NeuroImage **11,** 289–301 (2000)
doi:10.1006/nimg.2000.0544, available online at http://www.idealibrary.com on IDEAL®

How Good Is Good Enough in Path Analysis of fMRI Data?

Ed Bullmore,*,†,[1] Barry Horwitz,‡ Garry Honey,* Mick Brammer,* Steve Williams,* and Tonmoy Sharma*

*Institute of Psychiatry, King's College, London, United Kingdom; ‡National Institute on Deafness and other Communication Disorders, National Institutes of Health, Bethesda, Maryland; and †Department of Psychiatry, University of Cambridge, Addenbrooke's Hospital, Cambridge CB2 2QQ, United Kingdom

Received August 3, 1999

This paper is concerned with the problem of evaluating goodness-of-fit of a path analytic model to an interregional correlation matrix derived from functional magnetic resonance imaging (fMRI) data. We argue that model evaluation based on testing the null hypothesis that the correlation matrix predicted by the model equals the population correlation matrix is problematic because *P* values are conditional on asymptotic distributional results (which may not be valid for fMRI data acquired in less than 10 min), as well as arbitrary specification of residual variances and effective degrees of freedom in each regional fMRI time series. We introduce an alternative approach based on an algorithm for automatic identification of the best fitting model that can be found to account for the data. The algorithm starts from the null model, in which all path coefficients are zero, and iteratively unconstrains the coefficient which has the largest Lagrangian multiplier at each step until a model is identified which has maximum goodness by a parsimonious fit index. Repeating this process after bootstrapping the data generates a confidence interval for goodness-of-fit of the best model. If the goodness of the theoretically preferred model is within this confidence interval we can empirically say that the theoretical model could be the best model. This relativistic and data-based strategy for model evaluation is illustrated by analysis of functional MR images acquired from 20 normal volunteers during periodic performance (for 5 min) of a task demanding semantic decision and subvocal rehearsal. A model including unidirectional connections from frontal to parietal cortex, designed to represent sequential engagement of rehearsal and monitoring components of the articulatory loop, is found to be irrefutable by hypothesis-testing and within confidence limits for the best model that could be fitted to the data.

INTRODUCTION

Path analytic models are interesting to many neuroscientists because they can quantify functional relationships between multiple brain regions in terms of unidirectional connections and, historically, many models of integrated brain function have assumed hierarchical relationships between information-processing modules. To take a seminal example of this kind of model, consider the Wernicke–Lichtheim diagram for language (Lichtheim, 1885) (Fig. 1). Here three brain regions are specialized for clinicopathologically dissociable functions of hearing words (input processing), speaking words (output processing), and conscious or intelligent mediation between input and output. The sequential or directional flow of information between these regions, indicated diagramatically by single-headed arrows, is modelled by two possible paths: from input to output directly (simple repetition) or from input to output via a vaguely localised center for "conscious elaboration of concepts." This particular model may no longer be regarded as tenable. But the principles of neuromodular organisation it pioneered remain central to many contemporary models of working memory, reading, visual perception, and other complex functions (Fodor, 1983; Shallice, 1988), as well as dysconnectionist models of many psychiatric and neurological disorders (Geschwind, 1965; Paulesu *et al.*, 1996; Bullmore *et al.*, 1997).

Path analysis of functional imaging data offers an opportunity to test such models empirically. The theoretically anticipated connections between p brain regions are written down in the form of a path model matrix, and the numerical values of the q nonzero path coefficients in the matrix are estimated so as to minimize a measure of discrepancy between the observed interregional correlation matrix $\mathbf{C}$ and the correlation matrix predicted by the model $\Sigma(\theta)$. The question then arises: is the theoretical model good enough?

One way of addressing this question is by a test of the null hypothesis that the population correlation matrix Σ is equal to the correlation matrix predicted by the

[1] To whom correspondence should be addressed at e-mail: etb23@cam.ac.uk; fax: +44(0)1223 336581.

1053-8119/00 $35.00

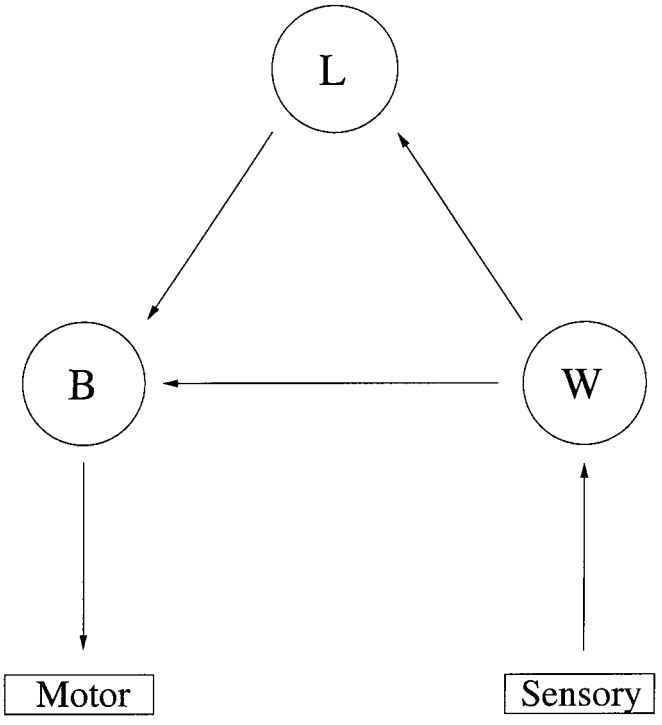

FIG. 1. The Wernicke–Lichtheim diagram. Wernicke's area of left temporoparietal cortex (W) is specialized for language perception; Broca's area of left inferior frontal cortex (B) is specialized for language production; Lichtheim's vaguely localized center (L) is specialized for conscious manipulation of language. The arrows drawn between regions represent unidirectional effects of one region on the function of another, mediated by anatomically defined connections, i.e., the arcuate fasciculus subserves the direct effect of Wernicke's area on Broca's area, $W \rightarrow B$. Historically, the empirical basis for this model was clinicopathological correlation between specific language dysfunctions or dysphasias and distinct lesions to gross brain anatomy at postmortem.

model: $H_0\ \Sigma = \Sigma(\theta)$. Under this hypothesis, some measures of discrepancy between observed and predicted correlation matrices are asymptotically distributed as chi square. More formally, the value of the maximum likelihood discrepancy function $F(\mathrm{C}, \Sigma(\theta))$ multiplied by $\nu - 1$, where ν is the number of independent observations on each variable, is approximately distributed as χ^2 on $\frac{1}{2}p(p+1) - q$ degrees of freedom. If the observed discrepancy has a small probability under this distribution, say $P < 0.1$, then the null hypothesis is formally refuted and we can be fairly sure that the model is *not* good enough. But what if the probability of the discrepancy between observed and predicted correlation matrices is larger than our arbitrary threshold for refutation, say $P = 0.21$? This might be upheld as a necessary condition for the model to be considered good enough but it is not by itself sufficient, on several grounds.

First, there are a number of technical reasons why the particular P value assigned to a given model should not be regarded absolutely. For example, P will generally get smaller as the number of data points n per voxel gets larger. This means that a model which might be regarded as acceptable based on analysis of 5 min of fMRI data could well be refuted if the same experimental stimuli were presented over the course of 10 min (assuming a constant interval between consecutive images). A related issue is that, due to temporal autocorrelation in fMRI time series, the number of independent observations will generally be less than the number of MR measurements per voxel: $\nu \leq n$. The correct scalar for calculation of chi square is probably ν rather than n, but there is as yet no general agreement about this, nor about the best method for estimating ν in a given time series, nor about the best procedure to follow if ν differs markedly between regional time series.

Another controversial area concerns prior estimation of residual variances. In a path analysis of p variables there are potentially $p \times (p-1)$ path coefficients to be estimated as well as p residual variances, usually denoted $\psi_i\ i = 1, 2, 3, \ldots, p$, which indicate the proportion of total variance in each region that is *not* accounted for by the effects of other regions. In path analysis of positron emission tomography (PET) data, it has often been the practice to assign an arbitrary value to regional residual variances *a priori*, i.e., to set $\psi = 0.4$ or 0.35 (McIntosh and Gonzalez-Lima, 1994; McIntosh *et al.*, 1994). This practice has the merit of reducing the number of model parameters to be estimated on the basis of the limited number ($k = \frac{1}{2}p(p+1)$) of nonredundant elements in the observed correlation matrix, thereby allowing greater freedom in specification of the connections between regions. However, it is well known that if the chosen value for ψ is close to 0 the minimum value of the discrepancy function for a given model will be greater than if the chosen value for ψ is close to 1 (McIntosh *et al.*, 1994). It would probably be preferable to have a less arbitrary, more empirical, procedure for estimating residual variances individually for each region, but this is not yet established. Meanwhile, different analysts making different assumptions about ψ might be led to very different conclusions about the adequacy of a given model to account for the same set of data.

Besides these issues concerning estimation of both ν and F, the chi square approximation for the null distribution of their product, $(\nu - 1) \times F(\mathrm{C}, \Sigma(\theta))$, entails several assumptions which may or may not be justified by the data. The most important of these is probably that the correlation matrix has been estimated on the basis of a "large" number of (independent) observations. To be a little more precise, the chi square approximation may be unsound if $\nu < 200$ (Bollen, 1988), which will often be the case for fMRI data acquired in less than 10 min. Under these circumstances, the chi square distribution tends to be overconservative, leading to excessively frequent rejection of the model because the assigned P value is incorrectly small.

In addition to these "small print" difficulties, there are broader problems with null hypothesis testing in path analysis of fMRI data. The logical basis of this approach is that the system under investigation is already well understood by the analyst, who can strongly predict a certain set of unidirectional connections between variables and wishes to affirm or refute those predictions once and for all. A more realistic account of the fMRI data analyst's predicament is that he or she will have only incomplete knowledge about the system. For example, it is currently uncertain how rapid, sequential engagement of a series of distinct processing centers at a neural level might be manifest at the level of slower haemodynamic changes measurable by fMRI (Horwitz and Sporns, 1994; Horwitz *et al.*, 1999). Thus even if there was good prior knowledge, e.g., from single unit recording studies in monkeys, that presentation of some stimulus caused discharge in region A, which preceded by several milliseconds a secondary increase in activity of region B, what are the implications for path modelling of fMRI data in humans? Does it unquestionably follow that the only model worth testing is one that has the variance in B unilaterally determined by A? The short answer, in our view, is no.

It seems much more natural, given the current state of knowledge about the biophysical basis of fMRI, to adopt an exploratory rather than confirmatory approach. This might mean separately fitting both possible one-path models (A → B and B → A) and using measures of model goodness to decide which of the two is better. Such a *relativistic* approach has indeed been preferred in many prior path analyses of functional imaging data (McIntosh and Gonzalez-Lima, 1994; McIntosh *et al.*, 1996; Büchel and Friston, 1997). However, it is important to be clear that tests of the null hypothesis that there is zero difference in goodness-of-fit between two models, based on the difference in their chi square criteria (often denoted χ^2_{diff}), are compromised by many of the same technical issues already identified as problematic in relation to null hypothesis testing of a single model.

In this paper, we develop an alternative inferential approach to path analysis of fMRI data, which may circumvent some of these problems. The method uses an automated search algorithm to find the best model for a given set of data and prior constraints. "Best" is operationally defined by a parsimonious fit index which incorporates both the discrepancy between observed and modelled data and its cost (in terms of the number of model parameters estimated). Bootstrapping is used to sample the distribution of best model fit indices. The goodness of any arbitrary model preferred on theoretical grounds can then be assessed relative to the confidence interval for goodness of the best model that can be fitted to the data.

The methods are introduced in the context of fMRI data acquired from 20 healthy volunteers performing a periodically designed semantic decision task. A more detailed description of these data and some preliminary analyses is provided in the next section of the paper; this is followed by a treatment of the path analytic methods and results.

DATA AND PRELIMINARY ANALYSES

Subjects and Data Acquisition

Functional MRI data were acquired from $m = 20$ right-handed healthy volunteers (11 male, 9 female; mean age = 44.4 years, standard deviation (SD) = 14.4 years) using a 1.5 T GE Signa system (General Electric, Milwaukee, WI) at the Maudsley Hospital, London UK. Written informed consent was provided by all subjects. The study was approved by the Bethlem Royal and Maudsley NHS Trust Ethics (Research) Committee.

Gradient-echo echoplanar T2*-weighted images depicting blood oxygen level dependent (BOLD (Ogawa *et al.*, 1990)) contrast were acquired from 14 noncontiguous near axial planes: TE = 40 ms, TR = 3 s, flip angle = 90°, number of signal averages = 1, slice thickness = 7 mm, slice skip = 0.7 mm, in-plane resolution = 3 mm. This multislice acquisition protocol introduces slight temporal offsets between MR data measured within the same TR in different axial planes. These effects were corrected by linear interpolation before recovery of regional fMRI time series.

Experimental Design

During scanning, each subject viewed a 5-min display of alternating 30-s epochs of two conditions, A and B. During the A (activation) condition, the subject was cued every 2.5 s by presentation of a word, e.g., "goat," to decide whether the word represented a living or nonliving object and to internally or subvocally articulate that decision. All words were in the high frequency range and between three and seven letters in length (Kucera and Francis, 1967). During the B (baseline) condition, the subject was asked simply to fixate on an isoluminant screen. Five cycles of alternation between epochs were presented in the course of each experiment; the B condition was always presented first.

Activation Mapping and Regional Time Series Analysis

Following estimation and correction of movement-related effects (Bullmore *et al.*, 1999b), periodic signal change at the (fundamental) frequency of alternation between A and B conditions was estimated by sinusoidal regression. The model was fit by a pseudogeneralised least squares (PGLS) procedure, treating the residuals of an initial (ordinary) least squares fit as a first order autoregressive (AR1) process (Bullmore *et al.*, 1996a). The standardized power of response at fundamental frequency was estimated at each voxel

and these power maps were registered in standard stereotaxic space (Talairach and Tournoux, 1988; Bullmore *et al.*, 1999b). Voxels demonstrating significant median power of response over all 20 subjects were identified by a permutation test with one-tailed probability of type 1 error $P < 0.0001$ (Brammer *et al.*, 1997) and displayed as a generic brain activation map (GBAM).

As shown in Fig. 2, there were five main regions of generic activation, consistent with the results of previous studies using the same methods (Bullmore *et al.*, 1999a):

- *Ventral extrastriate cortex* (VEC). Bilateral activation of fusiform, lingual, and middle occipital gyri (approximate Brodmann areas (BA) 18, 19), extending anteriorly to inferior temporal gyrus (BA 37).
- *Prefrontal cortex* (PFC). Predominantly left-sided activation of dorosolateral prefrontal cortex (BA 45, 9), extending posteriorly and superiorly to lateral premotor cortex (BA 6).
- *Supplementary motor area* (SMA). Medial activation of superior premotor cortex (BA 6).
- *Inferior frontal gyrus* (IFG). Predominantly left-sided activation of BA 44 and 45, corresponding to Broca's area on the left.
- *Inferior parietal lobule* (IPL). Predominantly left-sided activation of supramarginal and angular gyri (BA 39, 40), extending medially to precuneus (BA 7) and anteriorly and inferiorly to superior temporal gyrus (BA 21).

The motion-corrected fMRI time series volume for each subject was also registered in standard space and a set of five regional fMRI time series was recovered from the left hemisphere in each subject. Each regional time series was the average of the time series at the voxel indexed by regional coordinates in stereotactic space (see Table 1) and its eight nearest neighbors in 2-D (total regional volume = 0.57 cm^3). Following correction of head movement and multislice acquisition, there were 96 MR data points remaining in each time series. The ($m = 20 \times n = 96$) data matrix for a given region is denoted R_i. The rows of this matrix were standardized to zero mean and unit variance.

Principal components analysis was used to identify an "average" pattern of response to the experimental design over all subjects in each region (Fletcher *et al.*, 1999; Büchel *et al.*, 1999). Writing the singular value decomposition of $R_i = U_i L_i V'_i$, where $'$ denotes matrix transpose, the columns of V_i are the eigenvectors or "eigentimeseries" of R_i and the min $(n, m) = 20$ nonzero diagonal elements of $L_i = \lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_{20}$ are the

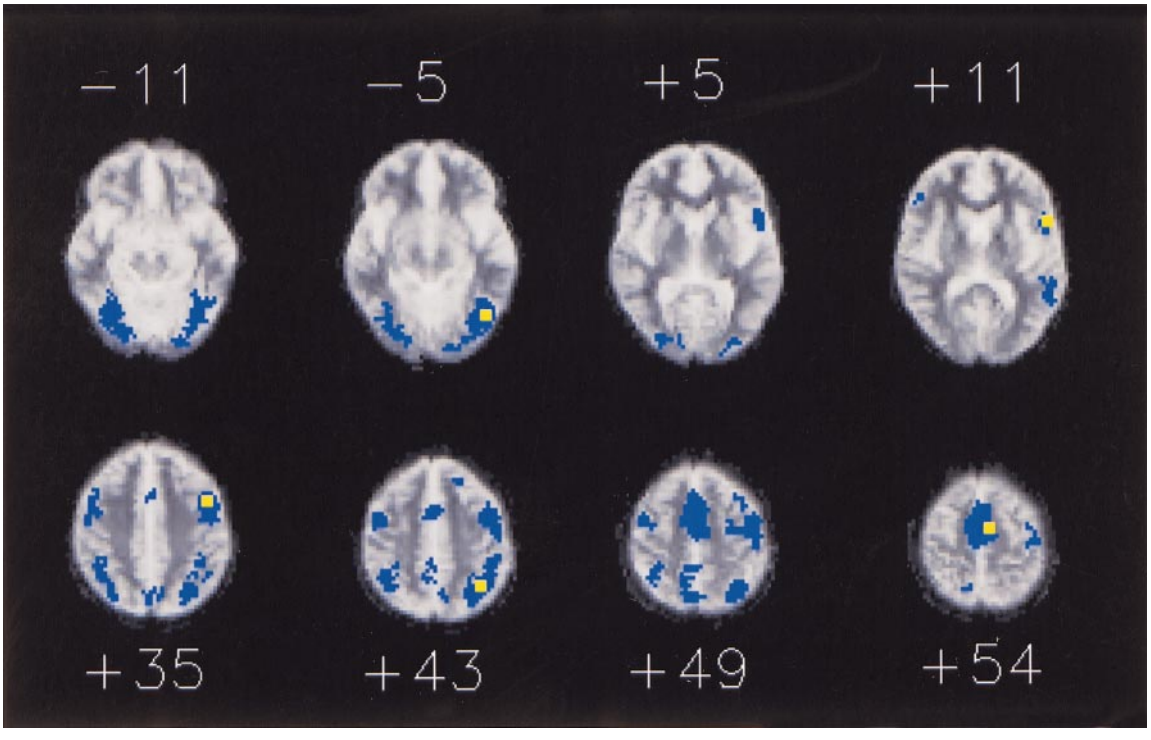

FIG. 2. Generic brain activation maps. Voxels demonstrating significant median power of response to the periodic experimental input function over all 20 subjects are colored red; voxelwise probability of false-positive error $P < 0.0001$. Blue squares indicate 3×3 voxel squares from which five regional fMRI timeseries were recovered from each subject's dataset. The z coordinate of each panel in the standard space of Talairach and Tournoux is given in millimeters. Following radiological convention, the right side of the brain is shown on the left side of each panel.

TABLE 1

Interregional Correlation Matrix, Residual Variances ψ, Effective Degrees of Freedom ν, and Talairach Coordinates $\{x, y, z\}$ for Five Left Hemispheric Brain Regions

	VEC	PFC	SMA	IFG	IPL
VEC	1				
PFC	0.661	1			
SMA	0.525	0.660	1		
IFG	0.486	0.507	0.437	1	
IPL	0.731	0.630	0.558	0.517	1
ψ	0.825	0.868	0.870	0.881	0.851
ν	19.4	29.3	37.4	40.9	24.2
$\{x, y, z\}$	−40, −65, −5	−40, 5, 35	−3, −5, 54	−48, 7, 10	−33, −50, 43

square roots of the eigenvalues of R_i (Bullmore *et al.*, 1996b).

The first eigentimeseries for each region was clearly dominated by a periodic trend at the experimental frequency of alternation between conditions; see Fig. 3.

The pairwise correlations $c_{i,j}$ between the first eigentimeseries for the ith and jth regions constituted the $(p \times p)$ interregional correlation matrix C that was later subject to path analysis; see Table 1. The residual variance for each region was estimated by the ratio between the first eigenvalue and the sum of eigenvalues:

$$\psi_i = 1 - \frac{\lambda_1^2}{\sum_{j=1}^{m} \lambda_j^2}. \quad (1)$$

To estimate the number of independent observations ν_i or effective degrees of freedom for each regional eigentimeseries, we first fitted a sinusoidal regression model to it by pseudogeneralised least squares, yielding a serially uncorrelated (white noise) process $\{\epsilon_t\}$ $t = 1, 2, 3, \ldots, n$ (Bullmore *et al.*, 1996a). We then derived an estimate of ν_i as the ratio of white noise (residual) variance to total eigentimeseries variance (first eigenvalue), multiplied by the nominal degrees of freedom or

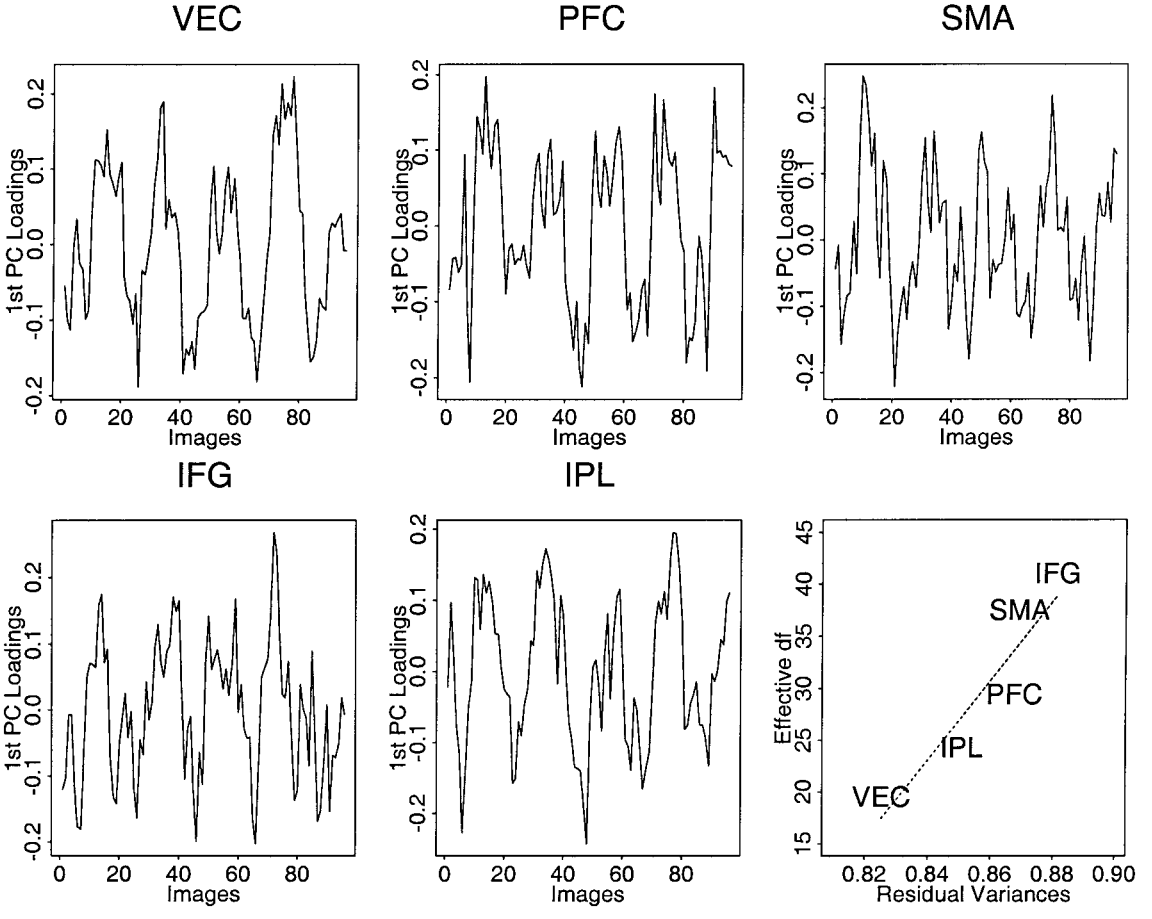

FIG. 3. The first eigenvector or eigentimeseries for each of five brain regions is shown in the first five panels (from left to right, top to bottom). Note that, in each region, the dominant mode of variance over all 20 subjects includes a periodic trend at the frequency of the experimental design. The sixth panel shows the strong association ($r = 0.92$) between the residual variance and effective degrees of freedom over all regions.

number of data points per voxel, n:

$$\nu_i = \frac{\mathrm{Var}\{\epsilon_t\}}{\lambda_1^2} \times n. \quad (2)$$

Residual variances and effective degrees of freedom given by these estimators for each of the five regions are shown in Fig. 3 and Table 1.

PATH ANALYSIS

Theoretical Model Specification

The first step in path analysis is formulation of a theoretically principled model. This can involve setting some possible path coefficients to zero on the grounds that there is no evidence for a direct, anatomical connection between the regions in question. Use of anatomical data in this way is complicated by the fact that much of it has been obtained by study of monkeys. And it is not always certain which area of, say, the macaque brain is homologous to a given area of human brain, especially if the human brain area is specialised for language or some other uniquely human function. Even in cases where translation between human and nonhuman primate anatomy is unproblematic or can be agreed by convention, the primate data are likely to show that each region is reciprocally interconnected to a large number of other regions. This means that anatomical data alone cannot often justify setting to zero just one of the two possible unidirectional connections between regions. Specification of the direction of effects must therefore depend on a functional analysis of the task, under anatomical constraints.

For these data, the only intrahemispheric pair of regions for which we could find *no* evidence of direct anatomical connectivity was VEC and SMA (Mesulam, 1990; McGuire *et al.*, 1991; Young, 1993). Adoption of this constraint leaves $q = (p \times (p - 1)) - 2 = 18$ path coefficients to be estimated on the basis of $k = 15$ nonredundant elements in the correlation matrix. This is clearly not an identified model, without further constraints on the number of path coefficients to be estimated. A subset of anatomically possible connections was therefore selected, as shown in Fig. 4. This model comprises $q = 6$ nonzero path coefficients. VEC and IPL are reciprocally connected to model the fact that these regions may share a common visual input and are strongly correlated in functional connectivity analysis of PET data acquired during a word-reading task (Horwitz *et al.*, 1998). Since the verbal stimuli used in this experiment are all familiar words, rather than pseudowords taxing phonological processes (Rumsey *et al.*, 1997), the main input analysis region is assumed to be VEC, and a connection is drawn forward from VEC to PFC to model the process of semantic analysis and decision. Then a series of connections is drawn from PFC to IPL via SMA and IFG. This may be understood as a simple model of an inner speech circuit or articulatory loop. SMA is putatively responsible for endogenously directing inner speech production by IFG, and IPL is assumed to function as a phonological

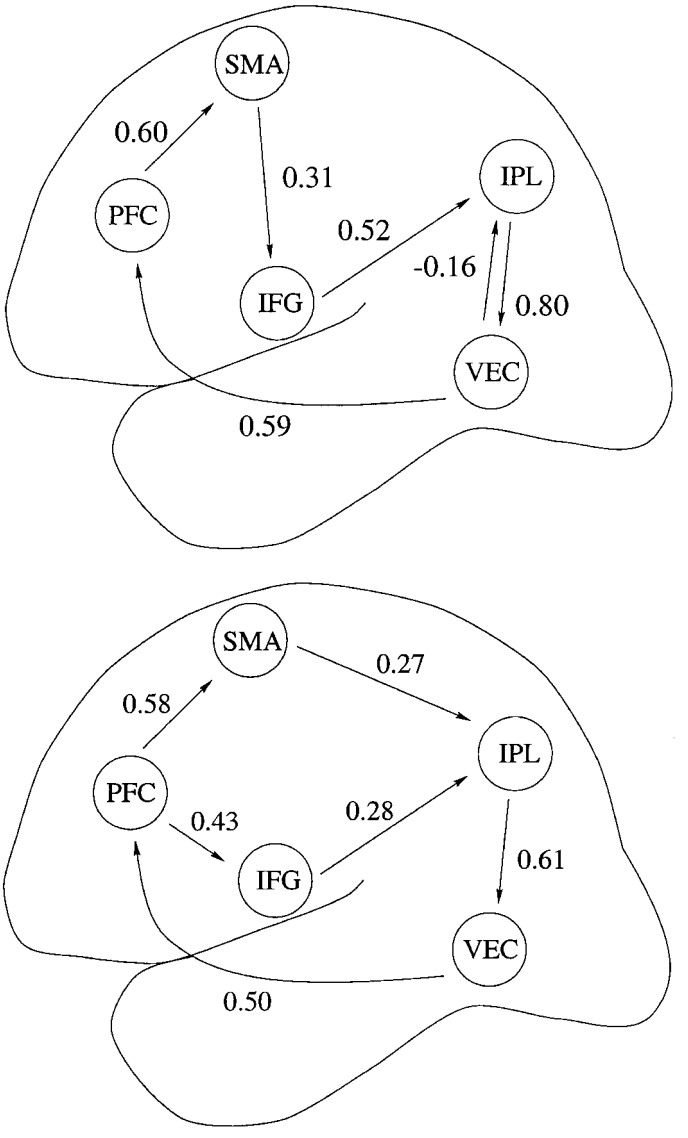

FIG. 4. Two path models fitted to the same data from 20 normal subjects performing a semantic decision and inner speech task. Top: The theoretically preferred path model. Maximum likelihood estimates of the $q = 6$ nonzero path coefficients are shown. Bottom: The best fitting path model that can be automatically identified to account for the data. "Best" is operationally defined as the maximum of Bollen's parsimonious fit index. The two models have several features in common, e.g., the number of nonzero coefficients and the direction of paths from VEC to PFC and from frontal to parietal areas. The models are also similar in terms of overall goodness-of-fit. From these and more formal inferential results described in the text, we conclude that it is not easy to find a much better model for these data than the theoretically preferred model.

buffer monitoring subvocal output from Broca's area. This modular decomposition is based on exploratory multivariate analyses of other fMRI data acquired under identical conditions (Bullmore *et al.*, 1996a), as well as previous PET studies of semantic decision and inner speech (Paulesu *et al.*, 1993; Kapur *et al.*, 1994; Paulesu *et al.*, 1996) and some nonhuman primate physiological data (Müller-Preuss and Jürgens, 1976; Frith, 1992).

Path Model Fitting

The objective is to find estimates for the q non-zero path coefficients $\{\theta_j\}$ $j = 1, 2, 3, \ldots, q$, which minimize a measure of discrepancy between the observed correlation matrix C and the correlation matrix predicted by the model $\Sigma(\theta)$. A convenient way of doing this is to write down the model as a set of simultaneous regression equations, for example:

$$\begin{pmatrix} \text{VEC} \\ \text{PFC} \\ \text{SMA} \\ \text{IFG} \\ \text{IPL} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & \theta_1 \\ \theta_2 & 0 & 0 & 0 & 0 \\ 0 & \theta_3 & 0 & 0 & 0 \\ 0 & 0 & \theta_4 & 0 & 0 \\ \theta_6 & 0 & 0 & \theta_5 & 0 \end{pmatrix} \times \begin{pmatrix} \text{VEC} \\ \text{PFC} \\ \text{SMA} \\ \text{IFG} \\ \text{IPL} \end{pmatrix} + \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \\ \psi_5 \end{pmatrix}$$

or more generally,

$$v = \mathsf{K}v + \psi, \tag{3}$$

where v denotes the $(p \times 1)$ vector of regional variances, ψ denotes the $(p \times 1)$ vector of residual variances, and K denotes the $(p \times p)$ path model matrix. Eq. (3) can be rearranged to show that

$$v = (\mathsf{I} - \mathsf{K}^{-1})\psi. \tag{4}$$

The correlation matrix predicted by the path model $\Sigma(\theta)$ is then given by the McArdle-McDonald equation (McArdle and McDonald, 1984):

$$\Sigma(\theta) = (\mathsf{I} - \mathsf{K}^{-1})\psi\psi'(\mathsf{I} - \mathsf{K}^{-1})'. \tag{5}$$

Estimates of the path coefficients are found by iteratively minimising the maximum likelihood (ML) discrepancy function,

$$F = \log|\Sigma(\theta)| + tr(\mathsf{C}\Sigma^{-1}(\theta)) - \log|\mathsf{C}| - q, \tag{6}$$

where $|.|$ denotes the determinant, and *tr* denotes the trace of a matrix.

These results were implemented using a general optimisation function, nlminb() in S-Plus (Venables and Ripley, 1994); they have also been implemented in MatLab as part of the SPM software package (Büchel and Friston, 1997).

For these data, the minimum of F multiplied by the mean number of independent observations over all regions ($\overline{\nu} = 30.3$) was 12.67. This has $P = 0.18$ under the null hypothesis that it is distributed as χ^2 on $k - q = 9$ *df*. The model is therefore not refuted. This means that the model provides a good account of the data but it leaves unanswered the question: how easy would it be to find a better-fitting model?

AUTOMATED SEARCH FOR THE "BEST" PATH MODEL

The automated search for the best fitting model starts from the worst fitting model—that is the "null model" in which all path coefficients are constrained or set to zero. The process is analogous to "step up" procedures for multiple regression model specification. The algorithm computes the Lagrangian multiplier (LM) for each constrained coefficient and allows the coefficient with the maximum LM to be nonzero. This $q = 1$ path model is fitted to the data and Lagrangian multipliers again computed for all constrained coefficients. The coefficient with maximum LM is allowed to be nonzero, this $q = 2$ model is fitted to the data, and the process is iterated until as many coefficients as possible are unconstrained.

The Lagrangian multipliers $\boldsymbol{L}$, or modification indices as they are called in LISREL and some other proprietary software packages (Loehlin, 1987), are a function of the first and second order partial derivatives of the ML discrepancy function with respect to the vector θ of all (constrained and unconstrained) path coefficients:

$$\boldsymbol{L} = \frac{\nu - 1}{2}\left(\frac{\partial F}{\partial \theta}\right)'\left(\frac{\partial^2 F}{\partial \theta \partial \theta'}\right)^{-1}\left(\frac{\partial F}{\partial \theta}\right). \tag{7}$$

The derivatives can be evaluated by a simple numerical method (Cudeck *et al.*, 1993) based on the equalities:

$$\begin{aligned} \frac{\partial^2 F}{\partial \theta_i \partial \theta_j} = h_{i,j} &\approx \frac{1}{2}\, tr\{\Sigma^{-1}(\theta)C_i\Sigma^{-1}(\theta)C_j\} \\ \frac{\partial F}{\partial \theta_i} &= \frac{1}{2}\, tr\{\Sigma^{-1}(\theta)[\Sigma(\theta) - \mathsf{C}]\Sigma^{-1}(\theta)C_i\}. \end{aligned} \tag{8}$$

The second order partial derivatives $h_{i,j}$ $i, j = 1, 2, 3, \ldots, p$ are elements of the symmetric $(p \times p)$ Hessian matrix H. The matrix C_i comprises the partial derivatives of the modelled covariance matrix with respect to the ith parameter. These matrices can be computed for

each parameter by finite forward differences, i.e.,

$$C_i = \frac{\Sigma(\theta + \boldsymbol{e}_i\eta) - \Sigma(\theta)}{\eta}, \quad (9)$$

where $\boldsymbol{e}_i$ denotes a vector of length $p \times (p-1)$ with only one nonzero element, which is unity at the ith position, and η is an arbitrarily small constant, e.g., $\eta = 10^{-4}$.

The maximum Lagrangian multiplier, at any step in the search, identifies the parameter estimate for which the local slope of the discrepancy function is steepest. Unconstrained (nonzero) model parameters should have Lagrangian multipliers equal to zero, since estimates for these parameters have been found by minimising the discrepancy function. Constrained parameters will only have Lagrangian multipliers equal to zero if the constraints are justified. The maximum Lagrangian multiplier may therefore be understood to identify the currently constrained parameter which is least justifiably set to zero.

Automatically unconstraining the parameter with maximum LM at each iteration of the search procedure may, however, lead to specification of an unidentified model. Simply restricting the number of unconstrained parameters q to be less than the number of nonredundant elements in the correlation matrix k is a necessary but not a sufficient condition of identifiability (Bollen, 1988). But identifiability *is* guaranteed if the model is recursive. (Note that recursive, in the technical sense used here, perhaps counterintuitively means there can be no reciprocal connections between any pair of regions.) An alternative way of stating this condition is that the model matrix K can be written in triangular form. To avoid fitting unidentified models, we therefore restricted the search for the maximum LM at each iteration to those currently constrained parameters which could be freed without making the model nonrecursive.

For $p = 5$, these numerical methods automatically specify and fit a series of recursive models in which the number of free (nonzero) path coefficients is incrementally increased from zero to ten. The minimum value of the discrepancy function, and therefore chi square, will monotonically decrease as the number of paths in the model is increased. So the "best" model, in terms of least discrepancy between observed and modelled correlation matrices, will always be the most elaborate model in the series. However, in the eyes of many analysts, an interesting model is probably one that explains as much as possible (of the observed data) for as little as possible cost (in number of model parameters). Chi square is not sensitive to the cost of model goodness but it can be combined with the number of model parameters q, or the degrees of freedom $k - q$, to provide various measures of goodness that are sensitive to its cost. One example is Akaike's information criterion A:

$$A = \chi^2_q + 2q. \quad (10)$$

As path coefficients are unconstrained, chi square for the model with q nonzero paths χ^2_q will decrease and $2q$ will increase. If the decrement $\chi^2_q - \chi^2_{q+1} < 2$, then A will have a minimum at q and this can be taken as an operational definition of the best model. An alternative measure is Bollen's (1986) parsimonious fit index ρ:

$$\rho = \frac{(\chi^2_0/k) - (\chi^2_q/\, k - q)}{\chi^2_0/k}. \quad (11)$$

Here χ^2_0 denotes chi square for the null model, in which all coefficients are set to zero. If the q path model provides a perfect fit, then ρ will have its maximum value of one. An operational definition of the best model in a series of nested models is therefore the model with maximum ρ. This metric has some advantages over Akaike's information criterion. For example, it generally exists in the unit interval (negative values are possible but in practice extremely unlikely) and it is independent of the number of observations ν, thereby facilitating comparison between models estimated on the basis of observed or resampled fMRI data which have different effective degrees of freedom.

Chi square, P value, Akaike's information criterion and Bollen's parsimonious fit index are shown for each model in the search series in Fig. 5. The best model, in terms of maximum probability and maximum ρ, has $q = 6$ nonzero path coefficients and is shown in Fig. 4. It clearly shares many features with the theoretically preferred model, including forward projections from VEC to PFC and from PFC to SMA. It is also notable that in both models the IPL receives projections from frontal areas, although the theoretically preferred series of connections SMA → IFG → IPL is replaced in the automatically identified model by parallel projections from both SMA and IFG to IPL.

BOOTSTRAPPING BEST MODEL GOODNESS

We can immediately see, by comparison of parsimonious fit indices, that the theoretically preferred model ($\rho = 0.72$) is not as good as the automatically optimized model ($\rho = 0.75$). However, the automated procedure may well have "capitalised on chance," that is to say it may have selected paths which happened to have large Lagrangian multipliers because of idiosyncratic properties of the sample rather than because they represented connections between regions that are consistently important in the population. This possibility may be discounted somewhat by the fact that the correlation

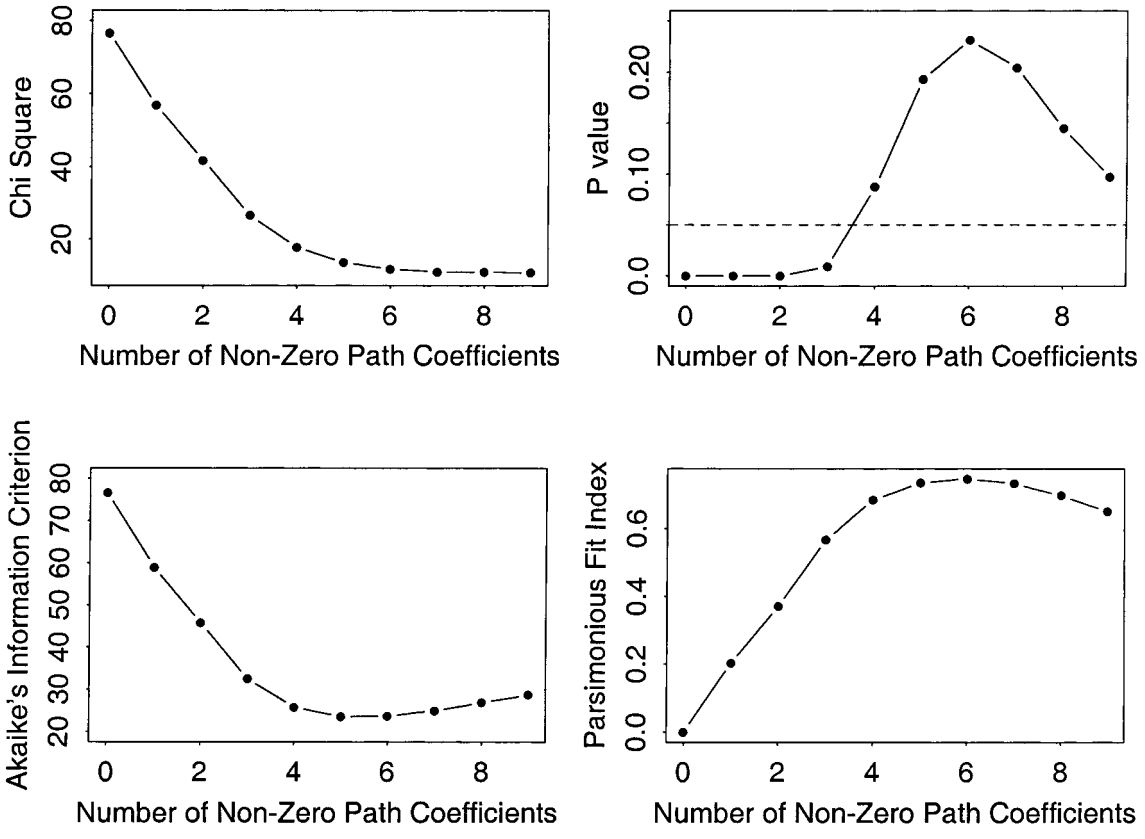

FIG. 5. Several measures of model goodness for each of a series of automatically specified path models. Top left, chi square monotonically decreases as the number of nonzero path coefficients in the model increases; top right, *P* value for chi square exceeds 0.05 once the number of nonzero coefficients q is greater than 3 and has a maximum when $q = 6$; bottom left, Akaike's information criterion has a minimum when $q = 5$; bottom right, Bollen's parsimonious fit index has a maximum when $q = 6$. Any merit function with a maximum or minimum may be used to provide an operational definition of the "best" model in a series. We have preferred Bollen's fit index because it is not conditional on the number of independent observations comprising each fMRI time series v and therefore facilitates comparison of model goodness between observed and resampled data.

matrix under investigation represents data pooled over 20 subjects by a preliminary principal components analysis. It may therefore seem unlikely that a particular interregional correlation will be substantially influenced by a single outlier among the subjects. Nevertheless, there will clearly be sampling variation in the correlation matrix and in the statistics derived from it, including best model fit indices.

To quantify the susceptibility of the automated search procedure to sampling variation, we used bootstrapping (Efron and Tibshirani, 1993). To do this, $\mathtt{m} = 20$ subjects were randomly sampled with replacement from the $m = 20$ subjects in the observed sample, and the fMRI time series for each resampled subject at the ith region was used to construct the $(\mathtt{m} \times n)$ bootstrapped data matrix R_i. Analysis of these data matrices then proceeded exactly as already described in detail for the observed data matrices R_i, to the point of automated path model search and optimal model identification. The whole process was repeated $\mathtt{B} = 200$ times.

Final products of this analysis include a bootstrap distribution for the parsimonious fit index of the best model. The 2.5 and 97.5% percentile values of this distribution can be taken as limits of a 95% confidence interval (CI). For best model goodness, operationally defined by the maximum of Bollen's fit index, the 95% CI estimated by this bootstrap procedure was [0.56, 0.97]. This interval includes the value of ρ for the theoretically preferred model fitted to the observed data.

DISCUSSION

It is good, but probably not good enough, to find an irrefutable path model for an fMRI dataset. If one attempts to conclude evaluation of model goodness at this point, questions may reasonably be asked about the dependence of the *P* value on asymptotic distributional assumptions and other details of the analysis, such as prior estimation of residual variances and effective degrees of freedom. It seems much better,

much more affirmative of the theoretically preferred model, if one can show that it is not only irrefutable but also that its goodness is within a confidence interval for goodness of the best model that can be fitted by identical methods to the data in hand. If so, we can legitimately say that the theoretical model *could be* the best model, and that this conclusion is not conditional on asymptotic assumptions underpinning the more conventional hypothesis-testing approach, nor on the estimators adopted to provide prior estimates of residual variances and effective degrees of freedom.

Of course, this approach to model evaluation also raises many questions. How is goodness of any particular model to be quantified? How can we expeditiously search through the very large number of possible models to find the one which is best? How can we construct a confidence interval for best model goodness?

Here we have considered two possible metrics of goodness: Akaike's information criterion and Bollen's parsimonious fit index. Both measures operationally define goodness in terms of a balance between the *adequacy* of the model to account for the data and *complexity* of the model in terms of the number of nonzero path coefficients estimated. The model which comes closest to satisfying these conflicting requirements will have minimum value of Akaike's criterion or maximum value of Bollen's fit index. We have preferred Bollen's index because it is not conditional on the mean number of degrees of freedom in the regional fMRI timeseries and it is therefore particularly suitable for comparing model goodness between observed and bootstrapped data, which will potentially differ in terms of the number of independent observations. Bollen's index can also appropriately be used on the basis of discrepancy functions other than the maximum likelihood function we have used. Since other discrepancy functions, such as the ordinary least squares function, require less intensive computation (fewer matrix inversions) this might be important in minimizing the computational cost of the bootstrapping procedure.

To identify the best model, by whatever criterion is adopted, we cannot realistically fit *all* possible models. We have to use a search strategy that reduces to a minimum the number of models that must be time-consumingly fitted to the data. It will also be preferable for the strategy to be implemented automatically, both to save time and to eliminate unreliability in optimal model identification due to human error. Various approaches have been advocated for automated and efficient covariance structural model specification; see Glymour *et al.* (1987) for review. The particular approach we have taken here is a "step up" search starting from the null model, in which all path coefficients are constrained to be zero, and iteratively proceeding to specify and fit a nested series of models by unconstraining the fixed parameter with the maximum Lagrangian multiplier at each step of the search. Lagrangian multipliers have been commonly used to guide model specification in path analysis and other forms of covariance structural equation modelling (Loehlin, 1987; Bollen, 1988) and they have the considerable advantage for this purpose of being quick to compute. Nevertheless, using the Lagrangian multipliers in the context of an automated search for the best model is relatively novel and has not previously been applied in the context of functional MRI or PET data analysis.

There are arguably at least two ways in which implementation of this strategy could be refined. First, there is the question of identifiability. Automated model searches will occasionally lead to specification of unidentified models, i.e., models that do not have unique solutions. Such models can often be recognized empirically by the fact that their parameters may have values that are at or beyond their boundary values. For example, a standardized path coefficient in an unidentified model may be assigned an "impossible" value of greater than 1 or less than -1. To avoid this problem, we have guaranteed identifiability of automatically specified models by constraining them to be recursive. This means that we have not (automatically) fitted models that include reciprocal connections between regions. This constraint is simple to implement and is justifiable by our prior interest in using path analysis to test modular models of brain organisation: because modular models, as narrowly defined (Fodor, 1983; Shallice, 1988), *are* recursive. However, there are other sufficient conditions of identifiability that could be used instead. For example, one could fit the model indicated by the maximum LM and then evaluate its Hessian; see Eq. (8). If the inverse of the Hessian is singular, the model is unidentified (Bollen, 1988), and the search would have to go back one step, unconstrain the parameter with the immediately submaximal Lagrangian multiplier, then try going forward again. This is potentially a more time-consuming guarantee of identifiability to adopt, but may be less restrictive on the kind of models that will be specified, than the recursive condition we have used. A second possible kind of refinement would be to allow the algorithm to "step down" as well as "step up." Once a path coefficient has been unconstrained by the current algorithm it will remain unconstrained, even if its role in accounting for the observed correlation matrix is substantially diminished by subsequently unconstrained coefficients. This creates the possibility of redundancy in model specification and it might be advisable for this reason to review nonzero parameter estimates at each iteration and try constraining any parameters with values close to zero. If such a path is truly important to the model, it will presumably be associated with a large Lagrangian multiplier once it is (again) constrained to zero, and

this will lead to it (again) being unconstrained on the next step-up.

The third question relates to assessment of sampling variation. Here we are trying to evaluate goodness of a theoretical model with respect to the sampling distribution of the fit index for the best model that can be fitted to the data. This is not a statistic for which there is obviously an asymptotic approximation in Normal theory. We have therefore obtained estimates of its standard error by repeatedly resampling with replacement, or boot-strapping, the originally observed fMRI time series data. This approach is conceptually simple and adaptable to statistics with theoretically intractable distributions. Its only major disadvantage is that it is computationally demanding. To identify the best model for each of 200 bootstrap samples required almost 12 h of central processing unit time on a Sun Ultra 1 workstation. This demand would increase sharply if the number of regions included in the model was increased; but of course if computational load became a major issue the code could be rewritten in a compiled language, rather than an interpreted language such as S-Plus or Matlab, or otherwise optimized for maximal efficiency (for example, by using an ordinary least squares discrepancy function as discussed above).

A final methodological question is how this approach relates to previous imaging applications of path analysis. Often path analysis has been used to test the hypothesis that there is a difference in interregional connectivity between two groups of subjects scanned under the same experimental conditions, or a difference within the same group of subjects scanned under two experimental conditions. A parametric test of these hypotheses based on the difference in fit indices for the two models χ^2_{diff} will be conditional on the validity of asymptotic assumptions and the choice of methods used to estimate residual variance and effective degrees of freedom. We can illustrate this by a simple thought experiment. Imagine, for example, that the covariance structure in the two groups truly is different but we have assumed (somewhat pessimistically) that the residual variance is 0.99 before estimating chi square separately in each group. This means that both chi squares will be small and the difference between them may well be too small to refute the null hypothesis. If the test procedure is repeated after our prior estimate of the residual variance is reduced to, say, 0.5 then both chi squares will be larger and the difference between them may now be large enough correctly to refute the null hypothesis. However, the computationally more intensive approach to inference advocated here could be generalised to a comparison between models that was less conditional on these important details. For example, one could compute χ^2_{diff} for the two groups of observed data and then after repeated reassignment of the data to two groups at random. This would allow a permutation test of the null hypothesis that χ^2_{diff} estimated in the observed data was not determined by differences between subjects related to the grouping variable; see Edgington (1995) and Good (1994) for general introductions to permutation testing. Alternatively, one could bootstrap the data in one group to construct confidence intervals for path coefficients and refer the coefficients estimated in the other group to those intervals. This would allow a more precise, path-by-path characterization of between-group difference than comparisons based on global goodness of fit measures. Both these tests should be robust to assumptions about (or estimation of) residual variance or effective degrees of freedom, provided the same assumption (or estimator) is used consistently to derive chi square for all observed and resampled groups. It is also worth noting that computational inference could be used to test null hypotheses about the covariance structure of data acquired from a single subject (Bollen and Stine, 1992) and to study interregional covariances in residual fMRI data after preliminary removal of experimentally determined effects.

We now briefly consider more substantive questions about the validity of the model. This is essentially a modular model for a linguistic task. Modular models have a long history in studies of brain and language, dating back to seminal work of the 19th century diagram-makers (Geschwind, 1967). But of course the longevity of modular language models does not guarantee their validity; it may instead reveal more about our limitations in conceiving of or empirically characterizing neural systems for language by more valid but more complex or emergent models. Nevertheless, this particular experiment can be construed as prescribing a series of component processes (word recognition, semantic analysis, and subvocal articulation) for each of which there is some prior evidence of modularity from neuropsychology and functional imaging. We presumed that ventral extrastriate and prefrontal cortices were responsible for word recognition and semantic analysis; a series of connections from prefrontal cortex via SMA to Broca's area were responsible for subvocal planning and articulation; and inferior parietal lobule was responsible for monitoring the subvocal output of Broca's area. The prior evidence for this model is probably strongest for the articulatory rehearsal and monitoring components of the system. There are behavioral data to suggest that articulatory and phonological storage functions are dissociable (Longoni *et al.,* 1993) and neuropsychological and functional imaging data to suggest that these functions are discretely localized in inferior frontal and inferior parietal cortices, respectively (Vallar and Baddeley, 1984; Paulesu *et al.,* 1993; Awh *et al.,* 1996). An effective connection from Broca's area to IPL during performance of a task demanding inner speech

is consistent with this literature but has not previously been demonstrated by functional imaging. However, a more secure functional interpretation of this and other path coefficients in the model will require further experiments in which, for example, the demands of rehearsal or monitoring might be parametrically adjusted in terms of word-length or phonemic similarity. It would be compatible with our interpretation of these data if the effective connection between inferior frontal and inferior parietal components of the articulatory loop was modulated by such manipulations of the articulatory load on the system.

Even assuming that our modular decomposition of this task is indeed correct, or at least reasonable on the basis of the available data, a more basic question remains: how does serial information processing at a neural level, in the order of milliseconds, map to an apparently analogous covariance structure at the much coarser temporal resolution of these fMRI data? This seems an important question in general for path analysis of fMRI data and one which may require convergent approaches from large scale neural modelling and biophysics to address.

ACKNOWLEDGMENTS

E. B. was supported by the Wellcome Trust. Thanks to friends and colleagues who welcomed E. B. and family to Washington, DC in Spring 1997. This work was presented at the 5th International Conference on Functional Mapping of the Human Brain, June 22–26, 1999, Düsseldorf, Germany.

REFERENCES

Awh, E., Jonides, J., Smith, E. E., Schumacher, E. H., Koeppe, R. A., and Katz, S. 1996. Dissociation of storage and rehearsal in verbal working memory: Evidence from positron emission tomography. *Psychol. Sci.* **7:**25–31.

Bollen, K. A. 1986. Sample size and Bentler and Bonett's non-normed fit index. *Psychometrika* **51:**375–377.

Bollen, K. A. 1988. *Structural Equations with Latent Variables*. Wiley, New York.

Bollen, K. A., and Stine, R. A. 1992. Bootstrapping goodness-of-fit measures in structural equation models. *Sociol. Methods Res.* **21:**205–229.

Brammer, M. J., Bullmore, E. T., Simmons, A., Williams, S. C. R., Grasby, P. M., Howard, R. J., Woodruff, P. W. R., and Rabe-Hesketh, S. 1997. Generic brain activation mapping in fMRI: A nonparametric approach. *Magn. Reson. Imag.* **15:**763–770.

Büchel, C., and Friston, K. J. 1997. Modulation of connectivity in visual pathways by attention: Cortical interactions evaluated with structural equation modelling and fMRI. *Cerebral Cortex* **7:**768–778.

Büchel, C., Coull, J., and Friston, K. J. 1999. The predictive value of changes in effective connectivity for human learning. *Science* **283:**1538–1541.

Bullmore, E. T., Rabe-Hesketh, S., Morris, R. G., Williams, S. C. R., Gregory, L., Gray, J. A., and Brammer, M. J. 1996a. Functional magnetic resonance image analysis of a large-scale neurocognitive network. *NeuroImage* **4:**16–33.

Bullmore, E. T., Brammer, M. J., Williams, S. C. R., Rabe-Hesketh, S., Janot, N., David, A. S., Mellers, J. D. C., Howard, R., and Sham, P. 1996b. Statistical methods of estimation and inference for functional MR image analysis. *Magn. Reson. Med.* **35:**261–277.

Bullmore, E. T., Frangou, S., and Murray, R. M. 1997. The dysplastic net hypothesis: An integration of developmental and dysconnectionist theories of schizophrenia. *Schizophrenia Res.* **28:**143–156.

Bullmore, E. T., Brammer, M. J., Williams, S. C. R., Curtis, V. A., McGuire, P. K., Morris, R. G., Murray, R. M., and Sharma, T. 1999a. Functional MR imaging of confounded hypofrontality. *Hum. Brain Mapp.* **8:**86–91.

Bullmore, E. T., Brammer, M. J., Rabe-Hesketh, S., Curtis, V., Williams, S. C. R., Sharma, T., and McGuire, P. K. 1999b. Methods for diagnosis and treatment of stimulus correlated motion in generic brain activation studies using fMRI. *Hum. Brain Mapp.* **7:**38–48.

Cudeck, R., Klebe, K. J., and Henly, S. J. 1993. A simple Gauss-Newton procedure for covariance structure analysis with high-level computer languages. *Psychometrika* **58:**211–232.

Edgington, E. S. 1995. *Randomisation Tests*. Marcel Dekker, New York.

Efron, B., and Tibshirani, R. J. 1993. *An Introduction to the Bootstrap*. Chapman & Hall, New York.

Fletcher, P., Büchel, C., Josephs, O., Friston, K. J., and Dolan, R. J. 1999. Learning-related neuronal responses in prefrontal cortex studied with functional neuroimaging. *Cerebral Cortex* **9:**168–178.

Fodor, J. A. 1983. *The Modularity of Mind*. MIT Press, Cambridge, MA.

Frith, C. D. 1992. *The Cognitive Neuropsychology of Schizophrenia*. Lawrence Erlbaum Associates, Hove, UK.

Geschwind, N. 1965. Disconnexion syndromes in animals and man. *Brain* **88:**237–294 and 585–644.

Geschwind, N. 1967. Wernicke's contribution to the study of aphasia. *Cortex* **3:**449–463.

Glymour, C., Scheines, R., Spirtes, P., and Kelly, K. 1987. *Discovering Causal Structure: Artifical Intelligence, Philosophy of Science, and Statistical Modeling*. Academic Press, New York.

Good, P. J. 1994. *Permutation Tests*. Springer Verlag, New York.

Horwitz, B., and Sporns, O. 1994. Neural modeling and functional neuroimaging. *Hum. Brain Mapp.* **1:**269–283.

Horwitz, B., Rumsey, J. M., and Donohoe, B. C. 1998. Functional connectivity of the angular gyrus in normal reading and dyslexia. *Proc. Nat. Acad. Sci. USA* **95:**8939–8944.

Horwitz, B., Tagamets, M-A., and McIntosh, A. R. 1999. Neural modeling, functional brain imaging and cognition. *Trends Cogn. Sci.* **3:**91–98.

Kapur, S., Rose, R., Liddle, P. F., Zipursky, R., Brown, G. M., Stuss, D., Houle, S., and Tulving, E. 1994. The role of the left prefrontal cortex in verbal processing: Semantic processing or willed action? *NeuroReport* **5:**2193–2196.

Kucera, H., and Francis, W. N. 1967. *Computational Analysis of Present Day American English*. Brown Univ. Press, Providence, RI.

Lichtheim, L. 1885. On aphasia. *Brain* **7:**433–484.

Loehlin, J. C. 1987. *Latent Variable Models: An Introduction to Factor, Path and Structural Analysis*. Lawrence Erlbaum Associates, Hillsdale, NJ.

Longoni, A. M., Richardson, J. T., and Aiello, A. 1993. Articulatory rehearsal and phonological storage in working memory. *Memory Cogn.* **21:**11–22.

McArdle, J. J., and McDonald, R. P. 1984. Some algebraic properties of the reticular action model for moment structures. *Br. J. Math. Stat. Psychol.* **37:**234–251.

McGuire, P. K., Bates, J. F., and Goldman-Rakic, P. S. 1991. Interhemispheric integration I: Symmetry and convergence of the corticocortical connections of the left and right principal sulcus and the left and right supplementary motor area in the rhesus monkey. *Cerebral Cortex* **1:**390–407.

McIntosh, A. R., and 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. R., Grady, C. L., Haxby, J. V., Ungerleider, L. G., and Horwitz, B. 1994. Network analysis of cortical visual pathways mapped with PET. *J. Neurosci.* **14:**655–666.

McIntosh, A. R., Grady, C. L., Ungerleider, L. G., and Horwitz, B. 1996. Changes in limbic and prefrontal functional interactions in a working memory task. *Cerebral Cortex* **6:**571–584.

Mesulam, M-M. 1990. Large scale neurocognitive networks and distributed processing for attention, language, and memory. *Ann. Neurol.* **28:**597–613.

Müller-Preuss, P., and Jürgens, U. 1976. Projections from the "cingular" vocalisation area in the squirel monkey. *Brain Res.* **103:**29–43.

Ogawa, S., Lee, T. M., Kay, A. R., and Tank, D. W. 1990. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proc. Natl. Acad. Sci. USA* **87:**9868–9872.

Paulesu, E., Frith, C. D., and Frackowiak, R. S. J. 1993. The neural correlates of the verbal component of working memory. *Nature* **362:**242–345.

Paulesu, E., Frith, U., Snowling, M., Gallagher, A., Morton, J., Frackowiak, R. S. J., and Frith, C. D. 1996. Is developmental dyslexia a disconnection syndrome? Evidence from PET scanning. *Brain* **119:**143–157.

Rumsey, J. M., Horwitz, B., Donohue, B. C., Nace, K., Maisog, J. M., and Andreason, P. 1997. Phonological and orthographic components of word recognition: A PET-rCBF study. *Brain* **120:**739–759.

Shallice, T. 1988. *From Neuropsychology to Mental Structure.* Cambridge Univ. Press, Cambridge.

Talairach, J., and Tournoux, P. 1988. *A Coplanar Stereotactic Atlas of the Human Brain*. Thieme Verlag, Stuttgart.

Vallar, G., and Baddeley, A. D. 1984. Fractionation of working memory: Neuropsychological evidence for a phonological store. *J. Learn. Verbal Behav.* **23:**151–161.

Venables, W. D., and Ripley, B. D. 1994. *Modern Applied Statistics with S-PLUS*. Springer Verlag, New York.

Young, M. P. 1993. The organization of neural systems in the primate cerebral cortex. *Proc. R. Soc. Series B* **252:**13–18.