Multivariate Statistical Analyses for Neuroimaging Data

Anthony R. McIntosh and Bratislav Mišić

Rotman Research Institute, Baycrest, Toronto, Ontario, Canada, M6A 2E1; email: rmcintosh@rotman-baycrest.on.ca

Annu. Rev. Psychol. 2013. 64:499–525

First published online as a Review in Advance on July 12, 2012

The *Annual Review of Psychology* is online at psych.annualreviews.org

This article's doi: 10.1146/annurev-psych-113011-143804

Keywords

multivariate, functional connectivity, effective connectivity, network

Abstract

As the focus of neuroscience shifts from studying individual brain regions to entire networks of regions, methods for statistical inference have also become geared toward network analysis. The purpose of the present review is to survey the multivariate statistical techniques that have been used to study neural interactions. We have selected the most common techniques and developed a taxonomy that instructively reflects their assumptions and practical use. For each family of analyses, we describe their application and the types of experimental questions they can address, as well as how they relate to other analyses both conceptually and mathematically. We intend to show that despite their diversity, all of these techniques offer complementary information about the functional architecture of the brain.

Contents

INTRODUCTION

The complex anatomical connectivity of the central nervous system suggests that interregional communication is a primary function. The notion that activity in individual regions is influenced by activity in other regions via direct or indirect projections has made the network organization of the brain a fundamental theme in neuroscience. A general systems theory has emerged that simultaneously emphasizes specialized local computation (functional specialization) and computation arising from interactions between regions (functional integration). Functional integration implies that the response properties of any one region must be studied with respect to the status of other regions in the network (neural context) (Bressler & McIntosh 2007; McIntosh 1998, 2000).

Univariate Versus Multivariate

Neuroimaging data mirror the complexity of the brain in the sense that signals can be recorded from a large number of spatially distributed sensors (e.g., voxels, electrodes) and often at multiple time points. Unsurprisingly, methods for statistical analysis of neuroimaging data have developed along distinct lines that focus either on functional specialization or integration.

The conventional "mass univariate" approach is optimal for identifying reliable task-dependent signal changes at the level of individual image elements (Friston et al. 1991). The general linear model (GLM) of the activity in each voxel ($\mathbf{Y}_{n\times 1}$) in n scans is a linear sum ($\boldsymbol{\beta}_{p\times 1}$) of p predictors ($\mathbf{X}_{n\times p}$) plus a residual error term, where each predictor is the time course of some effect

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}. \quad (1)$$

For instance, the predictor for the effect of a task on voxel activity is a vector of 1's at the time points when the task is on and 0's everywhere else. Ordinary least squares regression is used to estimate the β weights for each predictor. The weights indicate the contribution of the predictor to the variance of the voxel response. Depending on whether one or more contrasts are tested, either a t or F statistic is computed for each β weight. Critically, this type of analysis is performed separately for every image element, explicitly precluding the possibility that responses arise from coordinated dynamics among other elements. Conversely, multivariate statistical analyses take advantage of the spatial and temporal dependencies among image elements, enabling inference across space and/or time (Petersson et al. 1999). In the most general case, multivariate analyses enable one to capture spatiotemporal patterns of activity (e.g., time-varying networks) that have some functional significance (e.g., relate to a task contrast or behavior).

Exploratory Versus Confirmatory

Multivariate analyses can be broadly divided into those that are exploratory and those that are confirmatory. Exploratory techniques are primarily used to identify robust patterns of covarying neural activity [principal component analysis (PCA) and independent component

analysis (ICA)] and possibly relating these patterns to design variables and/or behavior [canonical correlation analysis (CCA), canonical variate analysis (CVA), partial least squares (PLS)]. They are usually data driven, and no explicit hypothesis needs to be specified about the contribution of individual brain regions or about the differentiation of conditions or groups. However, this does not imply that exploratory techniques are noninferential, and in most cases statistical inference is possible. In confirmatory analyses, an explicit model of regional interactions is formulated and tested to see whether it fits the data and/or whether it fits the observed data better than alternative models [structural equation modeling (SEM) and dynamic causal modeling (DCM)]. Thus, confirmatory techniques are used to test specific hypotheses.

The distinction between exploratory and confirmatory techniques parallels the distinction between functional and effective connectivity. Functional connectivity refers to statistical interdependencies between two remote regions (Friston et al. 1993). A functional connection between two regions does not imply that they are communicating directly, as their covariation could be due to common inputs from another source. Effective connectivity is defined as a directed causal influence of one region on another (Aertsen et al. 1989, Friston et al. 1993). Exploratory analyses extract distinct components of the covariance structure of the data. The networks of regions that they identify strongly covary with each other and possibly with some task effect and are interpreted as functional networks. Confirmatory analyses specify a model of interactions among units that takes into account external inputs and the anatomical substrate, and they are interpreted in terms of effective connectivity.

Multivariate Granger causality and the graph model are two techniques that do not perfectly fit in either category because they have both exploratory and confirmatory characteristics. For instance, Granger causality is a measure of causal influence, but it can be applied in an exploratory fashion to any number of pairs of regions. Likewise, the use of graph theoretic metrics is exploratory in the sense that they are usually calculated for all regions with no specific model, yet the measures used to estimate the connectivity of the graph may be measures of effective connectivity.

In the present article we give an overview of the most common strategies for multivariate analysis of neuroimaging data. By developing a simple taxonomy we hope to illustrate the potential applications of various methods and how they relate to each other. These techniques can be applied to most neuroimaging modalities, such as positron emission tomography (PET), near-infrared spectroscopy, functional magnetic resonance imaging (fMRI), local field potentials, electroencephalography (EEG), and magnetoencephalography (MEG).

PCA: principal component analysis

ICA: independent component analysis

CCA: canonical correlation analysis

PLS: partial least squares

SEM: structural equation modeling

DCM: dynamic causal modeling

Effective connectivity: a directed interaction between two regions

Functional connectivity: a systematic deviation from statistical independence between neural activity in two regions, which can be computed using any measure of association, such as correlation, phase locking, or mutual information

SVD: singular value decomposition

EXPLORATORY TECHNIQUES

Principal Component Analysis

The goal of PCA is to factorize a data matrix with many variables by creating a set of new variables, termed principal components (Pearson 1901). Each component is a linear combination of the original variables, and the weights are chosen such that the first component has the greatest possible variance, while each successive component also has the highest possible variance, under the condition that all components are mutually uncorrelated. This constraint ensures that the components capture unique, nonoverlapping portions of variance and that together they can perfectly reproduce the variance-covariance structure of the original data set. In the context of neuroimaging, PCA can be used to summarize a data set of several thousand voxels into a smaller number of principal components that may be interpreted as functional networks (Friston et al. 1993, Moeller & Strother 1991, Strother et al. 1995).

PCA is performed by subjecting a data matrix $\mathbf{X}_{n\times p}$ (n observations in the rows, p variables in the columns) to singular value decomposition (SVD) (Eckart & Young 1936). Although SVD is not a statistical analysis per

se, it is the engine behind PCA-, PLS-, and CCA-based techniques. The main difference between these techniques is what the matrix **X** represents (Worsley et al. 1997). SVD is an algebraic tool to deconstruct any given matrix **X** into its basic structure

$$\mathbf{X} = \mathbf{USV}', \tag{2}$$

where $\mathbf{U}_{n\times n}$ and $\mathbf{V}_{p\times n}$ are orthonormal matrices with rank equal to the original matrix **X,** and $\mathbf{S}_{n\times n}$ is a diagonal matrix of positive singular values. The vectors of **U** and **V** are termed the left and right singular vectors, respectively. An alternative way to perform PCA is to apply a spectral or eigen-decomposition to the covariance matrices of **X**. The eigen-decomposition of the voxel × voxel covariance matrix $\mathbf{X}'\mathbf{X}_{p\times p}$ yields **V**

$$\mathbf{X}'\mathbf{X} = (\mathbf{VS}'\mathbf{U}')(\mathbf{USV}') = \mathbf{V}(\mathbf{S}'\mathbf{S})\mathbf{V}', \tag{3}$$

while the spectral decomposition of the scan × scan covariance matrix $\mathbf{XX}'_{n\times n}$ yields **U**

$$\mathbf{XX}' = (\mathbf{USV}')(\mathbf{VS}'\mathbf{U}') = \mathbf{U}(\mathbf{SS}')\mathbf{U}'. \tag{4}$$

Both cases are made simpler by the fact that $\mathbf{U}'\mathbf{U} = \mathbf{I}$ and $\mathbf{V}'\mathbf{V} = \mathbf{I}$ because **U** and **V** are orthonormal. The vectors of **V** can be referred to either as the right singular vectors of **X** or the eigenvectors of **X′X**. The vectors of **U** can be referred to either as the left singular vectors of **X** or the eigenvectors of **XX′**. The squared singular value elements on the diagonals of the matrix **S′S** are also referred to as eigenvalues.

The rank of **X** is the smaller of its row and column dimensions and will determine how many components are extracted. In the illustration above, where **X** is a scan × voxel matrix, the total number of components will typically be equal to the number of scans rather than the number of voxels. In the case for group analysis, where **X** is a subjects × voxels matrix, the number of components will most typically be equal to the number of subjects. The *i*th principal component consists of the *i*th column vector of **U**, the *i*th column vector of **V**, and the *i*th singular value element of **S**. The vector **V**(*i*) contains weights that indicate the degree to which each voxel contributes to the component. Thus, **V**(*i*) is a spatial image (eigenimage) of that principal component. The corresponding vector **U**(*i*) also contains weights, but these indicate the degree to which the component is expressed in each scan (eigentimeseries, when the rows of **X** are scans). The contribution of the original variables to the component is often assessed in terms of loadings, but this term is ambiguous. It may refer to either the eigenvector weights or to the correlation between the original variables and components across the observations. The squared singular value associated with each component is proportional to the portion of variance accounted for by the component. The components are ordered by the magnitude of their squared singular values, from largest to smallest. Component scores indicate the value that each of the original *n* observations would have on the new variables. They are calculated by postmultiplying the data matrix by the eigenimage weights ($\mathbf{XV}_{n\times n}$).

The ability to concentrate as much variance as possible in as few components as possible makes PCA widely applicable in neuroimaging. The initial decomposition reduces the dimensionality of the data set from *p* (e.g., thousands of voxels) to *n* (e.g., hundreds of scans). Dimensionality can be reduced further by selecting the first *k* components and subtracting the variance associated with the rest, although there is no single best criterion to determine *k*. One option is to select the number of components that account for some minimum amount of total variance. This is a useful method of reducing dimensionality for computationally expensive techniques such as ICA. Another possible criterion is to keep all components up to the first discontinuity in the Scree plot of eigenvalues. Still another approach is to choose the number of components with respect to the generalizability of the decomposition (Hansen et al. 1999).

PCA aggregates variance in components by appropriately weighing original variables that tend to covary. As a result, PCA isolates prominent patterns of regional covariation that represent functional interactions (Friston et al. 1993). Although PCA is usually performed only on the neuroimaging data matrix without any

reference to the experimental design, functional networks extracted using PCA can be submitted to another analysis to test for task or group differences, such as CVA (Friston et al. 1996).

Scaled subprofile model. The scaled subprofile model (SSM) is an extension of PCA whereby the expression of dominant patterns of functional connectivity can be compared across groups or for individual subjects (Alexander & Moeller 1994, Moeller et al. 1987, Moeller & Strother 1991, Strother et al. 1995). In the SSM framework it is assumed that there exist meaningful patterns of regional interactions that are independent of the global mean activity and that their expression may differ across groups. To this end, SSM first attempts to remove the main effects of subject and region. The residual values in the data matrix, bereft of these main effects, should contain meaningful information about regional covariance patterns.

Data from multiple groups are combined in a single data matrix $\mathbf{X}_{n\times p}$ with n subjects in the rows and p voxels in the columns. The main difference between SSM and conventional PCA lies in the initial normalization. In PCA, $\mathbf{X}_{n\times p}$ is column centered by subtracting the mean of the column from each element in that column. In SSM, each element of $\mathbf{X}_{n\times p}$ is natural log transformed and then mean centered twice. First, the mean activity for each region across subjects is computed and subtracted (column centering). Second, the mean activity across all regions for each subject is computed and subtracted (row centering).

The twice-normalized data matrix is subjected to SVD and the results are interpreted in a manner identical to traditional PCA. The left singular vectors $\mathbf{V}(i)$ index the contribution of individual voxels to the topography of the component (group invariant subprofiles). As the analysis is blind to group membership, these profiles are group invariant, and the only way to differentiate groups is in terms of their expression. The degree to which a given pattern of regional covariation manifests in individual subjects is given by the subject-specific weights contained in the right singular vectors $\mathbf{U}(i)$ [subject scaling factor (SSF)].

As in traditional PCA, SSM reduces a large multivariate data set with thousands of voxel variables to a smaller number of topographic patterns. To investigate group and subject differences, SSFs for a specific pattern can be entered into a univariate analysis such as an analysis of variance. SSFs for a patient group can also be entered into a multivariate regression to investigate whether there exists any combination of functional connectivity patterns that predicts some specific behavioral trait.

Applications. Alexander et al. (1994) used SSM to investigate PET functional networks affected by Alzheimer's disease. They analyzed patients and healthy controls in the same step and found four consistent spatial patterns. The mean SSFs differed for the two groups on two of the four patterns, indicating that the expression of these patterns was significantly affected by the disease process. The first pattern indicated lower regional cerebral metabolism in bilateral parietal cortex and right superior temporal cortex and lower metabolism in anterior cingulate and orbital frontal cortex. Moreover, the patient SSFs correlated negatively with measures of visuospatial and attentional skills, indicating that the expression of the parietotemporal deficit pattern is related to poorer attentional performance. The second group-sensitive pattern revealed lower metabolism in the prefrontal, premotor, inferior parietal, medial temporal, and insular regions, as well as high metabolism in occipital association and calcarine regions. This pattern was related to deficits in verbal memory, language comprehension, and verbal fluency. Altogether, the analysis identified two functional networks affected by the disease with distinct behavioral consequences.

Independent Component Analysis

ICA is similar to PCA in the sense that both techniques seek to represent a large number of variables (e.g., voxels) in terms of a smaller

number of dimensions that can be interpreted as cohesive functional networks (Beckmann & Smith 2004; Calhoun et al. 2001a,b; McKeown et al. 1998a,b, 2003). Whereas principal components are assumed to be mutually uncorrelated both spatially and temporally, independent components are maximally statistically independent only in one domain.

In both PCA and ICA, the objective is to choose components that have minimal interdependencies among each other. In PCA, a new set of variables, which can be thought of as a set of axes, is created and oriented such that the projection of the data on the first axis has the greatest variance, the projection on the second axis has the second greatest variance, and so on. The key characteristics are that the axes are orthogonal and the components they represent are uncorrelated. The idea behind ICA is that the independent components were somehow mixed to give rise to the observed variables [e.g., blood oxygen–level dependent (BOLD) signal]. In other words, the neural activity measured in different voxels can be thought of as a linear combination of a smaller number of underlying independent component sources. By the central limit theorem, any linear mixture of independent variables (e.g., voxels) will be more "Gaussian" than the original variables themselves (e.g., independent components). Thus, ICA also seeks to create a new set of axes in which to represent the original data set, but the axes are oriented such that the projection of data points onto the axes is maximally non-Gaussian. Therefore, the new axes need not be orthogonal, and independent components can be said to be maximally statistically independent, but they are not necessarily uncorrelated like principal components.

In neuroimaging, independence can be imposed either in the spatial domain (spatial ICA) or in the temporal domain (temporal ICA). The choice between the two depends on the assumptions of the investigator and on the characteristics of the data set (McKeown et al. 2003). Spatial ICA is more commonly used for fMRI data because task-dependent activations are assumed to be relatively sparse in a volume of several thousand voxels. As a result, independent components isolate networks of coherent regions that overlap as little as possible. Conversely, temporal ICA is more commonly used for event-related potential (ERP) data because scalp measurements are assumed to be a mixture of several coactive sources (Makeig et al. 1997, 1999). Thus, the components may have overlapping topographies, but they should have distinct time courses, and so it is desirable for the components to be temporally independent.

The generative model for ICA can be written as

$$\mathbf{X}' = \mathbf{AS}. \tag{5}$$

Here $\mathbf{X}_{n\times p}$ is the observed data matrix with n scans in the rows and p voxels in the columns. $\mathbf{S}_{r\times n}$ is the source matrix with r sources and their activity in each of the n scans. Spatial ICA assumes the rows of $\mathbf{S}_{r\times n}$ are independent, whereas temporal ICA assumes the columns of $\mathbf{S}_{r\times n}$ are independent. The mixing matrix $\mathbf{A}_{p\times r}$ indicates how the sources were combined to produce the observed data. ICA is an iterative algorithm that simultaneously tries to estimate both $\mathbf{A}$ and $\mathbf{S}$ by maximizing the non-Gaussianity of either the rows or columns of $\mathbf{S}$. The final step of ICA is to project the original data to source space

$$\mathbf{A}^{-1}\mathbf{X} = \mathbf{S}, \tag{6}$$

where the unmixing matrix $\mathbf{A}^{-1}_{r\times n}$ is the inverse of $\mathbf{A}_{n\times r}$. The elements of each row vector of the unmixing matrix index the participation of each brain region and are effectively spatial maps, analogous to eigenimages. The row vectors of $\mathbf{S}$ reflect the activation of each component, analogous to eigentimeseries.

Independent components can be interpreted as the dominant functional networks or modes of activity that contribute to the observed neuroimaging data. In addition to a spatial distribution and a time course, each component will have a specific power spectrum and may timelock to certain stimuli or sensitively respond to some other experimental manipulation. Together, these attributes help to reveal the true character of the component. For instance, ICA

can identify artefactual sources of variability in a data set and remove only those components (Jung et al. 2000, Thomas et al. 2002). In fMRI, components with spatial activation ringing around the brain through the entire volume as well as a sudden spike in the time course indicate motion artifact (McKeown et al. 1998a). In EEG, lateral eye movements are marked by a stereotyped spatial distribution with most of the weights concentrated toward the frontal lateral positions on the scalp and a polarity reversal from one side to the other (Jung et al. 2000). Likewise, ICA can identify signal-carrying components, such as task-sensitive functional networks in fMRI (Calhoun et al. 2001c, 2008; McKeown et al. 1998a,b) or task-sensitive ERPs (Makeig et al. 1997, 1999).

A major difference between PCA and ICA is that ICA cannot be used to estimate how many sources of variability there are in the data. ICA is an iterative algorithm that seeks to maximize the independence of the components, but the number of components must be specified prior to the analysis. Because these algorithms are computationally expensive, in practice ICA is usually performed following dimensionality estimation and reduction with PCA (McKeown et al. 2003). Thus, only the first k largest principal components that capture some portion of variance (e.g., 99%) are kept, and the rest are discarded. The k components are then rotated by the ICA algorithm to maximize independence.

Group ICA. Up to this point we have described the extraction of independent components from single subjects, but how can components be made comparable across subjects to allow statistical inference? One approach is to first run the decomposition separately for each subject. Independent components common to most subjects can then be identified either by inspection (Calhoun et al. 2001a) or by a clustering analysis (Jung et al. 2001; Makeig et al. 2002; Onton et al. 2005, 2006). For the latter, some characteristic of the component is first selected (e.g., the spatial map), and the components are clustered into groups based on this characteristic.

An alternative approach is to decompose data from all subjects into the same independent components space. Here, data from all subjects are first concatenated or stacked together such that each subject is implicitly treated as an observation of the same underlying system (Calhoun et al. 2001b, Kovacevic & McIntosh 2007, Schmithorst & Holland 2004). If data are concatenated along the temporal dimension, subjects will have unique time courses but a common spatial map. If the data are concatenated along the spatial dimension, subjects will have unique spatial maps but common time courses. The matrix containing data from all subjects is then decomposed into independent components, which means that all subjects are now in the same space and can be compared directly. Statistical inference on the independent components is possible using univariate (Calhoun et al. 2001b) or multivariate analyses (Diaconescu et al. 2008, Kovacevic & McIntosh 2007, Mišić et al. 2010).

However, if data from individual subjects are concatenated prior to the analysis, components will have either the same spatial map or the same time course for all subjects. Ideally, the decomposition itself should account for between-subject variability as well, with each component having a subject-specific mode in addition to the spatial and temporal modes. The tensor ICA approach is a generalization of the ICA methodology to multiple dimensions, whereby the data are simultaneously decomposed into more than two modes (Beckmann & Smith 2005). In a typical fMRI experiment, these modes would be space, time, and subject. In this way, between-subject variability is estimated directly and allows subsequent between- and within-group analysis.

Applications. Damoiseaux et al. (2006) used tensor ICA to investigate the consistency with which resting-state functional networks manifest across subjects and scans. Resting-state BOLD signal was recorded from one group of subjects in two different sessions. Tensor ICA

was used to decompose the data along three domains: space, frequency, and subject. In the initial analysis, data from the two sessions were decomposed separately but yielded ten common spatial patterns with possible biological relevance. For example, they extracted networks composed of regions thought to be involved in executive processing, motor control and execution, memory, and the default mode. To assess the reliability of these networks, the authors created additional data sets with randomly chosen subjects and scans and then performed an ICA for each of these data sets. They found that all but one of the ten resting-state networks consistently appeared in the additional data. Moreover, regions that displayed the greatest signal change were also the least variable across the additional data sets.

Canonical Correlation Analysis

In a typical neuroimaging study, neural activity is not recorded in isolation but rather in the context of an experimental manipulation or together with some measure of behavior. Multivariate analyses are thus ideally suited to capture either the distributed patterns that respond to experimental manipulation or the patterns that optimally predict behavior. The goal of CCA is to relate two sets of data, $\mathbf{X}_{n\times p}$ and $\mathbf{Y}_{n\times q}$, with p and q variables in their respective columns and n observations in the rows (Hotelling 1936). For example, $\mathbf{X}_{n\times p}$ may represent activity in p voxels while $\mathbf{Y}_{n\times q}$ may represent either the experimental design (e.g., with dummy-coded vectors for membership in $q + 1$ categories) or q behavioral scales. CVA (Friston et al. 1996, Strother et al. 2002), linear discriminant analysis (LDA), and multivariate analysis of variance are special cases of CCA where the matrix $\mathbf{Y}$ codes for class membership (Kustra & Strother 2001). If either of the two sets contains data from only one variable, the analysis simplifies to a multiple regression.

The mathematical objective of CCA is to create pairs of new variables (canonical variates) that are linear combinations of the original variables in $\mathbf{X}$ ($\mathbf{XU}(i)_{n\times 1}$) and the variables in $\mathbf{Y}$ ($\mathbf{YV}(i)_{n\times 1}$) and that have the maximum correlation with each other. In addition, pairs of canonical variates are mutually orthogonal with all the other pairs. When $\mathbf{X}_{n\times p}$ and $\mathbf{Y}_{n\times q}$ are standardized, their between-set correlation matrix is given by $\mathbf{X'Y}_{p\times q}$. The between-set correlations are adjusted for within-set correlations ($\mathbf{X'X}_{p\times p}$ and $\mathbf{Y'Y}_{q\times q}$). The adjusted correlation matrix is then factorized using SVD

$$(\mathbf{X'X})^{-1/2}\mathbf{X'Y}(\mathbf{Y'Y})^{-1/2} = \mathbf{USV'}. \qquad (7)$$

Similar to principal components, each canonical variate pair is composed of an eigenvector $\mathbf{U}(i)_{p\times 1}$, which represents a pattern of neural activity that is maximally correlated with some differentiation of classes, captured by the corresponding eigenvector $\mathbf{V}(i)_{q\times 1}$. The singular values from the main diagonal of $\mathbf{S}_{q\times q}$ are the correlations between the two canonical variates. The squared canonical correlations (eigenvalues) index the proportion of variance shared by the canonical variates. Because the procedure optimizes the relationship between the two data sets, the first canonical correlation coefficient is guaranteed to be at least as large as the largest between-set correlation. The number of canonical variate pairs is determined by the rank r of the between-set correlation matrix, which is the smallest dimension in the two original matrices ($r = min\{n, p, q\}$). Statistical inference on the whole multivariate pattern is possible using several multivariate tests, such as Bartlett's χ^2 assessment of Wilks's Λ. The contributions of the original variables to a canonical variate are usually gauged—but not tested outright—by correlating each of the original variables with the canonical variate. The correlations (canonical loadings) are considered nontrivial if they exceed ± 0.3.

However, a typical neuroimaging data set has more variables (e.g., voxels) than observations (e.g., scans), and the matrix inverse $\mathbf{X'X}^{-1}$ does not exist because $\mathbf{X'X}$ is rank deficient. This can be rectified by reducing the number of variables prior to the analysis, such as first applying PCA and treating the principal components as the new variables (Friston et al. 1995, 1996). Friston et al. (1996) used this approach

to analyze a verbal fluency PET experiment. A letter was presented aurally every 2 seconds. In one condition, subjects had to simply repeat the letter (word shadowing), while in the other condition subjects had to respond with a word that began with that letter (word generation). The data were first reduced by PCA: Out of 60 components (5 subjects $\times$ 12 scans), the first 16 components were selected and entered into a CVA. The first canonical variate pair had a statistically significant canonical correlation and easily differentiated the word-shadowing and word-generation scans across all subjects. To obtain a statistical image depicting the total contributions of individual voxels to this effect, the authors multiplied the eigenvector weights from the PCA with the eigenvector weights from the CVA. The statistical image revealed substantial involvement of the anterior cingulate, Broca's area, and ventromedial prefrontal cortex in the differentiation of word shadowing and generation.

Partial Least Squares Analysis

The goal of PLS analysis is to relate two sets of data in a manner similar to CCA-based methods (Bookstein 1994, Krishnan et al. 2010, McIntosh et al. 1996, McIntosh & Lobaugh 2004, Wold 1982). For a neuroimaging experiment, the set with neural activity, $\mathbf{X}_{(n\times q)\times p}$, is organized as follows: the columns correspond to the p variables (e.g., voxels), while the rows correspond to the n participants nested within q experimental conditions. Depending on what the second set represents, PLS may be used to find spatiotemporal patterns that support a particular differentiation of conditions (task PLS) that optimally relate to behavior, or demographic measures (behavior PLS) that optimally relate to activity in a particular seed voxel (seed PLS), or some combination of these (multi-block PLS). In the following subsection we initially focus on task PLS but eventually describe the other variants as well.

Two approaches for computing PLS have been reported in the literature. In the contrast approach, the matrix $\mathbf{Y}_{(n\times q)\times(q-1)}$ contains orthonormal contrasts that code for the $q-1$ degrees of freedom in the design. The covariance matrix $\mathbf{X}'\mathbf{Y}_{p\times(q-1)}$ is then subjected to SVD. Here it is easy to see that PLS is conceptually and mechanically similar to CCA, with the important distinction that the covariance matrix is not corrected for within-set covariance prior to the decomposition. This characteristic makes PLS ideally suited for neuroimaging data because signal measured by various imaging modalities tends to have a high degree of spatial and/or temporal autocorrelation, leading to a rank-deficient matrix that cannot be inverted.

The alternative approach is to mean center the data matrix $\mathbf{X}_{(n\times q)\times p}$. This is the currently used approach, and it produces identical results save for scaling differences (McIntosh & Lobaugh 2004). Here, the within-task average is computed for each column to create a matrix $\mathbf{M}_{q\times p}$, which is then column-centered and subjected to SVD

$$\mathbf{M}'_{dev} = \mathbf{USV}'. \tag{8}$$

The decomposition yields a set of latent variables, each of which consists of the ith column vectors of $\mathbf{U}_{p\times q}$ and $\mathbf{V}_{q\times q}$, as well as the ith diagonal element of the diagonal matrix $\mathbf{S}_{q\times q}$. The left singular vector $\mathbf{U}(i)$ contains the element saliences (weights) that identify the voxels that collectively make the greatest contribution to the effects captured by the latent variable. The right singular vector $\mathbf{V}(i)$ contains the design saliences, which index the contribution of each task to the spatiotemporal pattern identified by the latent variable. The element and design saliences can be interpreted as the functional network and task contrast with maximal covariance. The strength of the relationship extracted by each latent variable is reflected by the relative size of the singular value. Note that the singular value is a covariance rather than a correlation (cf. canonical correlation). The proportion of between-set covariance accounted for by each latent variable is given by the ratio of the squared singular value to the sum of all other squared singular values. The expression of each latent variable can be calculated by taking the dot product of the singular

Permutation test: a significance test in which sampling without replacement is used to permute the labels of the observed data points in order to generate a sampling distribution of the test statistic under the null hypothesis that the labels are exchangeable

Bootstrapping: a procedure in which sampling with replacement of the original data points is used to assemble a sampling distribution for some parameter (e.g., the mean), usually for the purpose of constructing a confidence interval

vector and the original data, analogous to principal component scores. The projection of the original voxel activations on the element saliences ($\mathbf{XU}_{n\times q}$, brain scores) indicates the degree to which each of the n observations expresses the task effects.

From a practical point of view, PLS and its application to neuroimaging use a framework for nonparametric statistical inference at the level of the entire multivariate pattern as well as at the level of the element saliences and their individual contributions. Specifically, permutation tests (Edgington 1995) are used to test the significance of the latent variables, and bootstrapping (Efron & Tibshirani 1986) is used to estimate the standard errors of the element saliences.

To assess the significance of a latent variable, the rows (i.e., observations) of $\mathbf{X}_{n\times p}$ are randomly reordered (permuted). The new data are mean centered and subjected to SVD as before in order to obtain a new set of singular values. These singular values are effectively generated under the null hypothesis that there is no association between neural activity and the task. The procedure is then repeated many times (e.g., 1,000) to generate an entire sampling distribution of singular values under the null hypothesis. Because the singular value is proportional to the magnitude of the effect, the p-value is estimated as the probability that singular values from the distribution of permuted samples exceed the singular value from the original, nonpermuted data. In other words, the p-value is the probability of obtaining a singular value of this size under the null hypothesis that there is no association between the task and brain activity.

The contribution of individual element saliences to the latent variable is operationalized in terms of reliability or stability. Many bootstrap samples (e.g., 1,000) are generated by sampling with replacement subjects within conditions and then computing PLS for each sample to generate a distribution of saliences. The bootstrap distribution is then used to estimate the standard errors and confidence intervals for the saliences. The idea is to see which saliences are stable irrespective of the sample and which saliences are sensitive to which subjects in the sample. The ratio of the salience to its bootstrap-estimated standard error (bootstrap ratio) allows saliences that are both large and reliable to be selected. If the bootstrap distribution is normal, this ratio is approximately equivalent to a z-score. Stable saliences identify voxels that make a robust contribution to the latent variable.

So far we have described how data from a single time point are analyzed, but the analysis can easily be extended to the temporal domain by treating each data point (i.e., each voxel at each scan) as a separate variable. For example, if there are p voxels and t scans, the data matrix $\mathbf{X}_{(n\times q)\times(p\times t)}$ is constructed by nesting the scans within the voxels in the columns. Thus, the first column contains the amplitude of the first voxel in the first scan, the second column contains the amplitude of the first voxel in the second scan, and so on. The rest of the analysis is performed as described above. The only difference is that the left singular vectors contain saliences that describe a spatiotemporal pattern of neural activity. In each vector $\mathbf{U}(i)_{(p\times t)\times 1}$ the saliences are organized in the same way as in $\mathbf{X}_{(n\times q)\times(p\times t)}$, with the scans nested within the voxels. This simple innovation permits analysis of neuroimaging data with a time component, such as event-related fMRI (McIntosh et al. 2004) and ERPs (Lobaugh et al. 2001).

Behavior PLS. A notable variant of the analysis described above is one where the matrix $\mathbf{Y}_{(n\times q)\times b}$ contains b demographic and/or behavioral measures. In this case, the input matrix should be constructed such that it contains correlations between neural activity and behavior within each of the q experimental conditions. The condition-specific correlations between the submatrices $\mathbf{X}_{n\times p}$ and $\mathbf{Y}_{n\times b}$ are calculated first and then stacked to form the input matrix

$$\mathbf{Y}'\mathbf{X}_{behavior} = \begin{bmatrix} \mathbf{Y}_1'\mathbf{X}_1 \\ \mathbf{Y}_2'\mathbf{X}_2 \\ \vdots \\ \mathbf{Y}_q'\mathbf{X}_q \end{bmatrix}. \qquad (9)$$

Rather than finding patterns of neural activity that relate to some differentiation of conditions, the analysis now finds patterns of neural activity that relate similarly and differently across conditions (or groups) to behavior. Behavior saliences $\mathbf{V}(i)$ indicate the degree to which brain-behavior correlations are expressed for each task. Element saliences $\mathbf{U}(i)$ indicate the degree to which individual regions express these brain-behavior correlations.

Seed PLS. The machinery for behavior PLS can also be used to assess task-dependent changes in the correlation (i.e., functional connectivity) between one or more seed regions and the rest of the brain. The seeds may be selected either on theoretical grounds or using another statistical analysis, such as task PLS. The activity of each of the b seed voxels is entered into the matrix $\mathbf{Y}_{(n\times q)\times b}$, with the data from different conditions stacked on top of each other. The input matrix is constructed in a manner analogous to behavior PLS. Seed PLS analysis of this matrix identifies brain-seed correlations that are interpreted in terms of element and seed saliences. Seed saliences $\mathbf{V}(i)$ indicate the degree to which the seed regions—during different tasks—are functionally connected to the spatiotemporal patterns of neural activity depicted by element saliences $\mathbf{U}(i)$.

Multiblock PLS. Task, behavior, and seed PLS relate patterns of neural activity to one other "block" of data. Multiblock PLS is a way to simultaneously relate neural activity to two or more blocks of data (Caplan et al. 2007). For example, multiblock PLS can identify functional networks that support some differentiation of tasks and at the same time display robust functional connectivity with a particular seed region. The input matrix is constructed by column-wise concatenating the matrix of task means $\mathbf{M}_{dev}$ and the task/group dependent correlation of behavior/seed with brain activity (e.g., $\mathbf{Y}'\mathbf{X}_{behavior}$ and/or $\mathbf{Y}'\mathbf{X}_{seed}$):

$$\mathbf{Y}'\mathbf{X}_{multi} = \begin{bmatrix} \mathbf{M}_{dev} \\ \mathbf{Y}'\mathbf{X}_{behavior} \\ \mathbf{Y}'\mathbf{X}_{seed} \end{bmatrix}. \qquad (10)$$

The multiblock covariance matrix $\mathbf{Y}'\mathbf{X}_{multi}$ is then decomposed by SVD, and the singular vectors are interpreted as before. The left singular vectors $\mathbf{U}$ are a spatial pattern of voxels. The only difference from a standard analysis is that the right singular vectors $\mathbf{V}$ now contain saliences that capture both task and behavior/seed relationships. The weights in the columns of $\mathbf{V}$ identify LVs that reflect either a task effect, a behavior/seed effect, or the convergence of the two:

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_{task} \\ \mathbf{V}_{behavior} \\ \mathbf{V}_{seed} \end{bmatrix}. \qquad (11)$$

Applications. Grady et al. (2010) used both task and seed PLS to study the effect of aging on default mode and task-positive networks. During scanning, young and older adults performed four visual tasks. During the fifth "condition," participants were required to fixate on a centrally presented dot but did not have to respond. Both groups and all conditions were entered into a task PLS analysis, which revealed two statistically significant latent variables. The first latent variable separated the no-task fixation condition from the other four task conditions in both groups. The design saliences were negative for fixation and positive for the tasks. The regional pattern that showed consistently greater activity during the tasks (indexed by positive bootstrap ratios) resembled the task-positive network. The regional pattern that showed the opposite trend (greater activity during fixation, indexed by negative bootstrap ratios) resembled the default mode network. The authors concluded that the first latent variable isolated default mode and task-positive networks that were similar for young and older participants.

The authors then used seed PLS to investigate how functional connectivity within

the default mode and task-positive networks changes during aging. Connectivity in the default mode network was analyzed by choosing two prominent default mode regions from the task PLS analysis as seeds. Seed PLS analyses revealed high correlations with other default mode regions. Likewise, prominent task-positive regions were chosen as seeds to assess connectivity in the task-positive network. The seeds mainly correlated with other task-positive regions. To assess whether functional connectivity within either network differed between groups, brain scores were correlated with seed activity, separately for each group. This comparison showed that connectivity within the default mode network decreased with age, whereas connectivity within the task-positive network did not change.

Classification Techniques

Neuroimaging experiments are specifically designed to differentiate and isolate brain states. Neural activity collected during an experiment can be stratified into several classes according to the experimental design. Classes may represent different stimuli that were presented and tasks that were performed as well as responses or decisions made by subjects. Classes may also represent different groups of subjects. The goal of classification is to use imaging data to find brain states that can reliably predict class membership (O'Toole et al. 2007, Pereira et al. 2009). Multivariate classification techniques seek patterns of data features (e.g., voxel activations) that maximally separate the classes. If conventional analyses such as GLM can be thought of as an attempt to use the design variables to predict neural activity, then classifiers effectively do the opposite: They use patterns of neural activity to predict the experimental design (Pereira et al. 2009). In a typical analysis, data are divided into a training set and a test set. The classifier is calibrated using the training set and then used to predict class membership in the test set. Classification accuracy on the test set allows for statistical assessment of classifier performance. Predictive learning (Strother et al. 2002) and multivoxel pattern analysis (Haynes & Rees 2005, Norman et al. 2006) are also forms of classification.

Types of classifiers. Multivariate classifiers are a family of techniques defined more by purpose and application than by mechanics. Geometrically, the original p features (e.g., voxels) define a p-dimensional space, and each example (e.g., brain volume) represents a point in that space. Most classifiers identify patterns by learning a function that will take values of the features as an input and generate a class label. For linear classifiers this function is a simple linear combination of features that can be thought of as a hyperplane that maximizes the separation between points that belong to different classes (Campbell & Atchley 1981).

The flexible associative multivariate models previously described, such as CVA and PLS, are both examples of techniques that look for linear patterns to differentiate classes. One of the most commonly used classifiers in neuroimaging—LDA—is a special case of CVA with only two classes, which in turn is equivalent to CCA when the matrix $\mathbf{Y}$ codes for class membership (Kustra & Strother 2001). LDA is often framed in terms of maximizing the between-class covariance $\mathbf{B}$ (computed as the covariance of the class means) relative to the within-class covariance $\mathbf{W}$. The linear combinations that satisfy this condition are given by the eigenvectors of $\mathbf{W}^{-1}\mathbf{B}$, which are equivalent to the eigenvectors of the adjusted between-set correlation matrix for CCA (Equation 7). In choosing the best linear combination, multivariate classifiers will take into account the relationships between voxels. For example, the voxels that constitute the dominant pattern extracted by LDA were chosen because together they covaried with the differentiation of classes. Thus, discriminant functions can often be thought of as functional networks. The linear kernel support vector machine (SVM) is another salient example of a linear classifier. SVM focuses only on those examples that lie close to other classes and uses these to construct a discriminant function that maximizes the margin between the classes.

Nonlinear classifiers learn a nonlinear function of the features and are more diverse. Some nonlinear classifiers do not seek a hyperplane but more generally a surface to separate classes. For example, quadratic discriminant analysis assumes the discriminant function is a quadratic polynomial and computes a quadratic surface. Other methods change the space in which the data are represented. For example, SVMs with nonlinear kernels transform the original feature space to a higher-dimensional space where it is theoretically easier to construct separating hyperplanes.

Cross-validation. An important notion in classification is that the patterns extracted should be able to predict class membership of a new sample that the classifier has not previously encountered. The focus on prediction rather than explanation means that cross-validation plays a prominent role in multivariate classification for neuroimaging (Strother et al. 2002). The simplest approach would be to split the available data into halves, then train the classifier on one half and test it on the other. However, neuroimaging samples (scans, subjects) are scarce, and with a reduced data set the discriminant function may be too variable and may not generalize well to other cases. Ideally, the classifier should be trained with as many samples as possible.

This is made possible by k-fold cross-validation. The sample is randomly split into k subsamples, with each subsample containing the same number of examples for each class. In each of the k iterations ("folds"), $k-1$ subsamples are used to train the classifier, and the single remaining subsample is used for validation. The classification accuracies obtained from individual folds are combined and have effectively made use of the entire sample. If the training sets do not contain equal numbers of examples for each class, a classifier may not be able to learn how to discriminate among classes that are underrepresented. This is usually not a problem for studies with counterbalanced designs.

To determine the success of a classifier, one must assess whether the neuroimaging data significantly helped to predict class membership. The null hypothesis is that the same classification accuracy could have been achieved simply by chance. Imagine that the probability of successfully classifying a single example by chance is p, and the probability of incorrectly classifying it is $1-p$. For example, if there are q classes, $p = \frac{1}{q}$. This helps to construct the null probability for any situation where the classifier correctly classified k examples by chance, out of a total of n test examples. The probability of being successful k times ($Pr\{X = k\}$) is given by the binomial distribution

$$Pr\{X = k\} = \binom{n}{k} p^k (1-p)^{n-k}, \qquad (12)$$

where p^k is the probability of k successful classifications, $(1-p)^{n-k}$ the probability of $n-k$ unsuccessful classifications, and $\binom{n}{k}$ the number of possible ways this could occur. The corresponding p-value is $Pr\{X \geq k\}$ and can be calculated by integration.

Applications. Carlson et al. (2003) used fMRI to investigate whether the perception of different object categories is modular (i.e., performed by category-specific areas) or distributed among many areas. Subjects were presented with pictures of multiple object categories (e.g., faces, houses) in the context of a delayed-match-to-sample task, as well as passive viewing. The data were first reduced by PCA and only the first 40 components were retained, corresponding to 80% to 85% variance. An LDA classifier was trained and assessed using a variant of the k-fold cross-validation procedure described above. One set of exemplars was removed and a subset of the remaining data (randomly chosen by sampling with replacement) used as a training set. The classifier was then tested on the exemplars that had been held out. The authors found that the trained classifiers discriminated at levels significantly above chance. Moreover, patterns of activity that distinguished one category from others had little spatial overlap, in support of the modularity hypothesis.

CONFIRMATORY TECHNIQUES

Psychophysiological Interactions

PPI: psychophysiological interaction

If the correlation between two brain regions changes significantly under different experimental manipulations, this suggests an interaction between the psychological variable and the underlying physiology—a psychophysiological interaction (PPI) (Friston et al. 1997). In PPI, the activity of one brain region is regressed onto the activity of another brain region in different experimental conditions, and the change in slope is assessed. PPI seeks to explain the physiological response in one region in terms of an interaction between a task and physiological activity in another region by explicitly looking for regions whose correlation with the seed changes in response to the task. The idea is that activity in a seed region of interest may correlate with activity in other regions not due to the experimental manipulation, but rather simply by virtue of anatomical connections, common sensory inputs, or neuromodulatory influence. Thus, rather than looking for significant correlations, PPI looks for correlations that change significantly during a task.

The first step is to select a seed region and to extract its time course. As in seed PLS, the region may be chosen either on the basis of theory or by prior analysis such as task PLS. The PPI analysis is performed in the framework of a standard GLM. If the objective were to find regions whose activity correlates with the seed region, activity from the seed region could be entered as a predictor. Rather, the objective is to find regions whose activity depends on the interaction between the task and the seed region. To this end, an "interaction" predictor is created by taking the scalar product of the time course of the task (exactly as in a standard GLM) and the time course of activity in the seed region. GLM analysis with this interaction predictor would identify voxels whose correlation with the seed region is significantly higher in one task than in the other.

For every voxel in the brain Y, PPI is formulated as a regression equation that tries to predict Y from seed X, task A, and the interaction of the seed and the task (XA) (McIntosh & Gonzalez-Lima 1994):

$$Y = \beta_{y.z}X + \beta_{y.a}A + \beta_{y.xa}XA + \epsilon. \quad (13)$$

Notice that the interaction predictor is actually a product of two main effects: the seed time course and the experimental design. Voxels whose activity correlates only with the seed or only with the experimental design will still display some correlation with the interaction predictor. Therefore, the relationship between the interaction predictor and the target voxel is assessed as a semipartial correlation $\beta_{y.xa}$.

Applications. Stephan et al. (2003) investigated whether cognitive control and task execution show lateralization effects. During an fMRI experiment, participants were presented with four-letter nouns that had a red letter either in the second or third position. In the letter-decision task, they had to indicate whether the word contained the letter "A"; in the visuospatial-decision task, they had to indicate whether the red letter was on the left or the right. In a baseline task, the participants simply had to respond as quickly as possible to the appearance of the stimuli. In an initial GLM analysis, the authors contrasted the letter-decision and visuospatial-decision tasks with the baseline and found significant differences in the left and right anterior cingulate cortex (ACC), implicating these regions as the locus of cognitive control. The goal of the subsequent PPI analysis was to determine whether the coupling between the ACC and any other region in the brain significantly changed during task execution. The analysis revealed that the effective connectivity between left ACC and left inferior frontal gyrus significantly increased during letter decisions, whereas the connectivity between the right ACC and the left intraparietal sulcus increased during visuospatial decisions. The authors were able to conclude that cognitive control over regions involved in task execution is exerted within the same hemisphere.

Structural Equation Modeling

The goal of SEM is to construct a causal model and to test whether it is consistent with the data (Jöreskog et al. 1979, Loehlin 1987). In neuroimaging, SEM models are usually a subset of brain regions and the pattern of causal influence among them (McIntosh & Gonzalez-Lima 1991, 1994; McIntosh et al. 1994). The regions to be included in the model are either chosen a priori or from some exploratory analysis. The influences between regions are constrained anatomically, such that direct influence between two regions is possible only if there is a known white matter pathway between them. The models are then used to assess how interregional influence (i.e., effective connectivity) differs between experimental conditions or groups.

Each brain region is treated as a variable, and casual influences between regions are specified in terms of linear regression equations. As an example, consider the system identified by the path diagram in **Figure 1**. Here, putative anatomical projections (depicted by arrows) engender effective connections. Structural equations are essentially regression equations that define the sources of variance for each variable. In the present example, x_i represents the variance of each region. Each regional variance can be partitioned into variance explained by other regions, as well as an error or residual term (ψ_{x_i}). The residual terms may be thought of as exogenous influences from other brain regions that could not be included in the model, or the influence of a brain region upon itself. The strength of each connection is given by the regression weights p, q, r, s and t, also known as path coefficients. The causal order of the network is described by a system of structural equations:

$$\begin{aligned} x_1 &= \psi_{x_1} \\ x_2 &= p x_1 + \psi_{x_2} \\ x_3 &= q x_1 + r x_2 + \psi_{x_3} \\ x_4 &= s x_2 + t x_3 + \psi_{x_4}. \end{aligned} \tag{14}$$

The key idea behind SEM is that this system of equations assumes a particular causal order and can be used to generate an implied covariance matrix (McArdle & McDonald 1984). The implied covariance matrix is a prediction of the variances and covariances between regions, parameterized in terms of the path coefficients. Here we show the corresponding correlations for simplicity:

$$\begin{aligned} R_{x_1,x_2} &= p \\ R_{x_1,x_3} &= q + pr \\ R_{x_1,x_3} &= ps + prt + qt \\ R_{x_2,x_3} &= r + pq \\ R_{x_2,x_3} &= s + rt + pqt \\ R_{x_3,x_4} &= t + sr + qps. \end{aligned} \tag{15}$$

The covariance matrix is fitted to the empirical covariance matrix to estimate the path coefficients and residual variances. Typically, a method such as maximum likelihood estimation (MLE) or weighted least squares is used to establish a fit criterion that must be maximized. The model is initialized by guessing the values of the unknown parameters. At each iteration of the algorithm, the parameters are slightly altered and the fit of the implied covariance matrix to the empirical covariance matrix is reassessed. The procedure continues until there is no appreciable improvement in fit. Thus, known parameters (variances and covariances) are used to estimate unknown parameters (path coefficients and residual variances). SEM can be thought of as a method of using patterns of functional connectivity (covariances) to draw inferences about effective connectivity (path coefficients).

Notice that for any given connection (e.g., x_1 to x_3), the corresponding structural equation contains terms for the influence of other regions (e.g., the path coefficients for x_1 to x_2 and x_2 to x_3) in addition to the path coefficient for that connection. The resulting path coefficient has a meaning similar to a semipartial correlation in the sense that it reflects the influence of one region on another, with influences from all other regions on the sink region held constant. In practice, maximum likelihood estimates of path coefficients differ only slightly from least-squares estimates of semipartial regression coefficients mainly because the

former are calculated simultaneously, whereas the latter are calculated separately (McIntosh & Gonzalez-Lima 1994).

Model inference. The simplest application of SEM is one where a model is formulated and tested against the data. The discrepancy between the covariance matrix implied by the model and the empirical covariance matrix can be assessed using some goodness-of-fit test, such as the χ^2 statistic. A large χ^2 value indicates a significant departure from the empirical covariance matrix and indicates that there is sufficient evidence to reject the null hypothesis that the implied and empirical covariance matrices do not differ (i.e., the model is not consistent with the data).

However, there is no guarantee that the tested model will be the best-fitting model, and SEM can also be used to compare competing models. For example, if regions x_1 and x_2 were known to be part of a direct feedback loop, one may wish to test whether the effective connections between them are equal (symmetric) or unequal (asymmetric). The null model would constrain the connections to be equal by parameterizing each direction with the same path coefficient. The alternative model would assign different path coefficient parameters to the connections, allowing them to freely vary. An implied covariance matrix is generated for each model and statistically compared with the empirical covariance matrix using a χ^2 goodness-of-fit statistic. The models are then compared using the χ^2 difference test. The difference is computed by subtracting $\chi^2_{alternative}$ from χ^2_{null} and then assessed with respect to the differences in degrees of freedom for the two models. The test helps to determine whether the modification (i.e., additional parameter) significantly improves the fit of the model. In the present example, a significant difference test would imply that the path coefficients were significantly different in the two directions. Notice that the two models could not be distinguished in terms of functional connectivity, which is symmetric by definition (McIntosh & Gonzalez-Lima 1994).

Under the hierarchical testing scheme described above, SEM can be used to examine whether one or more causal influences change due to experimental manipulation by comparing models for two different conditions or groups of subjects. The simplest strategy would be to employ separate model runs and then describe where the models differ, although in this case there is no opportunity for statistical inference. The more common approach is to combine the models in a single multigroup or stacked run. The null hypothesis is that effective connections do not differ between groups, whereas the alternative hypothesis is that effective connections are group specific. Here, a null model is first constructed such that path coefficients of interest are constrained to be equal for both groups. In the alternative model these path coefficients are free to vary separately for each group. The alternative hypothesis is tested by generating an implied covariance matrix for each model and statistically comparing them with the empirical covariance matrix. An alternative χ^2 that is significantly lower than the null χ^2 implies that the path coefficients were statistically different for the two groups. An interesting situation arises if the omnibus test indicates a poor overall fit, but the difference test indicates a significant change from one task to another. SEM has been shown to be resilient in these situations because it can detect changes in effective connectivity even if the absolute fit of the model is inadequate (Protzner & McIntosh 2006).

Bullmore et al. (2000) describe an alternative approach to model selection, where the nodes of the network are specified a priori, but the paths are traced out in a data-driven manner. The starting point for the procedure is the null model with all path coefficients equal to zero. At each iteration, the path coefficient with the largest modification index (the improvement in model fit if that parameter were freed) is unconstrained and incorporated into the model. The addition of any path will improve the χ^2 value, so the fit of the new model is evaluated in terms of a parsimonious fit index, which is high for models with a well-fitting covariance matrix

and with the fewest paths (Bollen 1986). A path is added permanently only if it improves the fit index; otherwise, the path with the next highest modification index is unconstrained and evaluated. The algorithm continues until a model with the maximum fit index is identified that cannot be increased by adding any more paths. A confidence interval for the parsimonious fit index of the final model is formed by bootstrapping. Bootstrap samples are created by randomly sampling subjects with replacement, and the entire procedure is repeated to generate a null distribution for the fit index.

Applications. Nyberg et al. (1996) used SEM to study how effective connectivity changes during episodic memory retrieval relative to a baseline reading task. An initial univariate subtraction analysis revealed that activity in some regions increased relative to baseline, while activity in a number of regions actually decreased. One possible explanation for the decreased activity is that the baseline activity in those regions was higher than during the task. An alternative explanation is that activity in these regions was suppressed. The authors used SEM to test the hypothesis that the decrease in regional cerebral blood flow was the result of direct, active inhibition by regions with increased activity. They used a stacked-run SEM analysis to construct a null model in which path coefficients feeding back from activated to deactivated regions were constrained to be equal across tasks, as well as an alternative model in which the path coefficients were allowed to differ. The modification resulted in a significant χ^2 difference test, indicating significantly improved model fit. Moreover, the feedback paths were negative in both conditions but more negative in the episodic retrieval condition, suggesting increased inhibition. Thus, SEM was used to disambiguate the mechanism behind task-dependent changes in neural activity. Interestingly, the areas of decreased blood flow that were the targets of "inhibitory" effects were key constituents of the default mode network (Raichle et al. 2001), including medial frontal and retrosplenial cortices.

Dynamic Causal Modeling

DCM uses a Bayesian framework to estimate causal influences in a network of brain regions and how these influences change due to experimental manipulation (Friston et al. 2003). Neural activity in each region is modeled by differential equations that describe the local dynamics. The causal architecture of the network arises from interactions among regions. The interactions are specified by a set of coupling parameters that represent the efficacy of synaptic coupling and model effective connectivity. The biologically plausible causal model generates neural activity in real time. A forward model translates hidden neural dynamics at each region to measured responses (e.g., BOLD signal). Bayesian model inversion allows information from the experiment to be incorporated back into the causal model to get a better estimate of effective connections. Competing hypotheses about how experimental context modulates synaptic coupling are formulated as different models and compared to each other in terms of their relative likelihood given the data. The optimal model can further be characterized in terms of its coupling parameters. As we discuss below, DCM does not mandate any specific biophysical model of neural activity nor any specific forward model. Rather, DCM is a generic framework for inferring context-dependent changes in synaptic coupling at the neural level.

Causal model. The first stage is to define a model of causal order. Each region in the model consists of neuronal subpopulations that are intrinsically coupled to each other. Extrinsic coupling between different neuronal populations models a network of regions. The state or activity of each neuronal population is described by a set of stochastic or ordinary differential equations that relate the rate of change in activity (i.e., the future state) to the present state. The synaptic coupling among different populations allows terms for the current state of one population to be introduced in the equation describing the state of another population. The

coupling parameters can be thought of as rate constants that determine the speed with which one population influences another. Causal order is embodied in the ability of dynamics in one region to influence dynamics in another region. Experimental manipulations are modeled as external perturbations of the system. External inputs may induce either a change in coupling or a change in activity in a specific neuronal population. Therefore, the underlying causal model is a system of coupled differential equations,

$$\frac{\partial x}{\partial t} = f(x, u, \theta^c), \tag{16}$$

that describe how the rate of change of states x is a function of states of other populations (x), external inputs (u) and coupling parameters (θ^c). The coupling parameters θ^c are unknown and the purpose of DCM is to infer them, much like path coefficients in SEM. Notice that causality is engendered at the level of neural activity rather than at the level of the observed signal.

DCMs do not impose any one particular biophysical model of neural activity. The only stipulation is that the model is biologically plausible and adequately captures interactions between populations as well as externally induced perturbations. There is a well-developed literature on dynamical models of neural activity (Breakspear & Jirsa 2007), and the incorporation of new models into the DCM framework is an active topic of research (Daunizeau et al. 2009, Friston & Dolan 2010). For the present, we merely note that the type of model employed will depend on the imaging modality used in the experiment. Due to the slow and regionally variable hemodynamic response, fMRI does not provide sufficient information to estimate time delays in coupling between regions. Thus, DCMs for fMRI usually do not model conduction delays, whereas DCMs for EEG/MEG do. Recently developed DCMs for fMRI explicitly model excitatory and inhibitory subpopulations for each source (Marreiros et al. 2008). DCMs for EEG/MEG use more detailed neural mass models composed of populations of pyramidal cells as well as populations of excitatory and inhibitory interneurons (David et al. 2006).

Forward model. The spatiotemporal evolution of system dynamics at the neuronal level is described in real time. The second stage of DCM is to enable comparison with observed data by using a forward model to translate neuronal system states into measurements. The forward model is an explicit mapping (g) from neuronal activity (x) to some feature of the data (y)

$$y = g(\mathbf{x}, \theta^f). \tag{17}$$

The form of the forward model will depend on the imaging modality. For example, if the data are evoked responses, such as ERPs or event-related fields, the function g is the lead field matrix that models the propagation and subsequent volume conduction of electromagnetic fields through brain tissue, cerebrospinal fluid, skull, and skin. In this case, the additional unknown parameters θ^f introduced are the location and orientation of the source dipole (Kiebel et al. 2006). On the other hand, if the signal is BOLD contrast, the function g models how state changes at the neuronal level induce change in local blood flow, inflating blood volume and reducing deoxygenated hemoglobin (Buxton et al. 1998). In that case, the unknown parameters specify quantities such as the rate constants of vasodilatory signal decay and autoregulatory feedback by blood flow (Stephan et al. 2007).

Despite apparent differences, SEM and DCM have much in common. Both techniques seek to estimate context-dependent changes in effective connectivity. Through the prism of SEM, the intrinsic connectivity implemented in DCM can be thought of as the grand average effective connectivity across all conditions and the modulatory effects as the changes in the intrinsic connections due to experimental manipulation. From the perspective of DCM, SEM can be thought of as a special case in which the system is driven by noise rather than systematic exogenous inputs (**Figure 1**), while the interactions are linear and take place at the

level of the observations rather than at the neural level.

Model inversion. DCMs use a Bayesian framework for estimating the unknown parameters, and here we briefly outline the logic behind the approach. The parameters of interest are assumed to be random variables with some probability density. Before an experiment is performed, the parameters have a prior distribution, which reflects a priori knowledge about their values. Thus, unknown parameters are constrained either to an interval or to a fixed value. For example, one may have prior empirical knowledge about the likely range of values of some hemodynamic parameters. Likewise, one may make an explicit assumption that some coupling parameters are zero. After the experiment is performed, new information is obtained from the data and used to update the prior distribution. The new distribution of each parameter, which takes into account both prior beliefs and the available data, is called the posterior distribution. Estimation of the posterior distribution is essentially an optimization problem, and the priors can be thought of as soft constraints because they bias the parameter estimates.

Bayesian model inversion is a procedure that uses the observed data to update the model (i.e., estimate the parameters) in a way that maximizes the model evidence. This quantity, also known as the marginal likelihood of the model, is defined as the probability of the data given the model m. Model evidence is highest for models that explain the data as accurately as possible and at the same time have the fewest parameters. The unknown parameters from the causal and the forward model are denoted by $\theta = \{\theta^c, \theta^f\}$. The posterior density $p(\theta|y, m)$ is estimated by combining the prior density on the parameters $p(\theta|m)$ with the likelihood function $p(y|\theta, m)$,

$$p(\theta|y, m) \propto p(y|\theta, m)p(\theta|m), \qquad (18)$$

which follows from Bayes' rule.

Inference. DCMs allow statistical inference on models and on parameters. For example, alternative models correspond to alternative hypotheses about context-dependent changes in neural activity. Thus, model space should be constructed systematically and include only plausible models. Two models can be compared directly either by taking the ratio of their respective evidence (Kass & Raftery 1995) or the difference in their respective log evidence. A model with evidence more than 20 times greater than another model is considered stronger. This procedure (Bayesian model selection) can be used to make a wide variety of comparisons, such as DCMs with different inputs, different anatomical connections, or different priors. Models with different numbers of parameters can be compared directly because evidence takes into account model complexity. Once the optimal model is selected, specific parameters can be statistically assessed with respect to their posterior densities. For instance, the probability of exceeding some preset threshold can be evaluated directly from the posterior probability.

Because model inversion is done on a subject-by-subject basis, there is no guarantee that the same model will necessarily be optimal for all subjects. Therefore, for between-subjects (group-level) inference the investigator must decide whether or not to enforce the same model for all subjects (Stephan et al. 2010). If there is reason to believe that the process under study is homogeneous in the population (e.g., a sensory response in healthy subjects), one can multiply the evidence for a specific model (or add the log-evidence) for each subject to get group-level evidence for that model. This is effectively a fixed-effects assumption. Conversely, if there is reason to believe that the process under study is heterogeneous in the population (e.g., a cognitive process in a spectrum disorder), one can compute the ratio of the number of subjects who show positive evidence for a given model m_i relative to the number of subjects who show greater evidence for another model m_j (Stephan et al. 2007). This is a random-effects assumption.

Inference on model parameters will also depend on whether individual effects are assumed

to be fixed or random. In the fixed-effects case, the model is the same for all participants and posterior densities can be estimated for all participants. Thus, one approach to estimating group-average parameters is to compute a joint density for the subject-specific posterior estimates (e.g., Garrido et al. 2007). The probability of exceeding a particular threshold can be computed directly from the joint density, exactly as in the case of single-subject analysis. In the random-effects case, the models may be different for different subjects and thus a joint posterior probability cannot be computed. The most common approach is to treat the mode of each subject-specific posterior distribution as a summary statistic. Known as maximum a posteriori estimates, these can then be submitted to a traditional random-effects analysis of variance or *t*-test.

Applications. A study by Garrido et al. (2009) demonstrates how DCM can be used to study the mismatch negativity (MMN). The MMN is a pronounced negative potential in response to oddball stimuli (deviants) embedded in a stream of repeated stimuli (standards). The authors investigated whether the MMN is caused by comparison between sensory input and a memory trace of previous input or by local adaptation in primary auditory cortex as a result of repeating stimuli, or both. Regions for the network were selected on the basis of previous studies. Local adaptation and memory comparison were operationalized as changes in intrinsic (within a population) and extrinsic (between populations) coupling, respectively. Several models were constructed, involving no changes in coupling, changes in local coupling only, changes in extrinsic coupling only, or changes in both intrinsic and extrinsic coupling. The DCMs were inverted separately for each subject and log evidence was summed across subjects for each model in order to select the best one. The model that allowed changes in both intrinsic and extrinsic coupling had the highest total log evidence. The free effective connections for each subject-specific model were computed separately for each condition and compared. Separate *t*-tests for each connection showed that coupling increased for deviants relative to standards, indicating learning-related changes in coupling.

OTHER TECHNIQUES

Multivariate Granger Causality

A signal x can be said to cause another signal y if the past of x can predict the future of y (Granger 1969). If one is interested in such causal relationships in a whole network of brain regions, this framework can be extended to include multiple predictors, such that the present activity of all regions is being predicted by the past activity of all other regions. If causal effects are assumed to be linear, the problem can be formulated as a multivariate linear regression and is termed a multivariate vector autoregressive model (MVAR) (Goebel et al. 2003). By default, the model contains terms for every possible connection in the network, so each connection is tested to see which ones are nonzero. In this manner, a directed subnetwork depicting causal flow can be extracted without any a priori hypothesis about the connectivity between regions.

An MVAR model of order m seeks to predict the present (tth) values of p variables (e.g., brain regions) as a linear combination of their m previous values. The tth sample from the multivariate time series is represented by the p-dimensional vector $\mathbf{X}(t)$:

$$\mathbf{X}(t) = \sum_{i=1}^{m} \mathbf{A}(i)\mathbf{X}(t-i) + \mathbf{E}(t). \qquad (19)$$

The ith matrix $\mathbf{A}(i)$ is a $p \times p$ matrix of autoregressive coefficients, and $\mathbf{E}(t)$ is a vector of residuals. The current value of the jth voxel $x_j(n)$ is a linear combination of m past values of all voxels, with the jth column of each matrix $\mathbf{A}(i)$. This is the key aspect of the multivariate version of Granger causality. For any given connection, the influence of other nodes in the network is accounted for and partialled out. Thus, multivariate Granger causality measures whether the past of x helps to predict y over and

above other variables z. The coefficients $\mathbf{A}(i)$ can be estimated by ordinary least-squares, i.e., by minimizing the sum of squared errors between the predicted and observed values of $\mathbf{X}(t)$. The effective connections are construed as regression equations, and their significance can be assessed via the F test. An alternative approach would be to assess the autoregression coefficients with respect to an empirical null distribution created by bootstrapping (Roebroeck et al. 2005).

The concept of Granger causality has been modified and adapted to accommodate many types of data features and interregional relationships, and here we outline two such innovations. The first involves the application of Granger causality after the original time series data are transformed to the frequency domain (Kaminski et al. 2001). Spectral Granger causality [also known as the directed transfer function (DTF)] is then calculated for each frequency and can be interpreted as the proportion of total power in some signal y that can be attributed to signal x. This variant of Granger causality may be particularly useful for studying regional interdependencies in data with high temporal resolution, where many effects of interest are specific to a certain frequency band.

Statistical inference is more complicated in this case because the parametric distribution of spectral Granger causality is not fully understood. A common approach is to estimate an empirical null (surrogate) distribution instead. Surrogate data are generated by transforming the original time series into the frequency domain, randomizing the phase coefficients, and then transforming the signal back into the time domain (Theiler et al. 1992). The surrogate signal is identical to the original in all aspects save for the causal temporal dependencies, which are now destroyed. The surrogate data can now be subjected to the same spectral Granger causality analysis to generate an empirical null distribution for each DTF coefficient. A p value can be calculated for each connection by comparing the DTF obtained from the original data against the corresponding null distribution.

A second major innovation is to assess predictability in terms of conditional mutual information rather than regression. This approach develops the concept of Granger causality to include nonlinear causal interactions between regions and is known as transfer entropy (TE) (Schreiber 2000). Although TE was a bivariate measure in its initial formulation, it has been extended to the multivariate case such that confounding influences from intervening regions are accounted for (Vakorin et al. 2009). This approach does not assume any particular causal order (TE is computed for all pairwise connections) or any particular type of causal influence (TE is sensitive to linear and nonlinear effects).

Surrogate data: a constrained realization of the observed data in which one or more parameters are altered to reflect the null hypothesis. This can be used to assess the likelihood of observing the data under the null hypothesis

Applications. Deshpande et al. (2009) used multivariate Granger causality to delineate a series of effective connectivity networks while participants performed a hand-grip experiment and subsequently became fatigued. They characterized the effect of fatigue in terms of topological changes. Specifically, for each region, they calculated the number of incoming and outgoing connections (in- and out-degree) as well as the average shortest path to all other nodes in the network (eccentricity). They found that the onset of fatigue was concomitant with a decreased out-degree and eccentricity for primary sensory-motor areas and an increased eccentricity of the cerebellum, suggesting that the latter became a major driving influence in the network.

Graph Model

The techniques described so far try to detect patterns and changes in connectivity. However, the results are interpreted either as an increase or a decrease in functional or effective connectivity but are seldom considered in terms of how such changes impact the broader capacity of the network to process information. For example, a reduction in the functional connectivity of a small number of regions may streamline information flow through alternate paths and increase the overall capacity of the brain to integrate information, yet this level of

description is not possible with the methods we have described so far. The graph model is a way to describe and quantify the topology of brain networks as well as the topological role of specific nodes in the network (Bullmore & Sporns 2009, Rubinov & Sporns 2010, Sporns et al. 2000, Stam & Reijneveld 2007). In this framework, the whole brain is spatially discretized into a set of nodes that are interconnected by a set of edges. In structural brain networks the edges correspond to white matter projections, whereas in functional networks they represent some measure of pairwise association, such as a Pearson correlation coefficient, mutual information, or transfer entropy. The former can be delineated using chemical tracers (Kötter 2004, Stephan et al. 2001) or diffusion-weighted MRI (Gong et al. 2009, Hagmann et al. 2008, Iturria-Medina et al. 2007), whereas the latter can be derived using virtually any measure of neural activity, such as fMRI (Achard et al. 2006), EEG (Stam et al. 2007), or MEG (Bassett et al. 2006, Stam 2004).

Graph theoretic metrics may describe either the global topological properties of the whole network or the role of a specific node. Moreover, different measures explicitly index either integration or segregation. At the level of individual regions, one can measure their connectedness by counting the total number of connections they have (degree); their tendency to occupy positions along the shortest paths between regions (betweenness); or their redundancy, measured as the fraction of a node's neighbors that are also neighbors of each other (clustering). At the level of the whole network, one can measure the average shortest path length between all pairs of nodes (characteristic path length) or the average clustering in the network. At an intermediate level, one can also profile the community structure of the network by measuring whether the network can be subdivided (modularity) or the frequency with which certain combinations of nodes and edges occur (motifs). Global metrics produce a single value per graph, and in that case statistical assessment of task or group differences is possible using standard univariate tests. Inference on local measures is a bigger challenge because if a separate test is performed for every node one must control the probability of type I errors. Likewise, inference on edges necessitates some form of false discovery rate correction (Zalesky et al. 2010).

Applications. Hagmann et al. (2008) mapped the large-scale anatomical connectivity of the brain using diffusion spectrum imaging and profiled the topological properties of the network. They identified a set of regions in posterior medial and parietal cortex that constitutes a putative structural "core." These regions were hubs by virtue of the fact that they had high degree and betweenness centrality. A modularity analysis revealed that these regions were "'connector" hubs that primarily link multiple modules (as opposed to regions only within a single module). Using a technique called *k*-core decomposition, the authors iteratively removed nodes with degrees lower than *k* until none remained. They were able to strip away most of the cortex down to the core regions, which remained highly mutually interconnected. The analysis revealed a highly central and densely interconnected component along the posterior medial axis of the cortex, situated at the very top of the topological hierarchy.

CONCLUSION

Multivariate statistical analyses have had a tangible effect on theoretical developments in neuroscience. Univariate analyses allow us to ask which regions show changes in activity; exploratory and confirmatory multivariate analyses allow us to investigate which networks show changes in activity as well as how these networks show changes in activity. As a result, interactions among regions in the context of system-level dynamics are an active area of research in imaging neuroscience. The multivariate techniques we have described all represent specific but complementary models of how these interactions are instantiated in the brain.

SUMMARY POINTS

1. Multivariate statistical techniques facilitate network discovery by simultaneously taking into account activity from multiple regions. In this way, they allow inference about the activity of any given region in the context of the entire brain.
2. Two characteristics influence how neuroimaging data are analyzed. First, there are usually more variables (e.g., voxels) than observations (e.g., subjects). Two, there is often a high degree of spatial and/or temporal autocorrelation. As a result, many exploratory techniques (e.g., PCA, ICA, CCA, PLS) are geared toward simplifying or reducing the original data.
3. Most exploratory techniques are mathematically related and feature matrix factorization by SVD. They are used to extract patterns of covariation between all possible pairs of regions and are usually interpreted from the perspective of functional connectivity.
4. Confirmatory techniques involve the fitting and comparison of models that embody hypotheses about causal influences between regions. Thus, they are mainly used to make inferences about effective connectivity. Analyses such as SEM and DCM may be thought of as extreme versions of each other that differ mainly in how they model these influences.
5. Recent tools, such as graph theoretic metrics, characterize neural networks as systems and quantitatively describe the topological role of individual regions in the context of subnetworks and networks.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

Achard S, Salvador R, Whitcher B, Suckling J, Bullmore E. 2006. A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *J. Neurosci.* 26:63–72

Aertsen A, Gerstein G, Habib M, Palm G. 1989. Dynamics of neuronal firing correlation: modulation of "effective connectivity." *J. Neurophysiol.* 61:900–17

Alexander G, Moeller J. 1994. Application of the scaled subprofile model to functional imaging in neuropsychiatric disorders: a principal component approach to modeling brain function in disease. *Hum. Brain Mapp.* 2:79–94

Alexander G, Moeller J, Grady C, Pietrini P, Mentis M, Schapiro M. 1994. Association of cognitive functions with regional networks of brain metabolism in Alzheimer's disease. *Neurobiol. Aging* 15:S36

Bassett D, Meyer-Lindenberg A, Achard S, Duke T, Bullmore E. 2006. Adaptive reconfiguration of fractal small-world human brain functional networks. *Proc. Natl. Acad. Sci. USA* 103:19518–23

Beckmann C, Smith S. 2004. Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Trans. Med. Image* 23:137–52

Beckmann C, Smith S. 2005. Tensorial extensions of independent component analysis for multisubject fMRI analysis. *NeuroImage* 25:294–311

Bollen K. 1986. Sample size and Bentler and Bonett's nonnormed fit index. *Psychometrika* 51:375–77

Bookstein F. 1994. Partial least squares: a dose-response model for measurement in the behavioral and brain sciences. *Psycoloquy* 5(23):1

In this survey, the authors describe SSM as an adaptation of PCA to neuroimaging experiments and include a series of illustrative examples of how SSM can be used to analyze a wide variety of data sets from clinical populations.

The first application of tensor ICA to neuroimaging data.

Breakspear M, Jirsa V. 2007. Neuronal dynamics and brain connectivity. See Jirsa & McIntosh 2007, pp. 3–64
Bressler S, McIntosh A. 2007. The role of neural context in large-scale neurocognitive network operations. See Jirsa & McIntosh 2007, pp. 403–19
Bullmore E, Horwitz B, Honey G, Brammer M, Williams S, Sharma T. 2000. How good is good enough in path analysis of fMRI data? *NeuroImage* 11:289–301
Bullmore E, Sporns O. 2009. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* 10:186–98
Buxton R, Wong E, Frank L. 1998. Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. *Magn. Reson. Med.* 39:855–64
Calhoun V, Adali T, McGinty V, Pekar J, Watson T, Pearlson G. 2001a. fMRI activation in a visual-perception task: network of areas detected using the general linear model and independent components analysis. *NeuroImage* 14:1080–88
Calhoun V, Adali T, Pearlson G, Pekar J. 2001b. A method for making group inferences from functional MRI data using independent component analysis. *Hum. Brain Mapp.* 14:140–51
Calhoun V, Adali T, Pearlson G, Pekar J. 2001c. Spatial and temporal independent component analysis of functional MRI data containing a pair of task-related waveforms. *Hum. Brain Mapp.* 13:43–53
Calhoun V, Kiehl K, Pearlson G. 2008. Modulation of temporally coherent brain networks estimated using ICA at rest and during cognitive tasks. *Hum. Brain Mapp.* 29:828–38
Campbell N, Atchley W. 1981. The geometry of canonical variate analysis. *Syst. Zool.* 30:268–80
Caplan J, McIntosh A, De Rosa E. 2007. Two distinct functional networks for successful resolution of proactive interference. *Cereb. Cortex* 17:1650–63
Carlson T, Schrater P, He S. 2003. Patterns of activity in the categorical representations of objects. *J. Cogn. Neurosci.* 15:704–17
Damoiseaux J, Rombouts S, Barkhof F, Scheltens P, Stam C, et al. 2006. Consistent resting-state networks across healthy subjects. *Proc. Natl. Acad. Sci. USA* 103:13848–53
Daunizeau J, David O, Stephan K. 2009. Dynamic causal modelling: a critical review of the biophysical and statistical foundations. *NeuroImage* 58:312–22
David O, Kiebel S, Harrison L, Mattout J, Kilner J, Friston K. 2006. Dynamic causal modeling of evoked responses in EEG and MEG. *NeuroImage* 30:1255–72
Deshpande G, LaConte S, James G, Peltier S, Hu X. 2009. Multivariate Granger causality analysis of fMRI data. *Hum. Brain Mapp.* 30:1361–73
Diaconescu A, Kovacevic N, McIntosh A. 2008. Modality-independent processes in cued motor preparation revealed by cortical potentials. *NeuroImage* 42:1255–65
Eckart C, Young G. 1936. The approximation of one matrix by another of lower rank. *Psychometrika* 1:211–18
Edgington E. 1995. *Randomization Tests*. Boca Raton, FL: CRC Press
Efron B, Tibshirani R. 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat. Sci.* 1:54–75
Friston K, Buechel C, Fink G, Morris J, Rolls E, Dolan R. 1997. Psychophysiological and modulatory interactions in neuroimaging. *NeuroImage* 6:218–29
Friston K, Dolan R. 2010. Computational and dynamic models in neuroimaging. *NeuroImage* 52:752–65
Friston K, Frith C, Fiddle P, Frackowiak R. 1993. Functional connectivity: the principal-component analysis of large (PET) data sets. *J. Cereb. Blood Flow Metab.* 3:5–14
Friston K, Frith C, Frackowiak R, Turner R. 1995. Characterizing dynamic brain responses with fMRI: a multivariate approach. *NeuroImage* 2:166–72
Friston K, Frith C, Liddle P, Frackowiak R. 1991. Comparing functional (PET) images: the assessment of significant change. *J. Cereb. Blood Flow Metab.* 11:690–99
Friston K, Harrison L, Penny W. 2003. Dynamic causal modelling. *NeuroImage* 19:1273–302
Friston K, Poline J, Holmes A, Frith C, Frackowiak R. 1996. A multivariate analysis of PET activation studies. *Hum. Brain Mapp.* 4:140–51
Garrido M, Kilner J, Kiebel S, Friston K. 2009. Dynamic causal modeling of the response to frequency deviants. *J. Neurophysiol.* 101:2620–31

Introduces the method of PPI. The authors show that effective connectivity may be thought of as an experimentally induced modulation in the functional connectivity between two regions.

Garrido M, Kilner J, Kiebel S, Stephan K, Friston K. 2007. Dynamic causal modelling of evoked potentials: a reproducibility study. *NeuroImage* 36:571–80

Goebel R, Roebroeck A, Kim D, Formisano E. 2003. Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modeling and Granger causality mapping. *Magn. Reson. Imaging* 21:1251–61

Gong G, He Y, Concha L, Lebel C, Gross D, et al. 2009. Mapping anatomical connectivity patterns of human cerebral cortex using in vivo diffusion tensor imaging tractography. *Cereb. Cortex* 19:524–36

Grady C, Protzner A, Kovacevic N, Strother S, Afshin-Pour B, et al. 2010. A multivariate analysis of age-related differences in default mode and task-positive networks across multiple cognitive domains. *Cereb. Cortex* 20:1432–47

Granger C. 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37:424–38

Hagmann P, Cammoun L, Gigandet X, Meuli R, Honey C, et al. 2008. Mapping the structural core of human cerebral cortex. *PLoS Biol.* 6:e159

Hansen L, Larsen J, Nielsen F, Strother S, Rostrup E, et al. 1999. Generalizable patterns in neuroimaging: how many principal components? *NeuroImage* 9:534–44

Haynes J, Rees G. 2005. Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nat. Neurosci.* 8:686–91

Hotelling H. 1936. Relations between two sets of variates. *Biometrika* 28:321–77

Iturria-Medina Y, Canales-Rodríguez E, Melie-García L, Valdes-Hernandez P, Martinez-Montes E, et al. 2007. Characterizing brain anatomical connections using diffusion weighted MRI and graph theory. *NeuroImage* 36:645–60

Jirsa V, McIntosh A, eds. 2007. *Handbook of Brain Connectivity*. Berlin: Springer-Verlag

Jöreskog K, Sörbom D, Magidson J, Cooley W. 1979. *Advances in Factor Analysis and Structural Equation Models*. Cambridge, MA: Abt Books

Jung T, Makeig S, Humphries C, Lee T, Mckeown M, et al. 2000. Removing electroencephalographic artifacts by blind source separation. *Psychophysiology* 37:163–78

Jung T, Makeig S, Westerfield M, Townsend J, Courchesne E, Sejnowski T. 2001. Analysis and visualization of single-trial event-related potentials. *Hum. Brain Mapp.* 14:166–85

Kaminski M, Ding M, Truccolo W, Bressler S. 2001. Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance. *Biol. Cybern.* 85:145–57

Kass R, Raftery A. 1995. Bayes factors. *J. Am. Stat. Assoc.* 90:773–95

Kiebel S, David O, Friston K. 2006. Dynamic causal modelling of evoked responses in EEG/MEG with lead field parameterization. *NeuroImage* 30:1273–84

Kötter R. 2004. Online retrieval, processing, and visualization of primate connectivity data from the CoCoMac database. *Neuroinformatics* 2:127–44

Kovacevic N, McIntosh A. 2007. Groupwise independent component decomposition of EEG data and partial least square analysis. *NeuroImage* 35:1103–12

Krishnan A, Williams L, McIntosh A, Abdi H. 2010. Partial least squares (PLS) methods for neuroimaging: a tutorial and review. *NeuroImage* 56:455–75

Kustra R, Strother S. 2001. Penalized discriminant analysis of [15O]-water PET brain images with prediction error selection of smoothness and regularization hyperparameters. *IEEE Trans. Med. Imaging* 20:376–87

Lobaugh N, West R, McIntosh A. 2001. Spatiotemporal analysis of experimental differences in event-related potential data with partial least squares. *Psychophysiology* 38:517–30

Loehlin J. 1987. *Latent Variable Models: An Introduction to Factor, Path, and Structural Equation Analysis*. Mahwah, NJ: Erlbaum

Makeig S, Jung T, Bell A, Ghahremani D, Sejnowski T. 1997. Blind separation of auditory event-related brain responses into independent components. *Proc. Natl. Acad. Sci. USA* 94:10979–84

Makeig S, Westerfield M, Jung T, Covington J, Townsend J, et al. 1999. Functionally independent components of the late positive event-related potential during visual spatial attention. *J. Neurosci.* 19:2665–80

Makeig S, Westerfield M, Jung T, Enghoff S, Townsend J, et al. 2002. Dynamic brain sources of visual evoked responses. *Science* 295:690–94

Marreiros A, Kiebel S, Friston K. 2008. Dynamic causal modelling for fMRI: a two-state model. *NeuroImage* 39:269–78

McArdle J, McDonald R. 1984. Some algebraic properties of the reticular action model for moment structures. *Br. J. Math. Stat. Psychol.* 37:234–51

Develops the notion of neural context and why it is important to study interactions between brain areas rather than individual brain areas in isolation.

McIntosh A. 1998. Understanding neural interactions in learning and memory using functional neuroimaging. *Ann. N.Y. Acad. Sci.* 855:556–71

McIntosh A. 2000. Towards a network theory of cognition. *Neural Netw.* 13:861–70

McIntosh A, Bookstein F, Haxby J, Grady C. 1996. Spatial pattern analysis of functional brain images using partial least squares. *NeuroImage* 3:143–57

McIntosh A, Chau W, Protzner A. 2004. Spatiotemporal analysis of event-related fMRI data using partial least squares. *NeuroImage* 23:764–75

McIntosh A, 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

A tutorial and review of SEM and its application to neuroscience.

McIntosh A, Gonzalez-Lima F. 1994. Structural equation modeling and its application to network analysis in functional brain imaging. *Hum. Brain Mapp.* 2:2–22

McIntosh A, Grady C, Ungerleider L, Haxby J, Rapoport S, Horwitz B. 1994. Network analysis of cortical visual pathways mapped with PET. *J. Neurosci.* 14:655–66

An overview of the different types of PLS analysis. The authors also discuss the use of resampling techniques for significance testing and estimating reliability.

McIntosh A, Lobaugh N. 2004. Partial least squares analysis of neuroimaging data: applications and advances. *NeuroImage* 23:S250–63

McKeown M, Hansen L, Sejnowski T. 2003. Independent component analysis of functional MRI: What is signal and what is noise? *Curr. Opin. Neurobiol.* 13:620–29

McKeown M, Jung T, Makeig S, Brown G, Kindermann S, et al. 1998a. Spatially independent activity patterns in functional MRI data during the Stroop color-naming task. *Proc. Natl. Acad. Sci. USA* 95:803–10

The first application of ICA to fMRI data.

McKeown M, Makeig S, Brown G, Jung T, Kindermann S, et al. 1998b. Analysis of fMRI data by blind separation into independent spatial components. *Hum. Brain Mapp.* 6:160–88

Mišić B, Schneider B, McIntosh A. 2010. Knowledge-driven contrast gain control is characterized by two distinct electrocortical markers. *Front. Hum. Neurosci.* 3:78

Moeller J, Strother S. 1991. A regional covariance approach to the analysis of functional patterns in positron emission tomographic data. *J. Cerebr. Blood Flow Metab.* 11:A121–35

Moeller J, Strother S, Sidtis J, Rottenberg D. 1987. Scaled subprofile model: a statistical approach to the analysis of functional patterns in positron emission tomographic data. *J. Cerebr. Blood Flow Metab.* 7:649–58

Norman K, Polyn S, Detre G, Haxby J. 2006. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn. Sci.* 10:424–30

Nyberg L, McIntosh A, Cabeza R, Nilsson L, Houle S, et al. 1996. Network analysis of positron emission tomography regional cerebral blood flow data: ensemble inhibition during episodic memory retrieval. *J. Neurosci.* 16:3753–59

Onton J, Delorme A, Makeig S. 2005. Frontal midline EEG dynamics during working memory. *NeuroImage* 27:341–56

Onton J, Westerfield M, Townsend J, Makeig S. 2006. Imaging human EEG dynamics using independent component analysis. *Neurosci. Biobehav. Rev.* 30:808–22

O'Toole A, Jiang F, Abdi H, Pénard N, Dunlop J, Parent M. 2007. Theoretical, statistical, and practical perspectives on pattern-based classification approaches to the analysis of functional neuroimaging data. *J. Cogn. Neurosci.* 19:1735–52

A comprehensive introduction to the logic and mechanics of classification techniques and how they can be applied to neuroimaging experiments.

Pearson K. 1901. On lines and planes of closest fit to systems of points in space. *Philos. Mag.* 2:559–72

Pereira F, Mitchell T, Botvinick M. 2009. Machine learning classifiers and fMRI: a tutorial overview. *NeuroImage* 45:S199–209

Petersson K, Nichols T, Poline J-P, Holmes A. 1999. Statistical methods in functional neuroimaging. I. Non-inferential methods and statistical models. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 354:1239–60

Protzner A, McIntosh A. 2006. Testing effective connectivity changes with structural equation modeling: What does a bad model tell us? *Hum. Brain Mapp.* 27:935–47

Raichle M, MacLeod A, Snyder A, Powers W, Gusnard D, Shulman G. 2001. A default mode of brain function. *Proc. Natl. Acad. Sci. USA* 98:676–82

Roebroeck A, Formisano E, Goebel R. 2005. Mapping directed influence over the brain using Granger causality and fMRI. *NeuroImage* 25:230–42

Rubinov M, Sporns O. 2010. Complex network measures of brain connectivity: uses and interpretations. *NeuroImage* 52:1059–69

Schmithorst V, Holland S. 2004. Comparison of three methods for generating group statistical inferences from independent component analysis of functional magnetic resonance imaging data. *J. Magn. Reson. Imaging* 19:365–68

Schreiber T. 2000. Measuring information transfer. *Phys. Rev. Lett.* 85:461–64

Sporns O, Tononi G, Edelman G. 2000. Theoretical neuroanatomy: relating anatomical and functional connectivity in graphs and cortical connection matrices. *Cereb. Cortex* 10:127–41

Stam C. 2004. Functional connectivity patterns of human magnetoencephalographic recordings: a "small-world" network? *Neurosci. Lett.* 355:25–28

Stam C, Jones B, Nolte G, Breakspear M, Scheltens P. 2007. Small-world networks and functional connectivity in Alzheimer's disease. *Cereb. Cortex* 17:92–99

Stam C, Reijneveld J. 2007. Graph theoretical analysis of complex networks in the brain. *Nonlinear Biomed. Phys.* 1:1–19

Stephan K, Kamper L, Bozkurt A, Burns G, Young M, Kötter R. 2001. Advanced database methodology for the Collation of Connectivity data on the Macaque brain (CoCoMac). *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 356:1159–86

Stephan K, Marshall J, Friston K, Rowe J, Ritzl A, et al. 2003. Lateralized cognitive processes and lateralized task control in the human brain. *Science* 301:384–86

Stephan K, Penny W, Moran R, den Ouden H, Daunizeau J, Friston K. 2010. Ten simple rules for dynamic causal modeling. *NeuroImage* 49:3099–109

Stephan K, Weiskopf N, Drysdale P, Robinson P, Friston K. 2007. Comparing hemodynamic models with DCM. *NeuroImage* 38:387–401

Strother S, Anderson J, Hansen L, Kjems U, Kustra R, et al. 2002. The quantitative evaluation of functional neuroimaging experiments: the NPAIRS data analysis framework. *NeuroImage* 15:747–71

Strother S, Anderson J, Schaper K, Sidtis J, Liow J, et al. 1995. Principal component analysis and the scaled subprofile model compared to intersubject averaging and statistical parametric mapping: I. "Functional connectivity" of the human motor system studied with [15O] water PET. *J. Cerebr. Blood Flow Metab.* 15:738–53

Theiler J, Eubank S, Longtin A, Galdrikian B, Doyne Farmer J. 1992. Testing for nonlinearity in time series: the method of surrogate data. *Phys. D* 58:77–94

Thomas C, Harshman R, Menon R. 2002. Noise reduction in bold-based fMRI using component analysis. *NeuroImage* 17:1521–37

Vakorin V, Krakovska O, McIntosh A. 2009. Confounding effects of indirect connections on causality estimation. *J. Neurosci. Methods* 184:152–60

Wold H. 1982. Soft modelling: the basic design and some extensions. In *Systems Under Indirect Observation: Causality-Structure-Prediction*, ed. H Wold, K Joreskog, 2:1–54. Amsterdam: North Holland

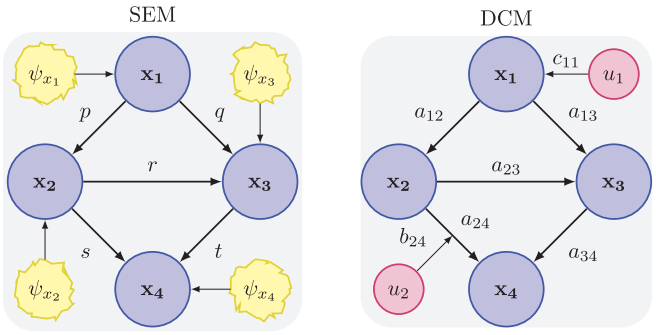
Worsley K, Poline J, Friston K, Evans A. 1997. Characterizing the response of PET and fMRI data using multivariate linear models. *NeuroImage* 6:305–19

Zalesky A, Fornito A, Harding I, Cocchi L, Yücel M, et al. 2010. Whole-brain anatomical networks: Does the choice of nodes matter? *NeuroImage* 50:970–83

Discusses advantages and disadvantages of univariate and multivariate models specifically from the perspective of neuroimaging. The authors argue that there is no universally correct analytic framework; rather, the choice of analysis depends on the experimental question.

A tutorial on how imaging data can be represented as graphs as well as the most commonly used graph theoretic metrics for cognitive neuroscience.

Thoroughly reviews the logic behind DCM and the ways in which it can be used to model and draw inferences about effective connectivity.



$$\text{SEM} \quad \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ p & 0 & 0 & 0 \\ q & r & 0 & 0 \\ 0 & s & t & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} + \begin{pmatrix} \psi_{x_1} \\ \psi_{x_2} \\ \psi_{x_3} \\ \psi_{x_4} \end{pmatrix}$$

$$\text{DCM} \quad \begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \end{pmatrix} = \left(\begin{pmatrix} 0 & a_{12} & a_{13} & 0 \\ 0 & 0 & a_{23} & 0 \\ 0 & 0 & 0 & a_{34} \\ 0 & 0 & 0 & 0 \end{pmatrix} + u_2 \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & b_{24} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \right) \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} + \begin{pmatrix} c_{11} & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$$

Figure 1

An example of typical structural equation modeling (SEM) (*left*) and dynamic causal modeling (DCM) (*right*). In SEM, causal order is specified by a system of linear regression equations with one set of path coefficients (β) and error terms (ψ). In DCM, causal order is specified by a system of differential equations parameterized in terms of synaptic couplings (**A**) as well as exogenous inputs (**u**) that may influence either the synaptic couplings between regions (**B**) or intrinsic activity in individual regions (**C**).

Annual Review of Psychology
Volume 64, 2013

Contents

Indexes

Errata

An online log of corrections to *Annual Review of Psychology* articles may be found at http://psych.AnnualReviews.org/errata.shtml