiquitous orthogonalization device singular value decomposition (SVD). SVD is an operation which decomposes the original time-series (\mathbf{M}) into two sets of orthogonal vectors (patterns in space and patterns in time). Let $\mathbf{M} = [\mathbf{m}^1 \dots \mathbf{m}^n]$ be a matrix of the normalized data (one column vector per voxel time-series) and:

$$[\mathbf{u}\mathbf{v}] = \text{SVD}[\mathbf{M}]$$

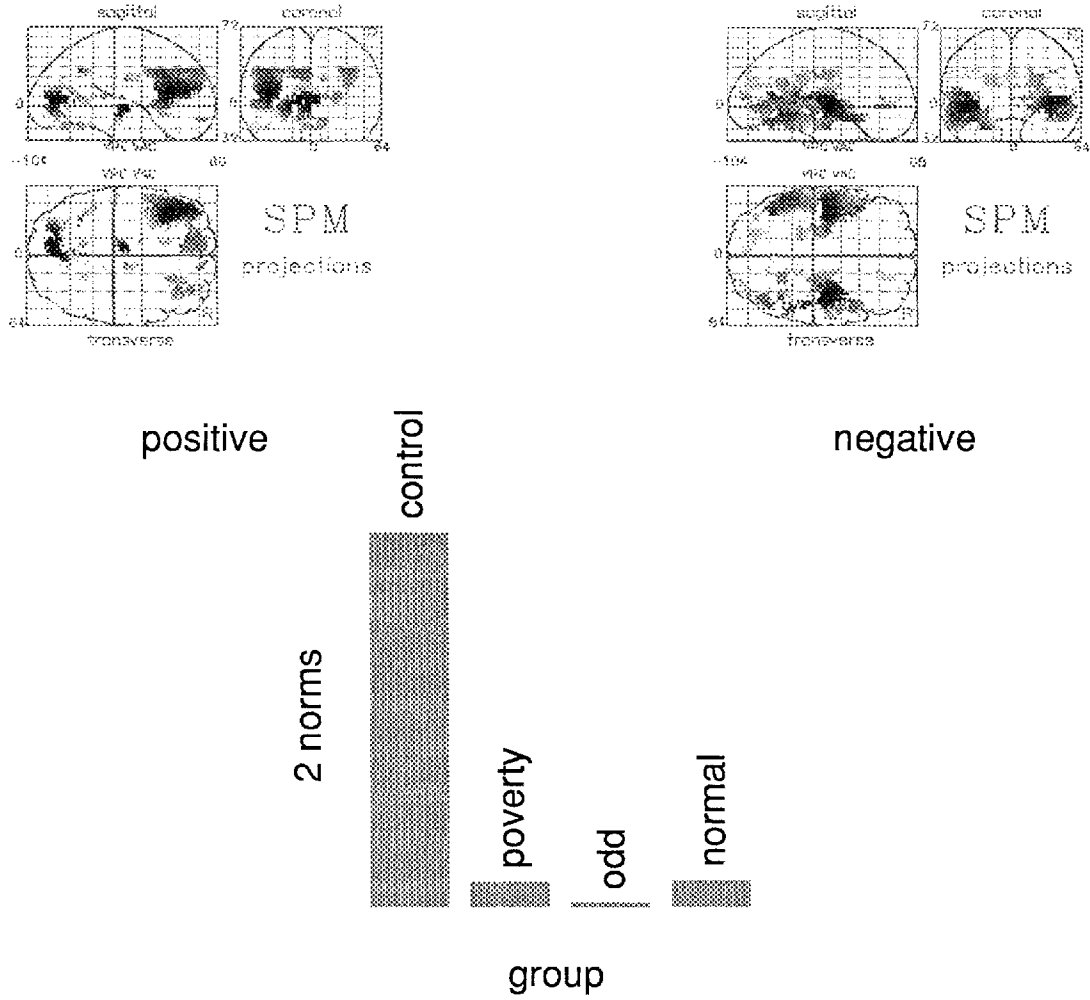


Fig. 1.

Analysis of fronto-temporal functional connectivity in normal subjects and schizophrenic patients. The maximum intensity projections (top) depict the positive and negative components of the pattern (\mathbf{p}) used to test for the covariance (functional connectivity) of interest. The 2-norm of this pattern was markedly attenuated in three groups of schizophrenic subjects compared to normal

subjects. The 2-norm was calculated according to Equation (1) where the data matrices (\mathbf{M}) represented the adjusted condition means for each of the four groups studied. The schizophrenic patients were divided into three groups according to their performance on FAS verbal fluency (poverty of words, odd inappropriate words, and relatively unimpaired performance).

such that:

$$\mathbf{M} = \mathbf{u} \cdot \mathbf{s} \cdot \mathbf{v}^T. \quad (2)$$

In this equation \mathbf{u} and \mathbf{v} are unitary orthogonal matrices (the sum of squares of each column is unity and all the column are uncorrelated) and \mathbf{s} is a diagonal matrix (only the leading diagonal has non-zero values) of decreasing singular values. A rearrangement of Equation (2) gives:

$$\mathbf{v}^T \cdot \mathbf{M}^T \mathbf{M} \cdot \mathbf{v} = \mathbf{s}^2. \quad (3)$$

Because $\mathbf{M}^T \mathbf{M}$ is the functional connectivity matrix (ρ where $\rho_{ij} = \mathbf{m}^{iT} \cdot \mathbf{m}^j$) the columns of $\mathbf{v} = [\mathbf{v}^1 \dots \mathbf{v}^n]$ are the eigenvectors of the functional connectivity matrix. The corresponding eigenvalues are given by the leading diagonal of \mathbf{s}^2 . Referring back to Equation (1) shows that the singular value of each eigenvector (or singular vector) is simply its 2-norm. Because SVD maximizes the largest singular value, the first eigenvector is the pattern which accounts for the greatest amount of the variance-covariance structure. In other words SVD and equivalent devices are simple and

powerful ways of decomposing a neuroimaging time-series into a series of orthogonal patterns that embody, in a stepdown fashion, the greatest amounts of functional connectivity. Each eigenvector defines a distributed brain system that can be displayed as an image. The distributed systems that ensue are called *eigenimages* or *spatial modes* and have been used characterize the spatiotemporal dynamics of neurophysiological time-series from several modalities; including multiunit electrode recordings [Mayer-Kress et al., 1991], EEG [Friedrich et al., 1991], MEG [Fuchs et al., 1992], PET [Friston et al., 1993a], and functional MRI [Friston et al., 1993c].

Many readers will notice that the eigenimages associated with the functional connectivity or covariance matrix are simply the principal components of the time-series [e.g., Friston et al., 1993a]. In the EEG literature one sometimes comes across the Karhunen-Loeve expansion, which is employed to identify spatial modes. If this expansion is in terms of the eigenvectors of the ρ (and it usually is), then the analysis is formally identical to the one presented above.

One might ask what the column vectors of \mathbf{u} correspond to in Equation (2). The vectors are the time-dependent profiles associated with each eigenimage. They reflect the extent to which an eigenimage is expressed in each experimental condition or over time. These vectors play an important role in the functional attribution of the distributed systems defined by the eigenimages and in setting up functional spaces of the sort used in multidimensional scaling. These and other points will be illustrated in the next two sections.

Mapping function into anatomical space— an illustration

The PET data matrix \mathbf{M} was subject to SVD as described in the previous section. The distribution of eigenvalues (Fig. 2 right) suggests only two spatial modes (distributed systems) are required to account for most of the observed variance-covariance structure. The first mode accounted for 68% and the second 16% of the variance. The first two eigenimages (\mathbf{v}^1 and \mathbf{v}^2) are shown in Figure 3 (right) along with the corresponding vectors in time (\mathbf{u}^1 and \mathbf{u}^2 —left). The first eigenimage (top right) has positive loadings in the anterior cingulate (BA 24,32), the left DLPFC (BA 46), Broca's area (BA 44), the thalamic nuclei, and in the cerebellum. Negative loadings were seen bitemporally and in the posterior cingulate (although not shown here the negative loadings were almost identical to

the top right hand panel in Figure 5). According to \mathbf{u}^1 the first mode is prevalent in the verbal fluency tasks with negative scores in word shadowing. The second spatial mode (Fig. 3—lower right) had its highest positive loadings in the anterior cingulate and bitemporal regions (notably Wernicke's area on the left). This mode appears to correspond to a highly non-linear, monotonic time effect with greatest prominence in the earlier conditions (Fig. 3—lower left).

The post hoc functional attribution of these spatial modes is usually based on their time-dependent profiles (\mathbf{u}^i). The first mode may represent an *intentional* system critical for the intrinsic generation of words in the sense that the key cognitive difference between the verbal fluency and word shadowing conditions is the intrinsic generation of (as opposed to extrinsically cued) word representations and implicit mnemonic processing. The second system, which includes the anterior cingulate, seems to be involved in habituation; possibly of attentional or perceptual set (see Wise [1989] for a brief discussion of set). One might infer this from previous PET studies on attention [Petersen et al., 1989; Pardo et al., 1990; Corbetta et al., 1991] and other ideas relating to the distribution of attentional systems [Posner et al., 1990].

The anterior cingulate is an interesting example of an area that belongs, coincidentally, to two orthogonal systems. In other words the activity of this area increases during the verbal fluency task and declines with time (possibly with acquisition of perceptual set). *Yet these two effects are totally independent.*

There is nothing "biologically" important about the particular spatial modes obtained in this fashion, in the sense that one could "rotate" the eigenvectors such that they were still orthogonal and yet have different eigenimages. The uniqueness of the particular solution given by SVD is that the first eigenimage accounts for largest amount of the variance-covariance and the second for the greatest amount that remains and so on. The reason that the eigenimages in the example above lend themselves to such a simple interpretation is that the variance introduced by experimental design (intentional) was substantially greater than that due to time (attentional) and both these sources were greater than any other effect. Other factors that ensure a parsimonious characterization of the time-series, with small numbers of well defined modes include 1) smoothness in the data and 2) using only voxels that evidenced a non-trivial amount of change during the scanning session.

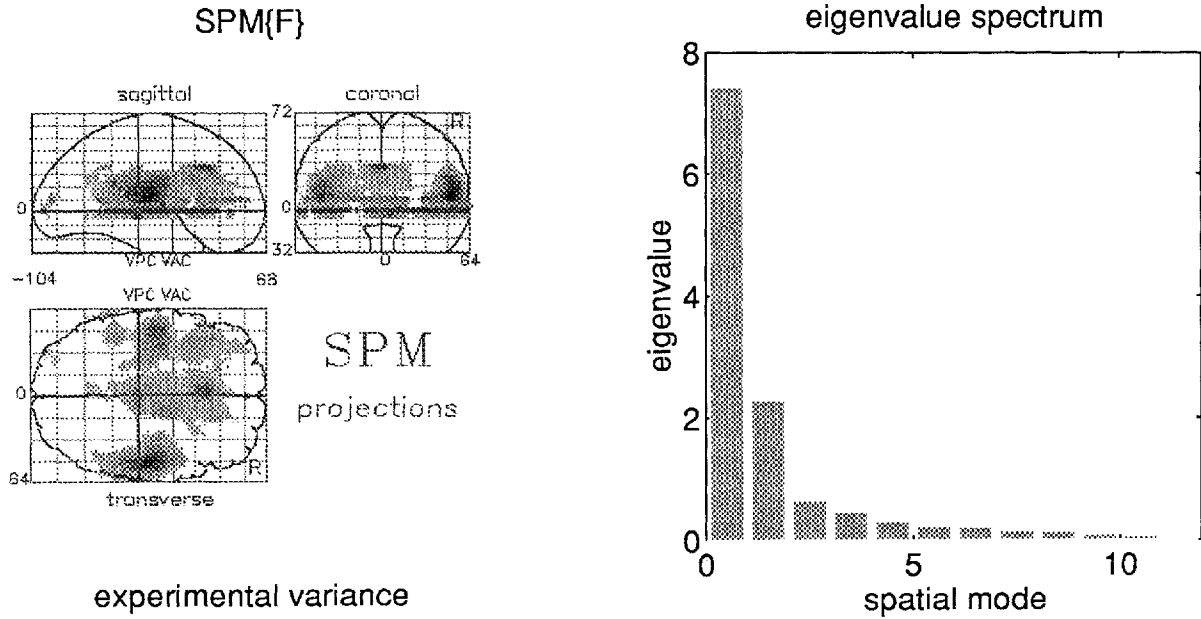


Fig. 2.

Experimentally introduced variance in the PET activation study of word generation. **Left:** Statistical parametric map (SPM [Friston et al., 1991b]) of the F ratio following an ANCOVA of the six-subject, 12-condition verbal fluency study. The maximum intensity projection display format is standard and provides three views of the brain in the stereotactic space of Talairach and Tournoux [1988] (from

the back, from the right, and from the top). **Right:** Eigenvalues (singular values squared) of the functional connectivity matrix reflecting the relative amounts of variance accounted for by the 11 spatial modes associated with this data. Only two eigenvalues are greater than unity and to all intents and purposes the changes characterizing this time-series can be considered two-dimensional.

Mapping anatomy into functional space— multidimensional scaling

In the previous section the functional connectivity matrix was used to define the associated eigenimages or spatial modes. In this section functional connectivity is used in a different way: Namely to constrain the proximity of two cortical areas in some functional space. The objective here is to transform anatomical space so that the distance between cortical areas is directly related to their functional connectivity. This transformation defines a new space whose topography is purely functional in nature. This space is constructed using multidimensional scaling or principal coordinates analysis [Gower, 1966].

Multidimensional scaling (MDS) is a descriptive method for representing of the structure of a system, on the basis of pairwise measures of similarity or confusability [Torgerson, 1958; Shepard, 1980]. The resulting multidimensional spatial configuration of the system's elements embody (in their proximity relationships) the comparative similarities. The technique was developed primarily in the analysis of

perceptual spaces. The proposal that stimuli be modeled by points in space, so that perceived similarity is represented by spatial distances, goes back to the days of Isaac Newton [1704]. The implementation of this idea is however relatively new [see Shepard, 1980]. A typical model underlying MDS can be summarized by:

$$F_{\text{mon}}(\delta_{ij}) \cong d_{ij}$$

$$d_{ij} = \sqrt{\sum (x_i^l - x_j^l)^2} \quad (4)$$

where $F_{\text{mon}}(\cdot)$ is a decreasing monotonic function. δ_{ij} is the measure of similarity (functional connectivity) between elements (voxels) i and j . d_{ij} is the distance between them in a Euclidean functional space. x_i^l is the l th coordinate of element (voxel) i in this new space. (\cong means equal, except for unspecified error terms.)

Imagine K measures from n voxels plotted as n points in a K -dimensional space (K -space). If they have been normalized to zero mean and unit sum of squares, these points will fall on a $K-1$ dimensional

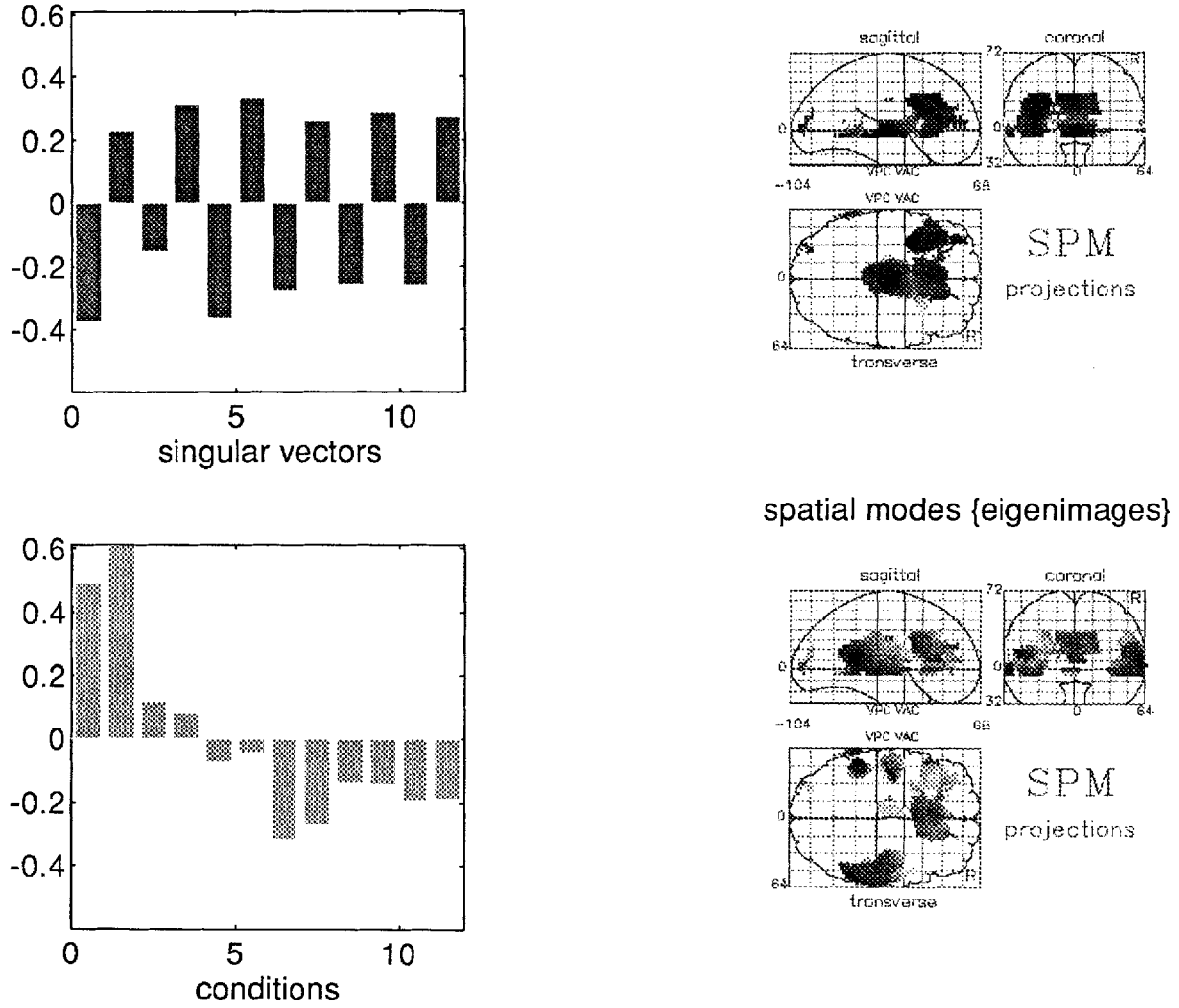


Fig. 3.

Spatial modes and singular vectors. **Left:** Time-dependent expression of the spatial modes (u^i) or alternatively eigenvectors of the distribution of points in a functional space (i.e., eigenvectors of MM^T). The attribution of the corresponding spatial modes, or direction in a functional space, depends on relating this vector to the tasks employed during the activation. The first vector (**top**) is clearly related to the difference between word generation (even-numbered conditions) and word shadowing (odd-numbered scans).

This difference is the intentional or intrinsic generation of word representations. The second vector (**bottom**) corresponds to some largely monotonic time effect we have labelled attentional. **Right:** The eigenimages or spatial modes corresponding to the first two eigenvectors of the functional connectivity matrix. These modes are the eigenvectors of $M^T M$. The eigenimages are displayed as a maximum intensity projection in standard SPM format. The color scale is arbitrary and each SPM is scaled to its maximum.

hypersphere. The closer any two points are to each other, then the greater their correlation or functional connectivity (in fact the correlation is the cosine of the angle subtended at the origin). The distribution of these points embodies the functional topography. A view of this distribution that reveals the greatest structure is simply obtained by rotating the points to maximize their apparent dispersion (variance). In other words one looks at the subspace with the largest “volume”

spanned by the principal axes of the n points in K -space. These principal axes are given by the eigenvectors of $M \cdot M^T$, i.e., the column vectors of u . From eqn(2):

$$M \cdot M^T = u \cdot s^2 \cdot u^T.$$

Let $X = [x^1 \dots x^r]^T$ be the matrix of desired coordinates derived by simply projecting the original data (M^T)

onto the axes defined by \mathbf{u} :

$$\mathbf{X} = \mathbf{M}^T \mathbf{u}. \quad (5)$$

Voxels that have a correlation of unity will occupy the same point in the MDS space. Voxels that have independent dynamics ($\rho_{ij} = 0$) will be $\sqrt{2}$ apart. Voxels that are negatively but totally correlated ($\rho_{ij} = -1$) will be maximally separated (by a distance of 2). Profound negative correlations denote a functional association that are modeled in the MDS functional space as diametrically opposed locations on the hypersphere. In other words two regions with profound negative correlations will form two “poles” in functional space.

This approach is equivalent to using the functional connectivity (ρ_{ij}) as the measure of similarity ($\delta_{ij} = \rho_{ij}$) where:

$$F_{\text{mon}}(\rho_{ij}) = d_{ij} = \sqrt{2} \cdot \sqrt{(1 - \rho_{ij})}. \quad (6)$$

This can be seen by noting that the points defined by the coordinates (\mathbf{x}_i^j) are an orthogonal rotation of the original data (\mathbf{m}_k^i —the k th element of \mathbf{m}^i). Because orthogonal rotation does not change Euclidean distances:

$$\begin{aligned} d_{ij} &= \sqrt{\sum (\mathbf{x}_i^j - \mathbf{x}_j^j)^2} = \sqrt{\sum (\mathbf{m}_k^i - \mathbf{m}_k^j)^2} \\ &= \sqrt{[\sum (\mathbf{m}_k^i)^2 + \sum (\mathbf{m}_k^j)^2 - 2 \cdot \sum (\mathbf{m}_k^i \mathbf{m}_k^j)]} = \sqrt{2} \cdot \sqrt{(1 - \rho_{ij})}. \end{aligned} \quad (7)$$

An illustration

Following normalization to *unit sum of squares over each column* \mathbf{M} was subject to singular value decomposition according to Equation (2) and the coordinates \mathbf{X} of the voxels in the MDS functional space computed as in Equation (5). Recall that only two eigenvalues exceed unity (Fig. 2—right) suggesting the functional space is essentially two dimensional. The locations of voxels in this two-dimensional subspace are shown in Figure 4 (lower row) by rendering voxels from different regions in different colours. The anatomical regions corresponding to the different colours are shown in Figure 4 (upper row). Anatomical regions were selected to include those parts of the brain which showed the greatest variance during the 12 conditions (see Fig. 2—left). Anterior regions (Fig. 4—right) included the mediodorsal thalamus (blue), the dorsolateral prefrontal cortex (DLPFC) and Broca’s area (red), and the anterior cingulate (green). Posterior regions (Fig. 4—left) included the superior temporal regions

(red), the posterior superior temporal regions (blue), and the posterior cingulate (green). The corresponding functional spaces (Fig. 4—lower row) reveal a number of things about the functional topography elicited by this set of activation tasks. First each anatomical region maps into a relatively localized portion of functional space. This preservation of local contiguity reflects the high correlations within anatomical regions, due in part to smoothness in the original data and to high degrees of intraregional functional connectivity. Secondly the anterior regions are almost in juxtaposition as are the posterior regions; however, the confluence of anterior and posterior regions form two diametrically opposing poles (or one axis). This configuration suggests an anterior–posterior axis with prefrontotemporal and cingulocingulate components. One might have predicted this configuration by noting that the anterior regions had high positive loadings on the first eigenimage (see top right panel of Figure 3) while the posterior regions had high negative loadings [see top right panel Figure 5 (which is very similar to the negative loadings)]. Thirdly within the anterior and posterior sets of regions certain generic features are evident. The most striking is particular ordering of functional interactions. For example the functional connectivity between the posterior cingulate (green) and superior temporal regions (red) is high and similarly for the superior temporal (red) and posterior temporal regions (blue), yet the posterior cingulate and posterior temporal regions show very little functional connectivity (they are $\sqrt{2}$ apart or equivalently subtend 90° at the origin).

These results are consistent with known anatomical connections. For example DLPFC–anterior cingulate connections, DLPFC–temporal connections, bitemporal commissural connections, and mediodorsal thalamic–DLPFC projections have all been demonstrated in nonhuman primates [e.g., Goldman-Rakic, 1986, 1988]. The mediodorsal thalamic region and DLPFC are so correlated that one is embedded within the other (purple area). This is pleasing given the known thalamocortical projections to the DLPFC.

At this point one might ask if absolute position in the MDS functional space has any meaning. For example, is the fact that the fronto-temporal axis is horizontal (as opposed to vertical) important? The answer is yes. Each dimension (\mathbf{u}^i) of the transformed space has the same specific functional attribution as the spatial mode (\mathbf{v}^i). Consequently one could designate the first horizontal dimension of the functional space (\mathbf{u}^1) as *intentional* (corresponding to the intentional or intrinsic generation of words) and the second

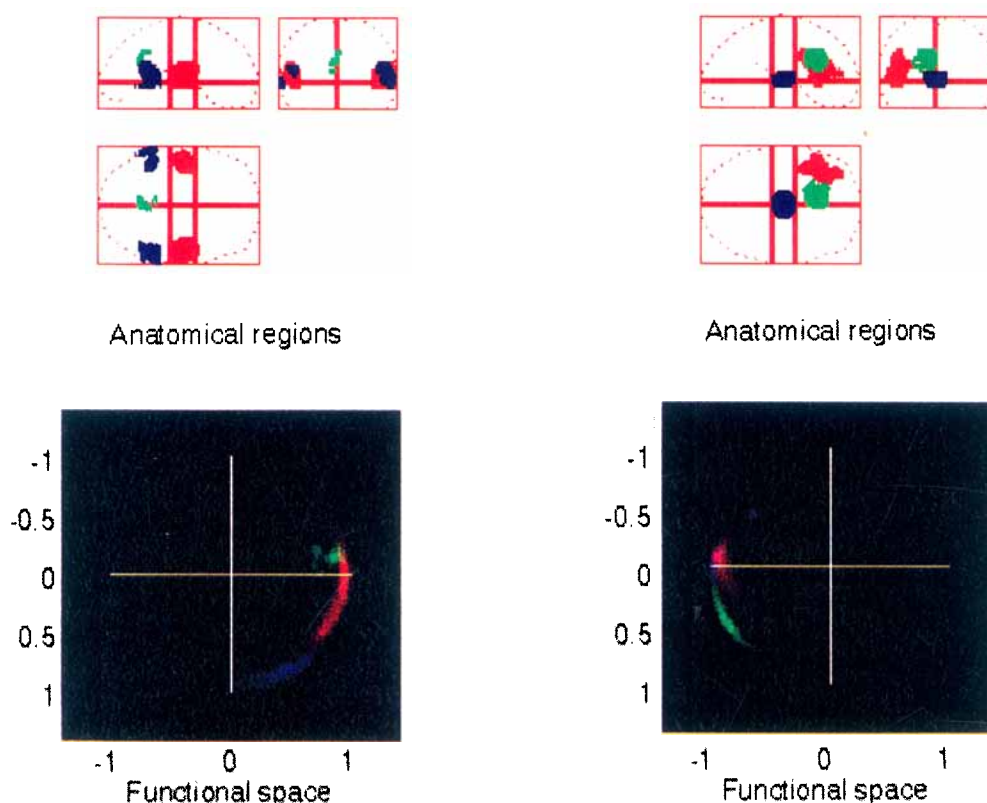


Fig. 4.

Classical or metric scaling analysis of the functional topography of intrinsic word generation in normal subjects. **Top:** Anatomical regions categorized according to their colour. The designation was by reference to the atlas of Talairach and Tournoux [1988]. **Bottom:** Regions plotted in a functional space following the scaling transformation. In this space the proximity relationships reflect the functional connectivity between regions. The colour of each voxel

corresponds to the anatomical region it belongs to. The brightness reflects the local density of points corresponding to voxels in anatomical space. This density was estimated by binning the number of voxels in 0.02 'boxes' and smoothing with a Gaussian kernel of full width at half maximum of three boxes. Each colour was scaled to its maximum brightness.

vertical dimension (u^2) as *attentional* (attentional changes or changes in perceptual set as the experiment proceeds).

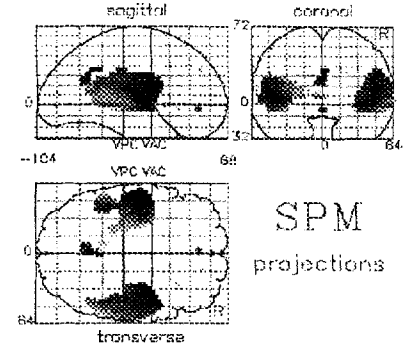
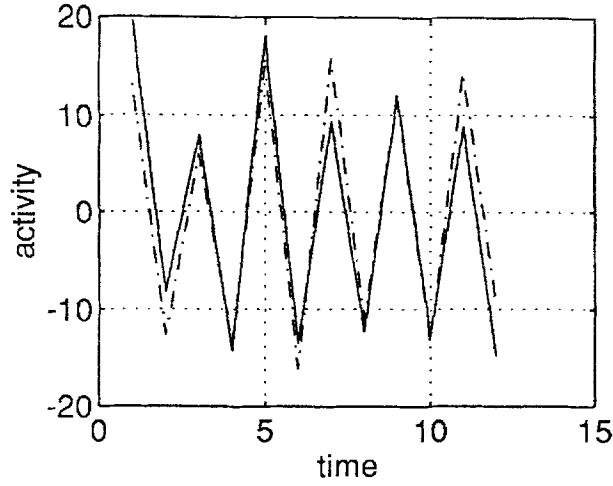
This perspective provides a slightly richer interpretation of the functional space in the following way: Functional connectivity (distance) between two regions can be partitioned into intentional (horizontal) and attentional (vertical) components. For example the horizontal proximity of the DLPFC (red) and anterior cingulate (green) is greater than their vertical proximity. In other words the functional connectivity between the DLPFC and anterior cingulate is dominated by the intentional aspects of the tasks used to elicit the functional interactions.

It is fortuitous that the data used to illustrate this approach were essentially two dimensional. If the

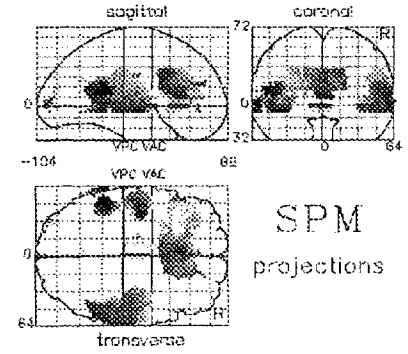
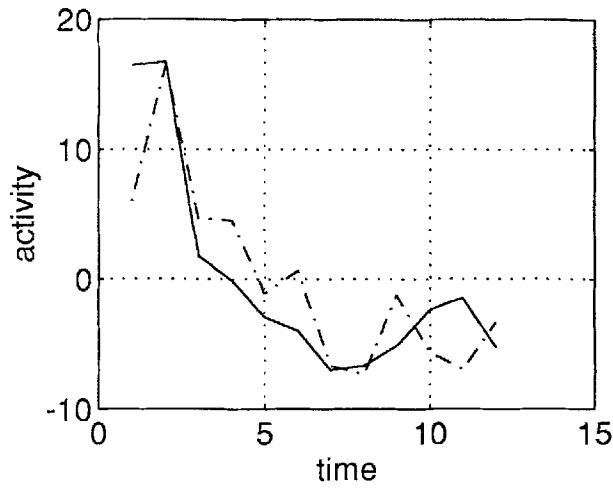
data have a greater dimensionality then it is necessary to provide several "views" of the functional space to completely capture the important relationships.

Functional connectivity between systems

Hitherto we have been dealing with the functional connectivity between two voxels. The same notion can, however, be extended to the functional connectivity between two systems by noting that there is no fundamental difference between the dynamics of one voxel and the dynamics of a distributed system or pattern. The functional connectivity between two systems is simply the correlation or covariance between their time-dependent activity. The time-depen-



1st pair



2nd pair

Fig. 5.

Systems subtending the greatest inter-hemispheric integration as defined by SVD. Upper row: The systems associated with the largest singular value. Lower row: The systems associated with the second largest singular value. **Left:** Activity of the system depicted on the left; right hemisphere, solid line; left hemisphere, broken line. **Right:** Images corresponding to the pairs of singular vectors in

Equation (10). Because each vector specifies the value of voxels in each hemisphere the two systems can be shown in one image. Although the bitemporal regions figure in both the first and second pairs the first pair implicates the *posterior* cingulate, whereas the second pair includes the *anterior* cingulate.

dent activity of a system or pattern \mathbf{p} is given by:

$$\mathbf{m}^p = \mathbf{M} \cdot \mathbf{p}$$

therefore:

$$\rho_{pq} = \mathbf{m}_q^T \cdot \mathbf{m}^p = \mathbf{q}^T \cdot \mathbf{M}^T \cdot \mathbf{M} \cdot \mathbf{p} \quad (8)$$

where ρ_{pq} is the functional connectivity between the systems described by vectors \mathbf{p} and \mathbf{q} . Note that different spatial modes have no functional connectivity (i.e., $\rho_{vivi} = 0$; $i \neq j$). Consider next functional con-

nectivity between two systems in separate parts of the brain, for example the right and left hemispheres. Here the data matrices (\mathbf{M}_p and \mathbf{M}_q) derive from different sets of voxels and eqn(8) becomes:

$$\rho_{pq} = \mathbf{m}_q^T \cdot \mathbf{m}_p = \mathbf{q}^T \cdot \mathbf{M}_q^T \cdot \mathbf{M}_p \cdot \mathbf{p}. \quad (9)$$

If one wanted to identify the intra-hemispheric systems that showed the greatest inter-hemispheric functional connectivity one would need to identify

the set of vectors \mathbf{p} and \mathbf{q} that maximize ρ_{pq} in Equation (9). SVD finds yet another powerful application in doing just this:

$$[\mathbf{USV}] = \text{SVD}\{\mathbf{M}_q^T \mathbf{M}_p\}$$

such that:

$$\mathbf{M}_q^T \mathbf{M}_p = \mathbf{U} \mathbf{S} \mathbf{V}^T$$

and

$$\mathbf{U}^T \mathbf{M}_q^T \mathbf{M}_p \mathbf{V} = \mathbf{S}. \quad (10)$$

The first columns of \mathbf{U} and \mathbf{V} represent the singular images that correspond to two systems that evidence the greatest amount of functional connectivity (the singular values in the diagonal matrix \mathbf{S}). In words SVD of the (generally asymmetric) covariance matrix based on the time-series from two anatomically separate parts of the brain yields a series of paired vectors (\mathbf{U}^i and \mathbf{V}^i) that, in a stepdown fashion, define pairs of brain systems that show the greatest functional connectivity.

To illustrate this the data matrix \mathbf{M} was partitioned into a matrix that included all the voxels from the right hemisphere (\mathbf{M}_p) and a matrix that included voxels from the left hemisphere (\mathbf{M}_q). These two matrices were subject to SVD as in Equation (10) and the first two pairs of singular vectors were rendered as images. Because each pair of vectors (\mathbf{U}^i and \mathbf{V}^i) are non-overlapping in space they can be displayed in the same anatomical projection (Fig. 5—upper right \mathbf{U}^1 and \mathbf{V}^1 and lower right \mathbf{U}^2 and \mathbf{V}^2). The corresponding time-dependent activities ($\mathbf{M}_p \mathbf{V}^i$ and $\mathbf{M}_q \mathbf{U}^i$) are plotted on the left.

This analysis reveals that the first pair of systems subtending the greatest interhemispheric integration (Fig. 5—upper row) are the bitemporal regions and posterior cingulate (the negative parts of these profile are not shown and included the left DLPFC). The covariances mediating this integration reflect the intentional aspects of the tasks and these pairs of patterns are almost indistinguishable from the first spatial mode. The second pair (Fig. 5—lower row) includes the bitemporal regions and the *anterior* cingulate. The corresponding profile of activities suggests these interactions could be designated as attentional (habituation of set—cf. the second spatial mode).

Functional connectivity and information theory

This, the final section on functional connectivity, places functional connectivity in the context of information theory. In the previous section we provided a way of assessing the functional connectivity between two single systems, each defined by a pattern or vector. In this section we consider the functional connectivity between two anatomical regions. The essential difference is that the dynamics of a voxel, system, or pattern is given by a single time-series. The dynamics of an anatomical region containing n voxels is given by n time-series. The extension of functional connectivity from uni- to multidimensional time-series requires a measure of mutual dependence called mutual information (MI). MI is a measure of the predictability of one set of multidimensional measurements given another; it an information theoretic measure that reflects mutual dependence. More formally:

$$\text{MI}_{pq} = \log(|\mathbf{M}_p^T \mathbf{M}_p| \cdot |\mathbf{M}_q^T \mathbf{M}_q| / |\mathbf{M}^T \mathbf{M}|^{1/2}) \quad (11)$$

where $|\cdot|$ denotes the determinant and \mathbf{M}_p and \mathbf{M}_q denote partitions of the data matrix as above. The determinant of a covariance matrix (e.g., $|\mathbf{M}_p^T \mathbf{M}_p|$) can be thought of as the multidimensional equivalent of variance. Because the determinant is unaffected by orthogonal rotation the determinants can be simply expressed in terms of the singular values. For example [from Eq. (2)]:

$$|\mathbf{M}^T \mathbf{M}|^{1/2} = |\mathbf{s}| = \prod \mathbf{s}^i. \quad (12)$$

In other words the mutual information between two spatially extended regions is the log of the products of singular values associated with all the spatial modes on considering the regions separately divided by the product obtained by considering the two regions together. Functional connectivity and mutual information both reflect the same thing. In the limiting case of just two voxels (or patterns) it is easy to show that:

$$\text{MI}_{pq} = -\log(1 - \rho^2)/2 \quad (13)$$

where ρ is the functional connectivity expressed as a correlation. MI has been used as a measure of intra-hemispheric functional connectivity to investigate interhemispheric integration in normal subjects and schizophrenic patients with some interesting results [Friston and Frith, 1994c].

Clearly this form of quantifying cortico-cortical integration is well suited to the characterization of putative disconnection syndromes. It should be noted that computing the MI requires that the matrices are nonsingular (the determinant is not equal to zero). This requires the number of time-series be smaller than the number of observations. In practice it is therefore necessary to apply Equation (11) to the dynamics of the spatial modes [Eq. (8)] as opposed to the original voxel time-series.

EFFECTIVE CONNECTIVITY

Models and validation

Functional connectivity is an operational definition. Effective connectivity is not. It depends on some model of the influence one neuronal system exerts over another. In this sense the validity of effective connectivity reduces to the validity of the model. We shall consider linear and nonlinear models and several approaches to establishing validity. Perhaps the simplest model of effective connectivity expresses the hemodynamic change at one voxel as a weighted sum of changes elsewhere. The weights or coefficients can then be identified with effective connectivity: For example, in matrix notation, one has:

$$\mathbf{m}^i = \mathbf{M} \cdot \mathbf{C}^i + \mathbf{e} \quad (14)$$

where \mathbf{C}^i is a column-vector of effective connectivities from all locations to the one in question (i). \mathbf{e} is an error term that is not correlated with any voxel time-series. If one selects a point (i) in the brain and asks; what is the effective connection strength between the location chosen and all other locations?, then one wants to know the values of all the elements of \mathbf{C}^i . The least squares solution for \mathbf{C}^i is [Binmore, 1982]:

$$\mathbf{C}^i = (\mathbf{M}^T \mathbf{M})^{-1} \cdot \mathbf{M}^T \cdot \mathbf{m}^i. \quad (15)$$

This solution can be regarded as a simple linear regression where the effective connectivity (C_{ij}) reflects the amount of rCBF variability, at i , attributable to rCBF changes at location j . Implicit in this interpretation is a mediation of this influence by neuronal connections with an effective strength equal to C_{ij} . There are three issues that deserve comment when estimating effective connectivity from neuroimaging data in this way. First these equations are linear, whereas cerebral physiology is not. However the

analysis of *functional connectivity* (see previous sections) suggested that the linear system was a sufficiently good approximation to account for observed changes in rCBF in a fairly comprehensive and meaningful way. Secondly the fact that there are only a few observations (12 in the PET study) but many (here 6,477) voxels means the set of inconsistent linear equations Equation (14) represents are underdetermined. The problem of underdetermination is dealt with finding a solution with the minimum 2-norm of all minimizers [Golub and Van Loan, 1991]. This is formally equivalent to solving the equation in the space defined by the spatial modes or eigenimages. Herein is a fundamental mathematical link between functional connectivity and effective connectivity when assessed in short time-series of spatially extended systems like images. Mathematically:

$$\mathbf{m}^i = \mathbf{M} \cdot \mathbf{v} \cdot \mathbf{a}^i + \mathbf{e} \quad (16)$$

is solved for \mathbf{a}^i , where $\mathbf{C}^i = \mathbf{v} \cdot \mathbf{a}^i$. As in previous sections $\mathbf{v} = [\mathbf{v}^1 \dots \mathbf{v}^r]$ is a matrix of r eigenimages and $\mathbf{M} \cdot \mathbf{v} = \mathbf{u} \cdot \mathbf{s}$. Only the r (or less) eigenvectors with non-zero (or large) eigenvalues are used, giving:

$$\mathbf{C}^i = \mathbf{v} \cdot \mathbf{s}^{-2} \cdot \mathbf{v}^T \mathbf{M}^T \cdot \mathbf{m}^i = \mathbf{v} \cdot \text{pinv}(\mathbf{M} \cdot \mathbf{v}) \cdot \mathbf{m}^i \quad (17)$$

where $\text{pinv}(\cdot)$ denote the pseudo inverse. Although this works mathematically, the question 'is the estimation of \mathbf{C}^i a good one?' remains.

Issues of validity

Because effective connectivity is not on operational definition (cf. functional connectivity) and is universally based on some model, issues of validity should always be addressed. These can be seen as comments on the validity of the model to account for the influences one neural system can be inferred to exert over another. In our work we separate these comments into those dealing with construct validity (does the model have validity in terms of another construct or framework, face validity (does the model capture what it is meant to), and predictive validity (does the model accurately predict the system's behaviour).

To illustrate how one might address the validity of the simple linear model above [Eq. (14)] consider construct validity in terms of anatomical connectivity. The premise here is that if the topography of effective connections conform to the known anatomical efferents to a particular area, one sort of construct validity would have established. Figure 6 (lower row) presents

maximum intensity projections of (the positive values of) C^i where the reference region (i) was placed in the left DLPFC (top). These estimates were derived using Equation (17) using the first and last eight PET scans of the word generation study (the reason for using these subsets relates to the examination of time-dependent changes described below). The key (darker) sources of effective connections include contiguous and extensive parts of the DLPFC, frontal operculum and Broca's area (BA 44), the anterior cingulate (BA 24, 32), a midline region centred on the medial dorsal thalamic nuclei, the retrosplenial cortex, and a more ventral area (including the brainstem), coextensive laterally with the parahippocampal gyri. These connections have been demonstrated anatomically (either as direct or indirect) in nonhuman primates [Goldman-Rakic, 1986, 1988]. Face validity can be established using the empirical data matrix (M) and an assumed effective connectivity (C^a). These simulations involve computing the expected rCBF (m^a) in a hypothetical region using C^a and Equation (14), and then working backwards to solve for C^a , using only M and m^a . Clearly this would be a rather circular exercise were it not for the underdetermination of the problem and the effect of simulated noise (e^a). Predictive validity can be assessed as the ability to predict changes in activity at some reference region on the basis of the observed changes elsewhere and effective connection strengths *estimated from independent data* (see Friston et al. [1994d] for an example using fMRI data).

Provisional experience suggests that the linear model [Eq. (14)] can be unnaturally well behaved. One explanation for this is that the dimensionality (the number of things which are going on) of the physiological changes over the brain can be very small. In fact the distribution of eigenvalues associated with the PET study suggest a dimensionality of two or three (Fig. 2—right) in the subset of voxels analyzed. In other words the brain responds to simple and well organized experiments in a simple and well organized way. In the example presented despite having measurements in 6,477 voxels there are only substantial changes in two or three spatial modes. As there are more than three observations the apparently underdetermined problem is, in one sense, *overdetermined*.

Finally it is important to realize that effective connectivity is being estimated in terms of the eigenimages ($[v^1 \dots v^r]$). If a particular pattern of connections has nothing in common with any of the eigenimages used in the least squares analysis, then these connections will "not be seen." In other words the pattern of rCBF correlations completely constrains the estimation of

effective connections which are, as a result, time and experiment-dependent.

The mathematical relationship between functional and linear effective connectivity

There is a very simple relationship between the functional connectivity and effective connectivity matrices if we are dealing with a linear model. By substituting M for m^i in Equation (17) we obtain:

$$\text{effective connectivity matrix} = C = v.v^T.$$

Compare this with the equivalent expression for the functional connectivity matrix

$$\rho = M^T.M = v.\lambda.v^T.$$

One way of interpreting this relationship is to note that linear effective connectivity can be represented by a series of orthogonal (noninteracting) spatial modes that are all self connected with unit strength. The functional "self" connectivity is simply the variance of each mode, namely the eigenvalues in the leading diagonal of λ ($= s^2$). The simplicity of effective connectivity assessed with a linear model is reflected in the fact that C is symmetric ($C_{ij} = C_{ji}$). Clearly the face validity of linear effective connectivity is somewhat limited by this and points to the importance of nonlinear models where asymmetry is allowed (see below).

CHANGES IN EFFECTIVE CONNECTIONS

In the next two sections we apply the concept of effective connectivity to two neurobiological questions. In both examples the questions can be formulated in terms of effective connection strength *changes* from one set of circumstances to another. The first example uses the PET data and shows that the changes in C_{ij} with time conform to those changes that would be predicted by an associative model of synaptic plasticity applied to the C_{ij} estimates. The second example demonstrates the presence of modulatory interactions between V2 and V1 using fMRI data. In this instance modulation is characterized by a change in "apparent" C_{ij} with activity intrinsic to the area being modulated. This can be seen by considering the C_{ij} in the absence of any intrinsic activity. As there is nothing to modulate the apparent C_{ij} will be zero. Although these models differ on many levels they share a common reliance on demonstrating a *change in connection strengths* and the same mathematical techniques.

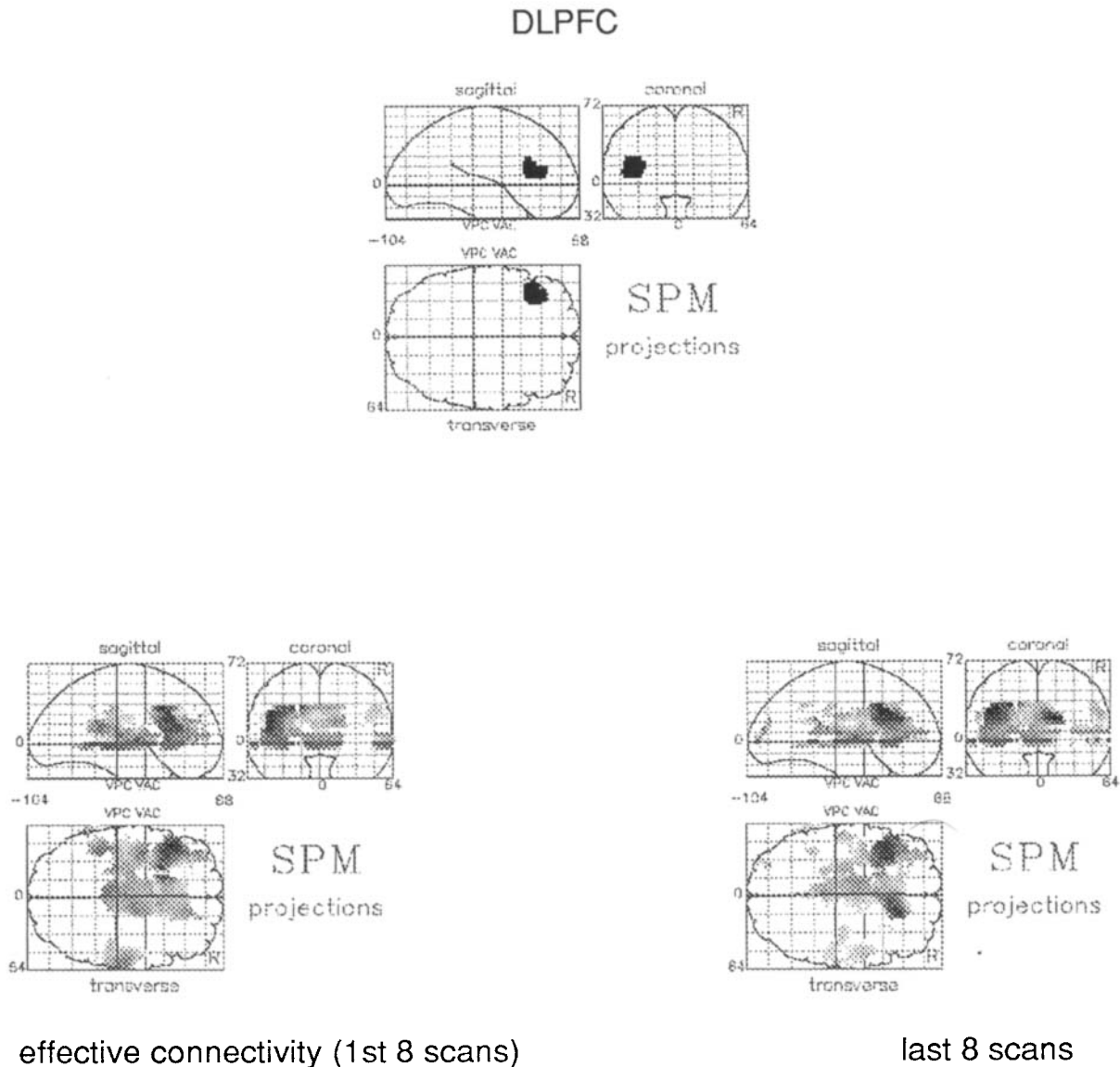


Fig. 6.

The estimated effective connectivity to a region in the dorsolateral prefrontal cortex (centred on $x, y, z = -40, 28, 16$ according to the atlas of Talairach and Tournoux [1988]). **Top:** the location of the reference region (all voxels within 12 mm of the reference location). **Lower right and left:** The mean effective connectivity

estimated as in eqn(17) on the basis of the first and last eight scans of the 12 scan time-series. The gray scale is arbitrary. The main regions highlighted include the left prefrontal cortex, midline and thalamic structures, and the anterior cingulate.

Time-dependent changes in linear effective connectivity

This section introduces the sort of electrophysiological evidence that has been used in support of associative plasticity in the brain and relates empirically determined equations to general rules which model synaptic change. We then see how these associative

models can be applied to rCBF data and, in particular, the changes in effective connectivity they predict. Finally provisional evidence is presented confirming that the predicted changes are indeed observed.

Lopez et al. [1990] have characterized associative long term modification of the ipsilateral and contralateral synapses formed by the bilateral entorhinal cortical (EC) projections to the dentate gyrus (DG) using

extracellular recording of population excitatory post-synaptic potentials. The contingencies of associated pre- and post-synaptic activation accurately predicted the altered synaptic responses of both ipsilateral and contralateral EC-DG pathways to stimulation. The authors proposed that the simplest equation consistent with these (and other electrophysiological observations) has the form:

$$\Delta C_{ij} = \langle f(m_i^t) \cdot m_j^t \rangle - \langle f(m_i^t) \cdot C_{ij} \rangle \quad (18)$$

(our notation where m_k^i and m_k^j are post and pre synaptic activity respectively). $f(\cdot)$ is some monotonic function and $\langle \cdot \rangle$ denotes averaging over time (t). This equation is a variant of Kohonen's [1982] equation. Assuming $f(\cdot)$ is linear Equation (18) becomes a special case of a more generic form (in matrix notation):

$$\Delta C^i = \partial M^T m^i - \phi(M, m^i, C^i) \quad (19)$$

used by Oja [1989] (our notation), where ΔC^i is the change in effective connectivity and $\partial M^T m^i$ is an *associative* or Hebbian term reflecting the correlation between pre- (M) and post- (m^i) synaptic activity. $\phi(M, m^i, C^i)$ is a *nonassociative*, decay or "forgetting" term which does not depend on an interaction between pre- and post-synaptic activity. The variables ∂ and ϕ are constants which govern the rate of associative plasticity and decay respectively.

Applying the notion of associative plasticity to rCBF data means that one expects to see an increase in effective connectivity between two brain regions if their rCBF goes up and down together. In other words if two regions have high activity at the same time, the effective connections between them will be consolidated (the mediation of this effect is discussed below).

The behavior predicted by associative or Hebbian changes of this sort is that the connections to any brain region will come to reflect the dominant pattern of correlations in inputs from the rest of the brain. The most dominant pattern of correlations is described by the first spatial mode (v_1). In other words the connections from the distributed brain regions defined by the first mode will be selectively enhanced relative to any competing pattern. The average strength of connections from the k th spatial mode is $a_k^i = v_k^T C^i$; this is simply the dot product or "overlap" between the mode and the pattern of afferent connection strengths. Alternatively a_k^i can be thought as a single *virtual connection* from the distributed brain system defined

by v_k . Associative plasticity would predict that the change in $a_1^i > a_2^i > a_3^i$ and so on.

The hypothesis that changes in effective connectivity would show this behavior can be tested by measuring C^i at the beginning of an experiment and some time later and estimating the change (ΔC^i). Figure 7 (left) shows ΔC^i where the two C^i were the estimates shown in Figure 6 (for the first and last eight scans of the 12 scan time-series). Again this estimation is in the space defined by the spatial modes (in this instance of the whole time series):

$$\begin{aligned} \Delta C^i &= v \cdot (\text{pinv}(M_2 \cdot v) \cdot m_2^i - \text{pinv}(M_1 \cdot v) \cdot m_1^i) \\ &= v \cdot (a_2^i - a_1^i) \end{aligned} \quad (20)$$

where the subscripts 1 and 2 denote different parts of the time series (in the example presented—the first and last eight scans).

The resulting profile of time-dependent changes in effective connectivity to the left DLPFC show increased effective connectivity from the anterior cingulate, frontal operculum, midline (thalamic) structures, and some temporal regions (Fig. 7—left). To see if this profile was consistent with an associative model of effective connectivity the connection strengths from the first three spatial modes were computed for the two sets of data. It can be seen that, as predicted, the first spatial mode has increased its contribution, whereas the second and third have reduced their (negative) contribution (Fig. 7—right). A more comprehensive analysis is presented in Friston et al. [1993b]; however it should be noted that the results here and in Friston et al. [1993b] are only used to illustrate the idea and methodology. Clearly a proper examination of time-dependent changes in effective connectivity would require exhaustive empirical study and a well validated experimental model of effective connectivity.

The relationship between synaptic and physiological plasticity

The results in the previous section suggest rCBF and electrophysiological measurements can show a degree of homology in terms of associative changes in effective connectivity. The use of a Hebbian model, to account for changes in rCBF, implies that when two brain regions are highly correlated, in terms of their physiology, then the effective connectivity between

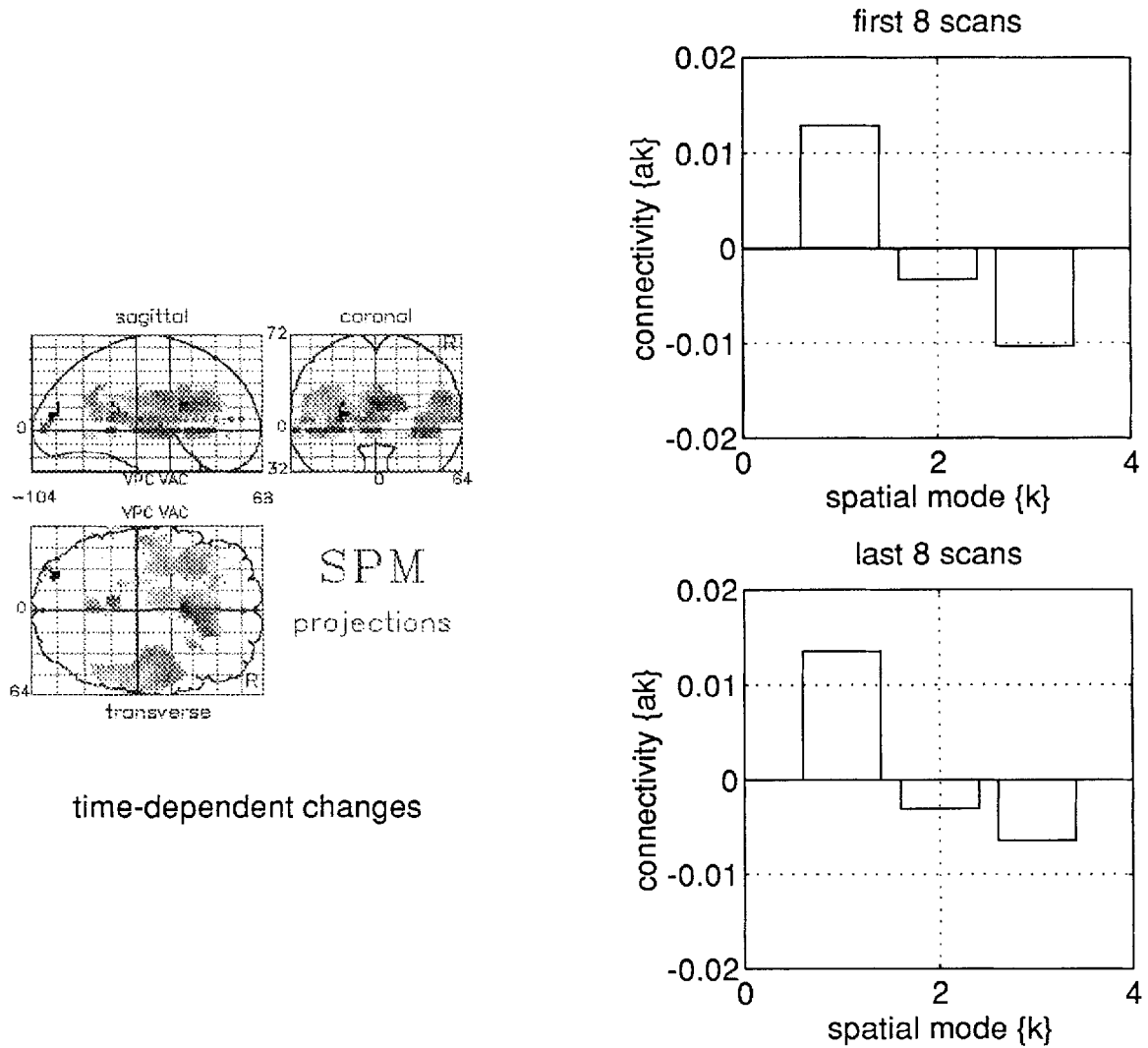


Fig. 7.

Time-dependent changes in effective connectivity. **Left:** Changes in effective connectivity between the estimates presented in Figure 6. The nature of these changes is characterized by the two graphs on the **right:** These represent the average connectivity from the first

three spatial modes (a_k in text). It can be seen that the first mode increases its contribution whereas the remaining two connectivities are attenuated. This is predicted by an associative model of changes in effective connectivity,

them should increase. To invoke synaptic plasticity as the underlying mechanism presupposes that two regions that have coincident high pool activity (reflected in their rCBF) contain neurons whose activity is highly coherent on a time scale of milliseconds: temporal summation of two EPSPs will only occur within 15 ms [Nelson et al., 1992]. This is important because this temporal summation may be instrumental in the induction of short and long term potentiation (STP and LTP).

Is it then reasonable to suppose that two regions with high pool activity will share a significant number

of neurons whose dynamic interactions occur within a time frame of milliseconds? We suggest it is. There are two lines of evidence in support of a cooperative interaction between fast dynamic correlations and slow covariation of pool activity: 1) Aertsen and Preissl [1991] have investigated the behavior of artificial networks, analytically and using simulations. They concluded that *short term effective connectivity* varies strongly with, or is modulated by, *pool activity*. Pool activity is the product of the number of neurons and their mean firing rate. The mechanism is simple; the efficacy of subthreshold EPSPs (excitatory post-

synaptic potentials) in establishing dynamic interactions is a function of postsynaptic depolarization, which in turn depends on the tonic background of activity. 2) The second line of evidence is experimental and demonstrates that the presence of fast interactions is associated with intermediate or long term correlations between distant neurons or neuronal groups. Nelson et al. [1992] have characterized effective connections between neurons or small groups of neurons, in BA 17 and BA 18 of cat extrastriate cortex. By cross-correlating activity they demonstrated that the most likely temporal relationship between spikes was a synchronous one. Furthermore the cross-correlograms segregated into three nonoverlapping groups with modal widths of 3 ms, 30 ms, and 400 ms. The short term correlation structures (3 and 30 ms) where almost always associated with the intermediate (400 ms) correlations. These observations suggest an interaction between short term (< 100 ms) and intermediate (100–1,000 ms) effective connectivity.

In summary co-activated regions will have increased rCBF and neuronal pool activity. Higher background discharge rates augment post-synaptic depolarization and susceptibility to fast dynamic interactions, STP and LTP, both within and between the regions co-activated. This sort of cooperative interaction between neuronal populations has been proposed as part of the theory of neuronal group selection [Edelman, 1978] and related work in this field makes explicit use of fast dynamic correlations [Sporns et al., 1989; Tononi et al., 1992].

Nonlinear models of effective connectivity

Reversible cooling experiments in monkey visual cortex, during visual stimulation, have demonstrated that neuronal activity in V2 depends on forward inputs from V1. Conversely neuronal activity in V1 is *modulated* by backward or reentrant connections from V2 to V1 [Schiller and Malpeli, 1977; Sandell and Schiller, 1982; Girard and Bullier, 1988]. In this section evidence is presented for a homologous asymmetry in reciprocal connections between V1 and V2 in human cortex, using physiological measurements obtained with functional MRI.

Nonlinear interactions between V1 and V2

Evidence for functional asymmetries is found in the work of Schiller and colleagues [Schiller and Malpeli, 1977; Sandell and Schiller, 1982] on interactions between V1 and V2. Retinotopically corresponding regions of V1 and V2 are reciprocally connected in the

monkey. V1 provides a crucial input to V2, in the sense that visual activation of V2 cells depends on input from V1. This dependency has been demonstrated by reversibly cooling (deactivating) V1 while recording from V2 during visual stimulation [Schiller and Malpeli, 1977; Girard and Bullier, 1988]. In contrast, cooling V2 has a more *modulatory* effect on V1 unit activity. “Most cells became less responsive to visual stimulation, while a few became more active during cooling.” The cells in V1 that were most affected by V2 deactivation were in the infragranular layers, suggesting V2 may use this pathway to modulate the output from V1 [Sandell and Schiller, 1982]. Similar conclusions about the return pathway between V5 and V2 were drawn by Girard and Bullier [1988]: because, in the absence of V1 input, these reentrant connections do not constitute an efficient drive to V2 cells, their role is most likely “to modulate the information relayed through area 17” (V1). The term *modulatory* is strictly functional and can be applied to either anatomically forward or backward connections.

A nonlinear model of effective connectivity

To examine the interactions between V1 and V2 we used a nonlinear model of effective connectivity, extended to include a modulatory interaction:

$$m_k^i = \Sigma(C_{ij}^O \cdot m_k^j + C_{ij}^M \cdot m_k^j m_k^i). \quad (21)$$

This model has two terms which allow for the activity in area i to be influenced by the activity in area j . The first represents an effect which depends only on afferent input from area j . This is the activity in j scaled by C_{ij}^O . The coefficient C_{ij}^O will be referred to as an *obligatory* connection strength, in the sense that a change in area j results in an obligatory response in area i . Conversely the second term reflects a modulatory influence of area j on area i . The coefficient determining the size of this effect (C_{ij}^M) will be referred to as a *modulatory* connection strength, because the overall effect depends on both the afferent input ($C_{ij}^O \cdot m_k^j$) and intrinsic activity (m_k^i).

This equation, or model, can be interpreted from two points of view; 1) by analogy with the nonlinear behaviour that characterizes voltage-dependent channels in electrophysiology or 2) in terms of classical (pharmacological) neuromodulation where post-synaptic responsiveness is modulated without a direct effect on post-synaptic membrane potential. The voltage-dependent analogy is obtained by considering m_k^i

as post-synaptic potential and m_k^i as a depolarizing current. According to Equation (21) a high C_{ij}^M reflects a *greater sensitivity to changes in input at higher levels of intrinsic activity*. In electrophysiological terms this translates as a change in post-synaptic depolarization, in response to a fixed depolarizing current, which increases with depolarization: This is a characteristic of voltage-dependent interactions [Haberly, 1991]. This intrinsic activity-dependent effect, determined by the value of C_{ij}^M , provides an intuitive sense of how to estimate C_{ij}^M . This estimation involves measuring the difference in sensitivity between states with high and low intrinsic activity at the location of interest:

Imagine one were able to “fix” the activity in V1 at a low level and measure the connectivity between V2 and V1 assuming a simple linear relationship [Eq. (14)]: a value for the sensitivity of V1 to V2 changes could be obtained, say C_1 . Now, if the procedure were repeated with V1 activity fixed at a high level, a second (linear) estimate would be obtained (C_2). In the presence of a substantial modulatory interaction between V2 and V1 the second estimate will be higher than the first. This is because the activity intrinsic to V1 is higher for the second estimate and V1 should be more sensitive to inputs from V2. In short $C_2 - C_1$ provides an estimate of the modulatory influence of V2 on V1. The activity of V1 can be fixed post hoc by simply selecting a subset of data in which the V1 activity is confined to some small range.

An estimation of the effective connection strengths of both an obligatory and modulatory nature can be obtained for the connections between *all* voxels and a reference location in the following way; for any reference location (i) assume that a subset of the time-series can be selected so that m_k^i is limited to some small range about its mean ($\langle m_k^i \rangle$). For this subset Equation (21) can be approximated by Equation (14) (omitting error terms for clarity):

$$\begin{aligned} m_k^i &\approx \sum C_{ij}^O m_k^j + \langle m_k^i \rangle \sum C_{ij}^M m_k^j \\ &\approx m^j \cdot (C^O + \langle m^i \rangle C^M) \\ &\approx m^j \cdot C. \end{aligned} \quad (22)$$

Where $C = C^O + \langle m^i \rangle C^M$. Now assume two such subsets are selected, one with a high mean ($\langle m^i \rangle_2$) and one with a low mean ($\langle m^i \rangle_1$) giving two solutions for C (C_2 and C_1), then:

$$C^M = (C_2 - C_1) / (\langle m^i \rangle_2 - \langle m^i \rangle_1) \propto \Delta C \quad (23)$$

(a similar expression for C^O can be derived—see Friston et al. [1994d]). ΔC is computed using exactly the same approach as in the section of time-dependent changes, namely using Equation (20). In the present application subsets 1 and 2 are subsets of the time-series selected on the basis of intrinsic activity (m_k^i) being high or low. This general approach to characterizing nonlinear systems with a piece-wise series of locally linear models has proved a fruitful strategy in many instances [Tsonis, 1992]. See Palus et al. [1991] for a conceptually related approach to multichannel EEG recordings.

The hypothesis that asymmetrical nonlinear V1 – V2 interactions would characterize cortical interactions in human visual cortex can be formulated in terms of C_{ij}^M : We predicted that 1) the modulatory component of effective connections to V1 would be regionally specific and include V2 ($C_{V1,V2}^M$ would be relatively high compared to other regions), and that 2) the forward modulatory influences from V1 to V2 ($C_{V2,V1}^M$) would be smaller than the reciprocal influences ($C_{V1,V2}^M$).

Regional specificity of modulatory connections to V1

A reference voxel was chosen in V1, according to the atlas of Talairach and Tournoux [1988], and the effective connection strengths $C_{V1,j}^O$ and $C_{V1,j}^M$ were estimated as described above. Values for $C_{V1,j}^O$ and $C_{V1,j}^M$ were generated for all voxels, allowing maps of $C_{V1,j}^O$ and $C_{V1,j}^M$ to be constructed. These maps provided a direct test of the hypothesis concerning the topography and regional specificity of modulatory influences on V1. The lower rows in Figure 8 are maps of $C_{V1,j}^O$ and $C_{V1,j}^M$ (for a reference in V1 on the right) and reflect the degree to which the area exerts an obligatory (left) or modulatory (right) effect on V1 activity. These maps have been thresholded at 1.64 after normalization to a standard deviation of unity. This corresponds to an uncorrected threshold of $P = 0.05$.

The obligatory connections to the reference voxel derive mainly from V1 itself, both ipsilaterally and contralaterally with a small contribution from contiguous portions of V2. The effective connectivity from contralateral V1 should not be overinterpreted given that 1) the source of many afferents to V1 (the lateral geniculate nuclei) where not included in the field of view and that 2) this finding can be more parsimoniously explained by “common input.” As predicted, and with remarkable regional specificity, the modulatory connections were most marked from ipsilateral

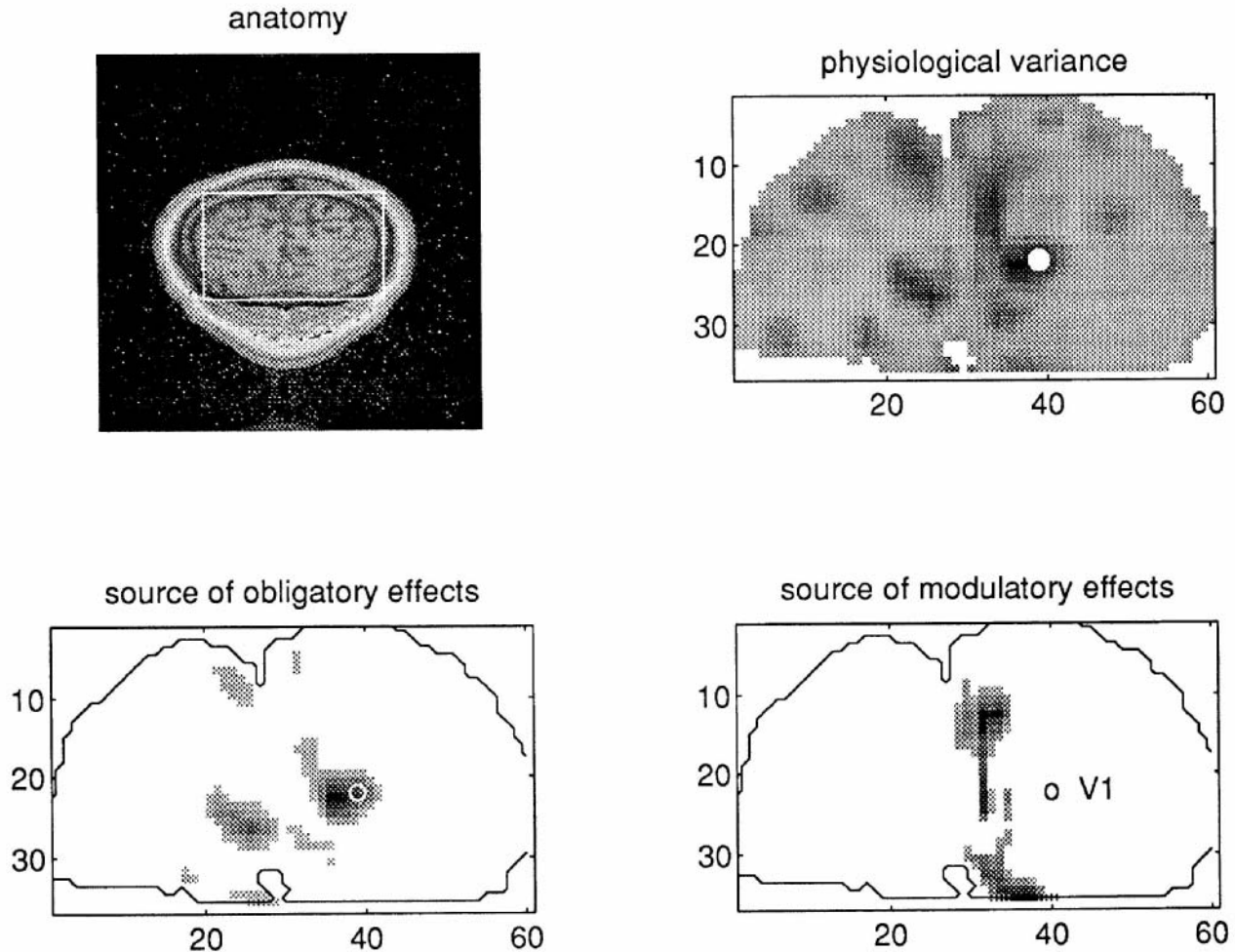


Fig. 8.

Maps of the estimates of obligatory and modulatory connection strengths to V1. **Top left:** Anatomical features of the coronal data used. This image is a high resolution anatomical MRI scan of the subject that corresponds to the fMRI slices. The box defines the position of a (36×60 voxel) subpartition of the fMRI time-series selected for analysis. **Top right:** The location of the reference voxel designated as V1 (white dot). This location is shown on a statistical parametric map of physiological variance (calculated for

each voxel from the time-series of 60 scans). The image has been scaled to its maximum. **Lower right and lower left:** Maps of $C\hat{V}_{i,j}$ and $C\hat{M}_{i,j}$. The images have been scaled to unit variance and thresholded at $P = 0.05$ (assuming, under the null hypothesis of no effective connectivity, the estimates have a Gaussian distribution). The reference voxel in V1 is depicted by a circle. The key thing to note is that V1 is subject to modulatory influences from ipsilateral and extensive regions of V2.

V2, dorsal and ventral to the calcarine fissure (Brodmann's area 18 according to the atlas of Talairach and Tournoux [1988]) (note that "common input" cannot explain interactions between V1 and V2 because the geniculate inputs are restricted to V1).

To address the functional asymmetry hypothesis the modulatory connection strengths between two extended regions (two 5×5 voxel squares) in ipsilateral V1 and V2 were examined. The estimates of effective connection strengths were based on hemodynamic changes in all areas and the subset of connec-

tions between the two regions were selected to compare the distributions of forward and backward modulatory influences. Figure 9 shows the location of the two regions (this time on the right) and the frequency distribution of the estimates for connections from the V1 box to the V2 box (broken line) and the corresponding estimates for connections from V2 to V1 (solid line). There is a remarkable dissociation, with backward modulatory effects (V2 to V1) being much greater than forward effects (V1 to V2). This can be considered a confirmation of the asymmetry hypothesis.

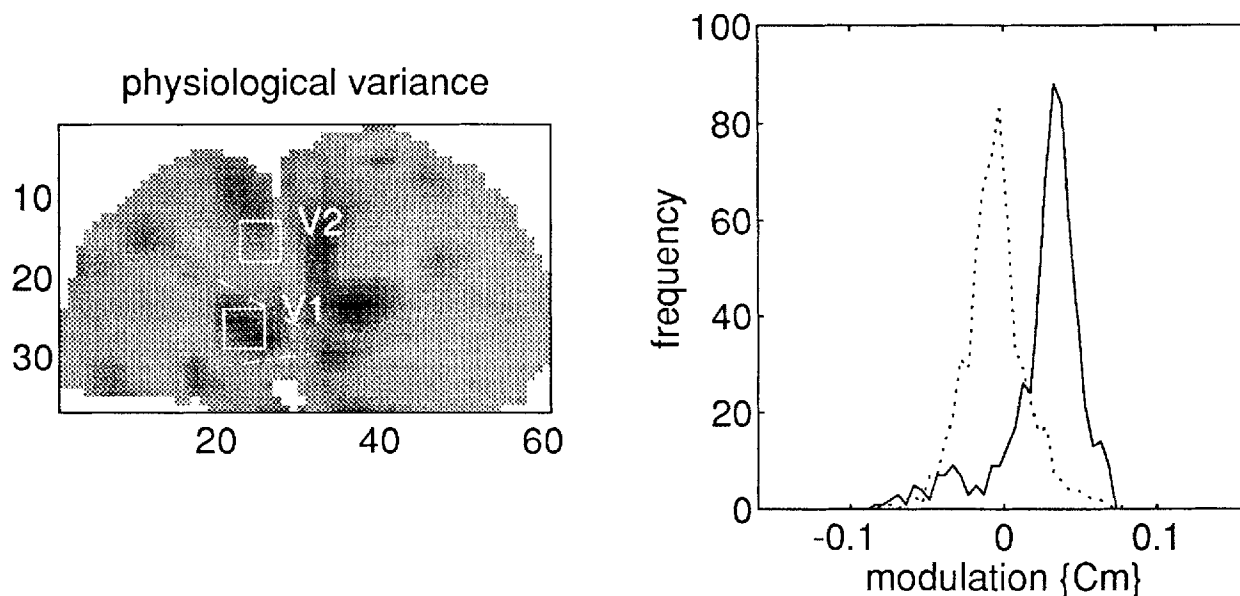


Fig. 9.

Graphical presentation of a direct test of the hypothesis concerning the asymmetry between forward and backward V1 - V2 interactions. **Left:** A map of physiological variance showing the positions of two boxes defining regions in V1 and V2. The broken lines correspond (roughly) to the position of the V1/V2 border according to the atlas of Talairach and Tournoux [1988]. The value of C_{ij}^M

were computed for all voxels in either box and Euclidean normalized to unity over the image. The frequency distribution of C_{ij}^M connecting the two regions is presented on the **right**. The backward connections (V2 to V1, solid line) are clearly higher than the corresponding forward connections (V1 to V2, broken line).

The biological mechanisms of modulatory dynamics

The measurements used in this study were *hemodynamic* in nature. This limits an interpretation at the level of *neuronal* interactions. However the analogy between the form of the nonlinear interaction between V1 and V2 activity and voltage-dependent connections is a strong one. It is possible that the modulatory impact of V2 on V1 is mediated by predominantly voltage-dependent connections. The presence of horizontal voltage-dependent connections within V1 has been established in cat striate cortex [Hirsch and Gilbert, 1991]. We know of no direct electrophysiological evidence to suggest that extrinsic backward V2 to V1 connections are voltage-dependent; however our results are consistent with this. An alternative explanation for modulatory effects, which does not necessarily involve voltage-dependent connections, can again be found in the work of Aertsen and Preissl [1991] cited in the previous section. Recall that they concluded effective connectivity varies strongly with, or is modulated by, pool activity. The mechanism related to the efficacy of subthreshold EPSPs in establishing dynamic interactions as a function of post-synaptic depolarization,

which in turn depends on the tonic background of activity. This clearly relates to the idea that sensitivity to afferent input increases with intrinsic activity (the original presentation of these results and a fuller discussion can be found in Friston et al. [1994d]).

CONCLUSION

This paper has reviewed the basic distinction between functional and effective connectivity in neuroimaging. Emphasis has been placed the points of contact between the diverse applications of these concepts and in particular the central role of eigenimages or spatial modes. The final sections demonstrate that nonlinear interactions can be characterized using simple extensions of linear models.

Most of the techniques presented here are both simple and established. Many have been used in the context of neuroimaging before. One important difference between the applications in this paper and other applications is that analyses such as eigenvector solutions, multidimensional scaling and linear regression approaches to effective connectivity are applied here to neuroimaging *time-series*. In this sense there is a greater correspondence between the current work

and the analysis of electrophysiological time-series than between this and previous analyses of covariances in neuroimaging data (where the covariances are across subjects not time). One might anticipate that the time-series perspective will dominate with the advent of functional MRI and other techniques that allow many repeated scans in quick succession.

An important development in recent years has been the use of structural equation modelling to examine interactions between brain regions [see McIntosh and Gonzalez-Lima, this issue]. In the context of the present framework structural equation modelling can be thought of as a (linear regression) model for effective connectivity. Structural equation modelling is particularly interesting in the sense that different models can be compared statistically: Furthermore control can be exerted over constraints on the model by specifying the sparsity structure of the connections and error variance.

Although a less than mature field the approach to neuroimaging data and regional interactions discussed by the many authors in these pages is an exciting endeavour that is starting to attract more and more attention.

REFERENCES

- Aertsen A, Preissl H (1991): Dynamics of activity and connectivity in physiological neuronal networks. In: Schuster HG (ed): *Non Linear Dynamics and Neuronal Networks*. New York: VCH Publishers Inc., pp 281–302.
- Binmore KG (1982): *Mathematical analysis*. 2nd ed. Cambridge: Cambridge University Press, p 218.
- Bleuler E (1913): *Dementia praecox or the group of schizophrenias*. Translated into English 1987. In: Cutting J and Shepherd M (eds): *The Clinical Routes of the Schizophrenia Concept*. Cambridge: Cambridge University Press.
- Corbetta M, Miezin FM, Dobmeyer S, Shulman GL, Petersen SE (1991): Selective and divided attention during visual discrimination of shape, color and speed. *Functional anatomy by positron emission tomography*. *J Neurosci* 11(8):2383–2402.
- Edelman GM (1978): Group selection and phasic reentrant signaling: A theory of higher brain function. In: Edelman GM, Mountcastle VB (eds): *The Mindful Brain*. MIT Press, pp 55–100.
- Fox PT, Mintun MA (1989): Non-invasive functional brain mapping by change distribution analysis of averaged PET images of $H^{15}O_2$ tissue activity. *J Nucl Med* 30:141–149.
- Friedrich R, Fuchs A, Haken H (1991): Modelling of spatio-temporal EEG patterns. In: Dvorak I and Holden AV (eds): *Mathematical Approaches to Brain Functioning Diagnostics*. New York: Manchester University Press.
- Friston KJ, Passingham RE, Nutt JG, Heather JD, Sawle GV, Frackowiak RSJ (1989): Localization in PET images: Direct fitting of the intercommissural (AC-PC) line. *J Cereb Blood Flow Metab* 9:690–695.
- Friston KJ, Frith CD, Liddle PF, Dolan RJ, Lammertsma AA, Frackowiak RSJ (1990): The relationship between global and local changes in PET scans. *J Cereb Blood Flow Metab* 10:458–466.
- Friston KJ, Frith CD, Liddle PF, Frackowiak RSJ (1991a): Plastic transformation of PET images. *J Comput Assist Tomogr* 15:634–639.
- Friston KJ, Frith CD, Liddle PF, Frackowiak RSJ (1991b): Comparing functional (PET) images: The assessment of significant change. *J Cereb Blood Flow Metab* 11:690–699.
- Friston KJ, Frith CD, Liddle PF, Frackowiak RSJ (1993a): Functional connectivity: The principal component analysis of large (PET) data sets. *J Cereb Blood Flow Metab* 13:5–14.
- Friston KJ, Frith CD, Frackowiak RSJ (1993b): Time-dependent changes in effective connectivity measured with PET. *Hum Brain Mapp* 1:69–80.
- Friston KJ, Jezzard P, Frackowiak RSJ, Turner R (1993c): Characterizing focal and distributed physiological changes with MRI and PET. In: *Functional MRI of the Brain*. Berkeley, CA: Society of Magnetic Resonance in Medicine, pp 207–216.
- Friston KJ, Herold S, Fletcher P, Silbersweig D, Cahill C, Dolan RJ, Liddle PF, Frackowiak RSJ, Frith CD (1994a): Abnormal fronto-temporal interactions in schizophrenia. In: Watson SJ (ed): *Biology of Schizophrenia and Affective Disease*. ARNMD Series Vol. 73, in press.
- Friston KJ, Liddle PF, Frith CD (1994b): Dysfunctional fronto-temporal integration in schizophrenia. *Neuropsychopharmacology* 10:538S.
- Friston KJ, Frith CD (1994c): Abnormal inter-hemispheric integration in schizophrenia: An analysis of neuroimaging data. *Neuropsychopharmacology* 10:719S.
- Friston KJ, Ungerleider LG, Jezzard P, Turner R (1994d): Characterizing modulatory interactions between V1 and V2 in human cortex with fMRI. Submitted.
- Fuchs A, Kelso JAS, Haken H (1992): Phase transitions in the human brain: Spatial mode dynamics. *Int J Bifurcation Chaos* 2:917–939.
- Gerstein GL, Perkel DH (1969): Simultaneously recorded trains of action potentials: Analysis and functional interpretation. *Science* 164:828–830.
- Gerstein GL, Bedenbaugh P, Aertsen AMHJ (1989): Neuronal assemblies. *IEEE Trans Biomed Eng* 36:4–14.
- Girard P, Bullier J (1988): Visual activity in area V2 during reversible inactivation of area 17 in the macaque monkey. *J Neurophysiol* 62:1287–1301.
- Gochin PM, Miller EK, Gross CG, Gerstein GL (1991): Functional interactions among neurons in inferior temporal cortex of the awake macaque. *Exp Brain Res* 84:505–516.
- Goldman-Rakic PS (1986): Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. In: Mountcastle VB (ed): *Handbook of Physiology*. Vol. 5. Baltimore: Williams and Wilkins, pp 373–417.
- Goldman-Rakic PS (1988): Topography of cognition: Parallel distributed networks in primate association cortex. *Ann Rev Neurosci* 11:137–156.
- Golub GH, Van Loan CF (1991): *Matrix computations*. 2nd ed. Baltimore and London: The Johns Hopkins University Press, pp 241–248.
- Gower JC (1966): Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* 53:325–328.
- Haberly LB (1991): Olfactory cortex. In: Shepard GM (ed): *The synaptic organization of the brain*. Oxford: Oxford University Press, p 331.
- Hirsch JA, Gilbert CD (1991): Synaptic physiology of horizontal connections in the cat's visual cortex. *J Neurosci* 11:1800–1809.

- Horwitz B (1990): Simulating functional interactions in the brain: A model for examining correlations between regional cerebral metabolic rates. *Int J Biomed Comput* 26:149–170.
- Horwitz B, Grady C, Haxby J, Schapiro M, Carson R, Herscovitch P, Ungerleider L, Mishkin M, Rapoport S (1991): Object and spatial visual processing: Intercorrelations of regional cerebral blood flow among posterior brain regions. *J Cereb Blood Flow Metab* 11:(Suppl 2):S380.
- Keys RG (1981): Cubic convolution interpolation for digital image processing. *IEEE Trans Acoustic, Speech and Signal Processing* 29:1153–1160.
- Kohonen (1982): Self organized formation of topologically correct feature maps. *Biol Cybern* 43:59–69.
- Lagreze HL, Hartmann A, Anzinger G, Shaub A, Deister A (1993): Functional cortical interaction patterns in visual perception and visuospatial problem solving. *J Neurol Sci* 114:25–35.
- Lopez HS, Burger B, Dickstein R, Desmond NL, Levy WB (1990): Associative synaptic potentiation and depression: Quantification of dissociable modifications in the hippocampal dentate gyrus favors a particular class of synaptic modification equations. *Synapse* 5:33–47.
- Mayer-Kress G, Barczys C, Freeman W (1991): Attractor reconstruction from event-related multi-electrode EEG data. In: Dvorak I and Holden AV (eds): *Mathematical Approaches to Brain Functioning Diagnostics*. New York: Manchester University Press.
- McIntosh AR, Gonzalez-Lima F (1991): Structural modeling of functional neural pathways mapped with 2-deoxyglucose: Effects of acoustic startle habituation on the auditory system. *Brain Res* 547:295–302.
- McIntosh AR, Grady CL, Ungerleider LG, Haxby JV, Rapoport SI, Horwitz B (1993): Network analysis of cortical visual pathways mapped with PET. *J Neurosci* 14:655–666.
- Moeller JR, Strother SC, Sidtis JJ, Rottenberg DA (1987): Scaled subprofile model: A statistical approach to the analysis of functional patterns in positron emission tomographic data. *J Cereb Blood Flow Metab* 7:649–658.
- More JJ (1977): The Levenberg-Marquardt algorithm: Implementation and Theory. In: Watson GA (ed): *Numerical Analysis. Lecture notes in mathematics* 630. Springer Verlag, pp 105–116.
- Nelson JL, Salin PA, Munk NMJ, Arzi M, Bullier J (1992): Spatial and temporal coherence in cortico-cortical connections: A cross-correlation study in areas 17 and 18 in the cat. *Vis Neurosci* 9:21–37.
- Newton I (1704): *Opticks*. Book 1, part 2, prop. 6 London: Smith and Walford.
- Oja E (1989): Neural networks, principal components, and subspaces. *Int J Neural Syst* 1:61–68.
- Pardo JV, Pardo PJ, Janer KW, Raichle ME (1990): The anterior cingulate cortex mediates processing selection in the Stroop attentional; conflict paradigm. *Proc Natl Acad Sci USA* 87:256–259.
- Palus M, Dvorak I, David I (1991): Remarks on spatial and temporal dynamics of EEG. In: Dvorak I, Holden AV (eds.): *Mathematical approaches to brain imaging diagnostics*. Manchester and New York: Manchester University Press, pp 369–385.
- Petersen SE, Fox PT, Posner ML, Mintun M, Raichle ME (1989): Positron emission tomographic studies of the processing of single words. *J Cog Neurosci* 1:153–170.
- Posner ML, Sandson J, Dhawan M, Shulman GL (1990): Is word recognition automatic? A cognitive-anatomical approach. *J Cog Neurosci* 1:50–60.
- Sandell JH, Schiller PH (1982): Effect of cooling area 18 on striate cortex cells in the squirrel monkey. *J Neurophysiol* 48:38–38.
- Schiller PH, Malpeli JG (1977): The effect of striate cortex cooling on area 18 cells in the monkey. *Brain Res* 126:366–369.
- Shepard RN (1980): Multidimensional scaling, tree-fitting and clustering. *Science* 210:390–398.
- Spinks TJ, Jones T, Bailey DL, Townsend DW, Grootnook S, Bloomfield PM, Gilardi MC, Casey ME, Sipe B, Reed J (1992): Physical performance of a positron tomograph for brain imaging with retractable septa. *Phys Med Biol* 37:1637–1655.
- Sporns O, Gally JA, Reeke GN, Edelman GM (1989): Reentrant signalling among simulated neuronal groups leads to coherence in their oscillatory activity. *Proc Natl Acad Sci USA* 86:7265–7269.
- Talairach P, Tournoux J (1988): *A stereotactic coplanar atlas of the human brain*. Stuttgart: Thieme.
- Tononi G, Sporns O, Edelman GM (1992): Reentry and the problem of integrating multiple cortical areas: simulation of dynamic integration in the visual system. *Cerebral Cortex* 2:310–335.
- Torgerson WS (1958): *Theory and methods of scaling*. New York: Wiley.
- 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* M1–10:505–512.
- Tsonis AA (1992): *Chaos: From theory to applications*. New York and London: Plenum Press, p 221.
- Wise SP (1989): Frontal cortical activity and motor set. In: Ito M (ed): *Neural Programming*. Tokyo: Japanese Scientific Societies Press. p 26.