Annu. Rev. Psychol. 2005. 56:57–87
doi: 10.1146/annurev.psych.56.091103.070311
Copyright © 2005 by Annual Reviews. All rights reserved
First published online as a Review in Advance on August 24, 2004

MODELS OF BRAIN FUNCTION IN NEUROIMAGING

Karl J. Friston
Wellcome Department of Cognitive Neurology, University College London, London, WC1N 3BG, United Kingdom; email: k.friston@fil.ion.ucl.ac.uk

Key Words dynamic, inference, causal, Bayesian, fMRI, hemodynamics, connectivity

■ **Abstract** Inferences about brain function, using neuroimaging data, rest on models of how the data were caused. These models can be quite diverse, ranging from conceptual models of functional anatomy to nonlinear mathematical models of hemodynamics. However, they all have to be internally consistent because they model the same thing. This consistency encompasses many levels of description and places constraints on the statistical models, adopted for data analysis, and the experimental designs they embody. The aim of this review is to introduce the key models used in imaging neuroscience and how they relate to each other. We start with anatomical models of functional brain architectures, which motivate some of the fundaments of neuroimaging. We then turn to basic statistical models (e.g., the general linear model) used for making classical and Bayesian inferences about where neuronal responses are expressed. By incorporating biophysical constraints, these basic models can be finessed and, in a dynamic setting, rendered causal. This allows us to infer how interactions among brain regions are mediated.

CONTENTS

INTRODUCTION 58
ANATOMIC MODELS 58
  Functional Specialization and Integration 58
  Functional Specialization and Segregation 59
STATISTICAL MODELS OF REGIONAL RESPONSES 60
  Statistical Parametric Mapping 60
  The General Linear Model 61
  Experimental Design 64
  Classical and Bayesian Inference 67
  Dynamic Models 69
  Biophysical Models 71
MODELS OF FUNCTIONAL INTEGRATION 77
  Functional and Effective Connectivity 77
  Dynamic Causal Modeling 79
CONCLUSION 84

0066-4308/05/0203-0057$14.00

INTRODUCTION

Understanding the brain depends on conceptual, anatomical, statistical, and causal models that link ideas about how it works to observations and experimental data. The aim of this review is to highlight the relationships among the sorts of models that are employed in imaging neuroscience. These relationships reflect an increasing mechanistic finesse as one moves from simple statistical models used to identify where evoked brain responses are expressed (cf. neo-phrenology) to models of how neuronal responses are caused (e.g., causal modeling). In the near future, models of representational inference and learning may be used as observation models to confirm some fundamental hypotheses about what the brain is doing (e.g., predictive coding). We review a series of exemplar models that cover conceptual models, motivating experimental design, to detailed biophysical models of coupled neuronal ensembles that enable questions to be asked at a physiological and computational level.

Anatomical models of functional brain architectures motivate the fundaments of neuroimaging. We start by reviewing the distinction between functional specialization and integration and how these principles serve as the basis for most analyses of neuroimaging data. We then turn to simple statistical models (e.g., the general linear model) used for making classical and Bayesian inferences about functional specialization in terms of where neuronal responses are expressed. Characterizing a region-specific effect rests on estimation and inference. Inferences in neuroimaging may be about differences expressed when comparing one group of subjects to another or, within subjects, inferences may be about changes over a sequence of observations. They may pertain to structural differences (e.g., in voxel-based morphometry; Ashburner & Friston 2000) or neurophysiological indices of brain functions (e.g., functional magnetic resonance imaging, or fMRI). The principles of data analysis are very similar for all these applications. We focus on the analysis of fMRI time-series because this covers most of the issues encountered in other modalities. By incorporating biophysical constraints, simple observation models can be rendered biologically more realistic and, in a dynamic framework, causal. This allows us to infer how interactions among brain regions are mediated. This sort of characterization speaks to functional integration and relies on the notions of functional and effective connectivity.

ANATOMIC MODELS

Functional Specialization and Integration

The brain appears to adhere to two fundamental principles of functional organization, functional specialization and functional integration, where the integration within and among specialized areas is mediated by effective connectivity. The distinction relates to that between "localizationism" and "[dis]connectionism" that dominated thinking about cortical function in the nineteenth century. Since the early anatomic theories of Gall, the identification of a particular brain region with a specific function has become a central theme in neuroscience. However, functional localization per se was not easy to demonstrate: For example, a meeting that took place on August 4, 1881, addressed the difficulties of attributing function to a cortical area, given the dependence of cerebral activity on underlying connections (Phillips et al. 1984). This meeting was entitled "Localization of Function in the Cortex Cerebri." Goltz (1881), although accepting the results of electrical stimulation in dog and monkey cortex, considered that the excitation method was inconclusive, in that movements elicited might have originated in related pathways, or current could have spread to distant centers. In short, the excitation method could not be used to infer functional localization because localizationism discounted interactions, or functional integration among different brain areas. It was proposed that lesion studies could supplement excitation experiments. Ironically, it was observations on patients with brain lesions some years later (see Absher & Benson 1993) that led to the concept of "disconnection syndromes" and the refutation of localizationism as a complete or sufficient explanation of cortical organization. Functional localization implies that a function can be localized in a cortical area, whereas specialization suggests that a cortical area is specialized for some aspects of perceptual or motor processing, and that this specialization is anatomically segregated within the cortex. The cortical infrastructure supporting a single function may then involve many specialized areas whose union is mediated by the functional integration among them. In this view, functional specialization is only meaningful in the context of functional integration and vice versa.

Functional Specialization and Segregation

The functional role of any component (e.g., cortical area, subarea, or neuronal population) of the brain is defined largely by its connections (Passingham et al. 2002). Certain patterns of cortical projections are so common that they could amount to rules of cortical connectivity. "These rules revolve around one, apparently, overriding strategy that the cerebral cortex uses—that of functional segregation" (Zeki 1990). Functional segregation demands that cells with common functional properties be grouped together. This architectural constraint necessitates both convergence and divergence of cortical connections. Extrinsic connections among cortical regions are not continuous but occur in patches or clusters. This patchiness has, in some instances, a clear relationship to functional segregation. For example, V2 has a distinctive cytochrome oxidase architecture consisting of thick stripes, thin stripes, and interstripes. When recordings are made in V2, directionally selective (but not wavelength or color-selective) cells are found exclusively in the thick stripes. Retrograde (i.e., backward) labeling of cells in V5 is limited to these thick stripes. All the available physiological evidence suggests that V5 is a functionally homogeneous area that is specialized for visual motion. Evidence of this nature supports the notion that patchy connectivity is the anatomical infrastructure that mediates functional segregation and specialization. If it is the case that neurons in a given cortical area share a common responsiveness, by virtue of their extrinsic connectivity, to some sensorimotor or cognitive attribute, then this functional segregation is also an anatomical one.

In summary, functional specialization suggests that challenging a subject with the appropriate sensorimotor attribute or cognitive process should lead to activity changes in, and only in, the specialized areas. This is the anatomical and physiological model upon which the search for regionally specific effects is based. We deal first with models of regionally specific responses and later return to models of functional integration.

STATISTICAL MODELS OF REGIONAL RESPONSES

Statistical Parametric Mapping

Functional mapping studies are usually analyzed with some form of statistical parametric mapping. Statistical parametric mapping entails the construction of spatially extended statistical processes to test hypotheses about regionally specific effects (Friston et al. 1991). Statistical parametric maps (SPMs) are image processes with voxel values that, under the null hypothesis, are distributed according to a known probability density function, usually the student's T or F distributions. These are known colloquially as T- or F-maps. The success of statistical parametric mapping is due largely to the simplicity of the idea. Namely, one analyzes every voxel using any standard (univariate) statistical test, usually one testing for activation or regression on some explanatory variable. The resulting statistical parameters are assembled into an image—the SPM. SPMs are interpreted as spatially extended statistical processes by referring to the probabilistic behavior of random fields (Adler 1981; Friston 1991; Worsley et al. 1992, 1996). Random fields model both the univariate probabilistic characteristics of an SPM and any nonstationary spatial covariance structure under the null hypothesis. "Unlikely" excursions of the SPM are interpreted as regionally specific effects, attributable to the sensorimotor or cognitive process that has been manipulated experimentally.

Over the years, statistical parametric mapping (Friston et al. 1995b) has come to refer to the conjoint use of the general linear model (GLM) and Gaussian random field (GRF) theory to analyze and make classical inferences about spatially extended data through statistical parametric maps. The GLM is used to estimate some parameters that could explain the spatially continuous data in exactly the same way as in conventional analysis of discrete data. GRF theory is used to resolve the multiple-comparisons problem that ensues when making inferences over a volume of the brain. GRF theory provides a method for adjusting p values for the search volume of an SPM to control false positive rates. It plays the same role for continuous data (i.e., images or time-series) as the Bonferonni correction for a family of discontinuous or discrete statistical tests.

Below we consider the Bayesian alternative to classical inference with SPMs. This rests on conditional inferences about an effect, given the data, as opposed to classical inferences about the data, given the effect is zero. Bayesian inferences about spatially extended effects use posterior probability maps (PPMs). Although less established than SPMs, PPMs are potentially very useful, not least because they do not have to contend with the multiple-comparisons problem induced by classical inference (see Berry & Hochberg 1999). In contradistinction to SPM, this means that inferences about a given regional response do not depend on inferences about responses elsewhere. Before looking at the models underlying Bayesian inference, we first consider estimation and classical inference in the context of the GLM.

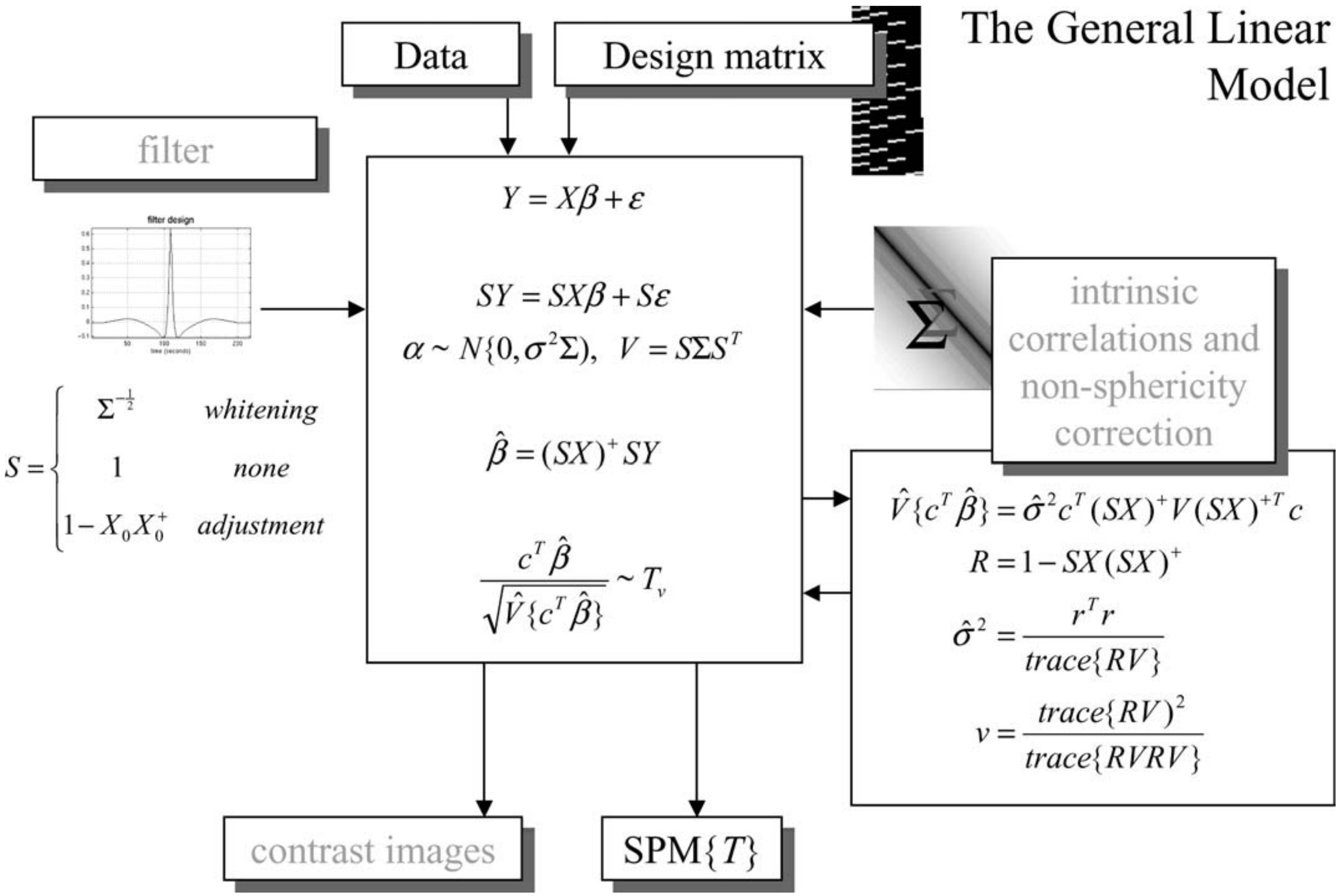
The General Linear Model

Statistical analysis of imaging data corresponds to (*a*) modeling the data to partition observed neurophysiological responses into components of interest, confounds, and error, and (*b*) making inferences about interesting effects using the variances of the partitions. A brief review of the literature may give the impression that there are numerous ways to analyze positron emission tomography (PET) and fMRI time-series, with a diversity of statistical and conceptual approaches. This is not the case. With very few exceptions (see Nichols & Holmes 2002 for an overview), every analysis is a variant of the general linear model. These include (*a*) simple T-tests on scans assigned to one condition or another, (*b*) correlation coefficients between observed responses and boxcar stimulus functions in fMRI (Bandettini et al. 1993), (*c*) inferences made using multiple linear regression, (*d*) evoked responses estimated using linear time invariant models, and (*e*) selective averaging to estimate event-related responses. Mathematically, all are identical and can be implemented with the same equations and algorithms. The only thing that distinguishes among them is the design matrix encoding the experimental design.

The general linear model is an equation,

$$y = X\beta + \varepsilon, \tag{1}$$

expressing the observed response y in terms of a linear combination of explanatory variables in the matrix X plus a well-behaved error term (i.e., an independently and identically distributed Gaussian random variable). The general linear model is variously known as "analysis of [co]variance" or "multiple regression" and subsumes simpler variants, like the T-test for a difference in means, to more elaborate linear convolution models such as finite impulse response (FIR) models. The matrix X that contains the explanatory variables (e.g., designed effects or confounds) is called the design matrix. Each column of the design matrix corresponds to some effect one has built into the experiment or that may confound the results. These are referred to as explanatory variables, covariates, or regressors. The example in Figure 1 relates to an fMRI study of visual stimulation under four conditions. The effects on the response variable are modeled in terms of functions of the presence of these conditions (i.e., box or stick functions smoothed with components of a hemodynamic response function; in Figure 1, two components were used for each


The General Linear Model
Data
Design matrix
filter
filter design
time (seconds)
$Y = X\beta + \varepsilon$
$SY = SX\beta + S\varepsilon$
$\alpha \sim N\{0, \sigma^2\Sigma\}, \quad V = S\Sigma S^T$
$\hat{\beta} = (SX)^+ SY$
$\frac{c^T\hat{\beta}}{\sqrt{\hat{V}\{c^T\hat{\beta}\}}} \sim T_v$
$S = \begin{cases} \Sigma^{-\frac{1}{2}} & \textit{whitening} \\ 1 & \textit{none} \\ 1 - X_0X_0^+ & \textit{adjustment} \end{cases}$
Σ
intrinsic correlations and non-sphericity correction
$\hat{V}\{c^T\hat{\beta}\} = \hat{\sigma}^2 c^T (SX)^+ V (SX)^{+T} c$
$R = 1 - SX(SX)^+$
$\hat{\sigma}^2 = \frac{r^T r}{trace\{RV\}}$
$v = \frac{trace\{RV)^2}{trace\{RVRV\}}$
contrast images
SPM{T}

of the four conditions, giving eight regressors in total). The relative contribution of each of these columns to the response is controlled by the parameters β. These are estimated using standard least squares, and inferences about the parameter estimates are made using T or F statistics, depending upon whether one is looking at a particular linear combination (e.g., a subtraction), or all of them together.

The design matrix can contain covariates or indicator variables that take values of 0 or 1, to indicate the presence of a particular level of an experimental factor. Each column of X has an associated but unknown parameter. Some of these parameters will be of interest (e.g., the effect of a sensorimotor or cognitive condition or the regression coefficient of hemodynamic responses on reaction time). The remaining parameters will be of no interest and pertain to nuisance or confounding effects (e.g., the effect of being a particular subject or the regression slope of regional activity on global activity). Inferences about the parameter estimates are made using their estimated variance. This allows one to test the null hypothesis that some particular linear combination (e.g., a subtraction) of the estimates is zero using an SPM{T}. The T statistic obtains by dividing a contrast or compound (specified by contrast weights) of the parameter estimates by the standard error of that compound. Sometimes, several contrasts of parameter estimates are jointly interesting; for example, when using polynomial (Büchel et al. 1996) or basis function expansions of some experimental factor. In these instances, the SPM{F} is used and is specified with a matrix of contrast weights that can be thought of as a collection of "T contrasts" that one wants to test together.

In most analyses, the design matrix contains indicator variables or parametric variables encoding the experimental manipulations. These are formally identical to classical analysis of [co]variance (i.e., AnCova) models. An important instance of the GLM, from the perspective of fMRI, is the linear time invariant model. Mathematically this is no different from any other GLM. However, it explicitly

←

Figure 1 The general linear model. This model is an equation expressing the response variable Y in terms of a linear combination of explanatory variables in a design matrix X and an error term with assumed or known autocorrelation Σ. In fMRI, the data can be filtered with a convolution or residual-forming matrix (or a combination) S, leading to a generalized linear model that includes (intrinsic) serial correlations and applied (extrinsic) filtering. Different choices of S correspond to different estimation schemes as indicated on the *upper left*. The parameter estimates obtain in a least squares sense using the pseudoinverse (denoted by $+$) of the filtered design matrix. Generally, an effect of interest is specified by a vector of contrast weights c that give a weighted sum or compound of parameter estimates referred to as a contrast. The T statistic is simply this contrast divided by the estimated standard error (i.e., the square root of its estimated variance). The ensuing T statistic is distributed with v degrees of freedom. The equations for estimating the variance of the contrast and the degrees of freedom are provided in the *right-hand panel* and accommodate the nonsphericity implied by Σ.

treats the data sequence as an ordered time-series and enables a signal processing perspective that can be useful.

The equations summarized in Figure 1 can be used to implement a vast range of statistical analyses. The issue is therefore not so much the mathematics but the formulation of a design matrix *X* appropriate to the study design and inferences that are sought. Before considering general linear models as biophysical or causal models of brain responses, we focus on the design matrix as a device to specify experimental design. Here the explanatory variables encode treatment effects that we assume are expressed in a linear and instantaneous fashion in the data, without reference to any particular mechanism.

Experimental Design

This section considers the different sorts of designs employed in neuroimaging studies. Experimental designs can be classified as single factor or multifactorial designs; within this classification, the levels of each factor can be categorical or parametric.

CATEGORICAL DESIGNS, COGNITIVE SUBTRACTION, AND CONJUNCTIONS The tenet of cognitive subtraction is that the difference between two tasks can be formulated as a separable cognitive or sensorimotor component and that regionally specific differences in hemodynamic responses, evoked by the two tasks, identify the corresponding functionally specialized area. Early applications of subtraction range from the functional anatomy of word processing (Petersen et al. 1989) to functional specialization in extrastriate cortex (Lueck et al. 1989). The latter studies involved presenting visual stimuli with and without some sensory attribute (e.g., color and motion). The areas highlighted by subtraction were identified with homologous areas in monkeys that showed selective electrophysiological responses to equivalent visual stimuli.

Cognitive conjunctions (Price & Friston 1997) can be thought of as an extension of the subtraction technique, in the sense that they combine a series of subtractions. In subtraction, one tests a single hypothesis pertaining to the activation in one task relative to another. In conjunction analyses, several hypotheses are tested to determine whether all the activations, in a series of task pairs, are expressed conjointly. Consider the problem of identifying regionally specific activations due to a particular cognitive component (e.g., object recognition). If one can identify a series of task pairs whose differences have only that component in common, then the region that activates, in all the corresponding subtractions, can be associated with the common component. In short, conjunction analyses allow one to disclose context-invariant regional responses.

PARAMETRIC DESIGNS The premise behind parametric designs is that regional physiology will vary systematically with the degree of cognitive or sensorimotor processing or deficits thereof. Examples of this approach include the PET experiments of Grafton et al. (1992) that demonstrated significant correlations between hemodynamic responses and the performance of a visually guided motor tracking task. On the sensory side, Price et al. (1992) demonstrated a remarkable linear relationship between perfusion in periauditory regions and frequency of aural word presentation. This correlation was not observed in Wernicke's area, where perfusion appeared to correlate, not with the discriminative attributes of the stimulus, but with the presence or absence of semantic content. These relationships or neurometric functions may be linear or nonlinear. Using polynomial regression, in the context of the GLM, one can identify nonlinear relationships between stimulus parameters (e.g., stimulus duration or presentation rate) and evoked responses. To do this one usually uses a SPM{F} (see Büchel et al. 1996).

The example provided in Figure 2 illustrates both categorical and parametric aspects of design and analysis. These data were obtained from an fMRI study of visual motion processing using radially moving dots (Chawla et al. 1999). The stimuli were presented over a range of speeds using isoluminant and isochromatic stimuli. To identify areas involved in visual motion, a stationary dots condition was subtracted from a moving dots conditions (see the contrast weights, *upper right*). To ensure significant motion-sensitive responses, using color and luminance stimuli, a conjunction of the equivalent subtractions was assessed under both viewing contexts. Areas V5 and V3a are seen in the ensuing SPM{T}. The T values in this SPM are simply the minimum of the T values for each subtraction. Thresholding this SPM ensures that all voxels survive a threshold in each subtraction separately. This conjunction SPM has an equivalent interpretation; it represents the intersection of the excursion sets, defined by the threshold of each component SPM. This intersection is the essence of a conjunction.

The responses in left V5 are shown in the *lower panel* of Figure 2 and speak to a compelling inverted "U" relationship between speed and evoked response that peaks at around 8° per second. It is this sort of relationship that parametric designs try to characterize. Interestingly, the form of these speed-dependent responses was similar using both stimulus types, although luminance cues elicit a greater response. From the point of view of a factorial design, there is a main effect of stimulus (isoluminant versus isochromatic), a main (nonlinear) effect of speed, but no speed by stimulus interaction.

MULTIFACTORIAL DESIGNS Factorial designs are more prevalent than single-factor designs because they enable inferences about interactions. At its simplest, an interaction represents a change in a change. Interactions are associated with factorial designs where two or more factors are combined in the same experiment. The interaction term assesses the effect of one factor on the effect of the other. Factorial designs have a wide range of applications. An early application, in neuroimaging, examined physiological adaptation and plasticity during motor performance by assessing time by condition interactions (Friston et al. 1992). Factorial designs have an important role in the context of cognitive subtraction and additive factors logic by virtue of being able to test for interactions, or context-sensitive activations

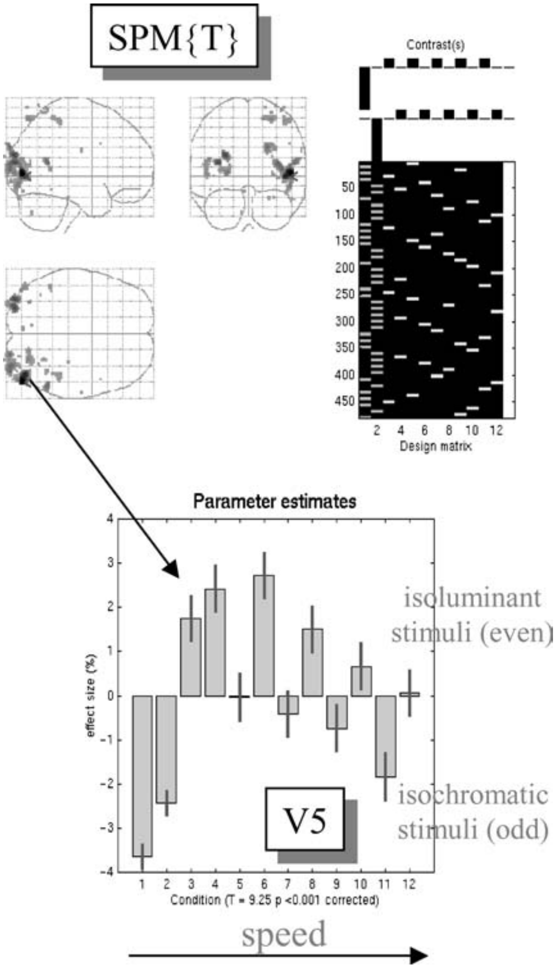


Figure 2 (*Top right*): An image representation of the design matrix. The vectors of contrast weights define the linear compounds of parameters tested. The contrast weights are displayed over the column of the design matrix that corresponds to the effects in question. The design matrix here includes condition-specific effects (boxcars convolved with a hemodynamic response function). Odd columns correspond to stimuli shown under isochromatic conditions and even columns model responses to isoluminant stimuli. The first two columns are for stationary stimuli and the remaining columns are for stimuli of increasing speed. The final column is a constant term. (*Top left*) A maximum intensity projection of the SPM{T} conforming to the standard anatomical space of Talairach & Tournoux (1988). The T values here are the minimum T values from both contrasts, thresholded at $p = 0.001$ uncorrected. The most significant conjunction is seen in left V5. (*Lower panel*) Plot of the condition-specific parameter estimates for this voxel. The T value was 9.25 ($p < 0.001$, corrected according to GRF theory). This example is based on data from Chawla et al. 1999.

(i.e., to demonstrate the fallacy of "pure insertion"; see Friston et al. 1996c). These interaction effects can sometimes be interpreted as (*a*) the integration of the two or more (cognitive) processes or (*b*) the modulation of one (perceptual) process by another.

In summary, the design matrix encodes the causes of observed data and, in particular, treatment effects caused by changes in the level of various experimental factors. These factors can have categorical or parametric levels and most experiments nowadays use multiple factors to test for both main effects and interactions. Before turning to mechanistically more informed formulations of the general linear model we consider briefly the two sorts of inferences that can be made about the parameter estimates.

Classical and Bayesian Inference

To date, inference in neuroimaging has been restricted largely to classical inferences based upon statistical parametric maps. The statistics that comprise these SPMs are essentially functions of the data. The probability distribution of the chosen statistic, under the null hypothesis (i.e., the null distribution), is used to compute a p value. This p value is the probability of obtaining the statistic, or the data, given that the null hypothesis is true. If sufficiently small, the null hypothesis can be rejected and an inference is made. The alternative approach is to use Bayesian or conditional inference based upon the posterior distribution of the activation given the data (Holmes & Ford 1993). This necessitates the specification of priors (i.e., the probability distribution of the activation). Bayesian inference requires the posterior distribution and therefore rests upon a posterior density analysis. A useful way to summarize this posterior density is to compute the probability that the activation exceeds some threshold. This computation represents a Bayesian inference about the effect, in relation to the specified threshold. By computing a posterior probability for each voxel, we can construct posterior probability maps or PPMs that are a useful complement to classical SPMs (Friston et al. 2002, Friston & Penny 2003).

The motivation for using conditional or Bayesian inference is that it has high face validity. This is because the inference is about an effect, or activation, being greater than some specified size that has some meaning in relation to underlying neurophysiology. This contrasts with classical inference, in which the inference is about the effect being significantly different from zero. The problem for classical inference is that trivial departures from the null hypothesis can be declared significant, with sufficient data or sensitivity. From the point of view of neuroimaging, posterior inference is especially useful because it eschews the multiple-comparisons problem. In classical inference, one tries to ensure that the probability of rejecting the null hypothesis incorrectly is maintained at a small rate, despite making inferences over large volumes of the brain. This induces a multiple-comparisons problem that, for continuous spatially extended data, requires an adjustment or correction to the p value using GRF theory as mentioned above. This correction means that classical inference becomes less sensitive or powerful with large search volumes. In contradistinction, posterior inference does not have to contend with the multiple-comparisons problem because there are no false-positives. The probability that activation has occurred, given the data, at any particular voxel is the same, irrespective of whether one has analyzed that voxel or the entire brain. For this reason, posterior inference using PPMs represents a relatively more powerful approach than classical inference in neuroimaging.

HIERARCHICAL MODELS AND EMPIRICAL BAYES PPMs require the posterior distribution or conditional distribution of the activation (a contrast of conditional parameter estimates) given the data. This posterior density can be computed, under Gaussian assumptions, using Bayes' rule. Bayes' rule requires the specification of a likelihood function and the prior density of the model's parameters. The models used to form PPMs and the likelihood functions are the same as in classical SPM analyses, namely the GLM. The only extra bit of information that is required is the prior probability distribution of the parameters. Although it would be possible to specify them using independent data or some plausible physiological constraints, there is an alternative to this fully Bayesian approach. The alternative is empirical Bayes, in which the variances of the prior distributions are estimated directly from the data. Empirical Bayes requires a hierarchical observation model where the parameters and hyperparameters at any particular level can be treated as priors on the level below. There are numerous examples of hierarchical observation models in neuroimaging. For example, the distinction between fixed- and random/mixed-effects analyses of multisubject studies relies upon a two-level hierarchical model (Friston et al. 2002). However, in neuroimaging there is a natural hierarchical observation model that is common to all brain mapping experiments. This is the hierarchy induced by looking for the same effects at every voxel within the brain (or gray matter). The first level of the hierarchy corresponds to the experimental effects at any particular voxel and the second level of the hierarchy comprises the effects over voxels. Put simply, the variation in a particular contrast, over voxels, can be used as the prior variance of that contrast at any particular voxel. Generally, hierarchical linear models have the following form:

$$\begin{aligned} y &= X^{(1)}\beta^{(1)} + \varepsilon^{(1)} \\ \beta^{(1)} &= X^{(2)}\beta^{(2)} + \varepsilon^{(2)}. \\ \beta^{(2)} &= \ldots \end{aligned} \tag{2}$$

This is the same as Equation 1, but now the parameters of the first level are generated by a supraordinate linear model and so on to any hierarchical depth required. These hierarchical observation models are an important extension of the GLM and are usually estimated using Expectation Maximization (Dempster et al. 1977). In the present context, the response variables comprise the responses at all voxels and $\beta^{(1)}$ are the treatment effects about which we want to make an inference. Because we have invoked a second level, the first-level parameters embody random effects and are generated by a second-level linear model. At the second level, $\beta^{(2)}$ is the average effect over voxels and $\varepsilon^{(2)}$ is voxel-to-voxel variation. By estimating the variance of $\varepsilon^{(2)}$, one is implicitly estimating an empirical prior on the first-level parameters at each voxel. This prior can then be used to estimate the posterior probability of $\beta^{(1)}$ being greater than some threshold at each voxel. An example of the ensuing PPM is provided in Figure 3 along with the classical SPM.

In this section we have seen how the GLM can be used to test hypotheses about brain responses and how, in a hierarchical form, it enables empirical Bayesian or conditional inference. In the next section, we deal with dynamic systems and how they can be formulated as GLMs. These dynamic models take us closer to how experimental manipulations actually cause brain responses, and represent the next step in working toward causal models of brain responses.

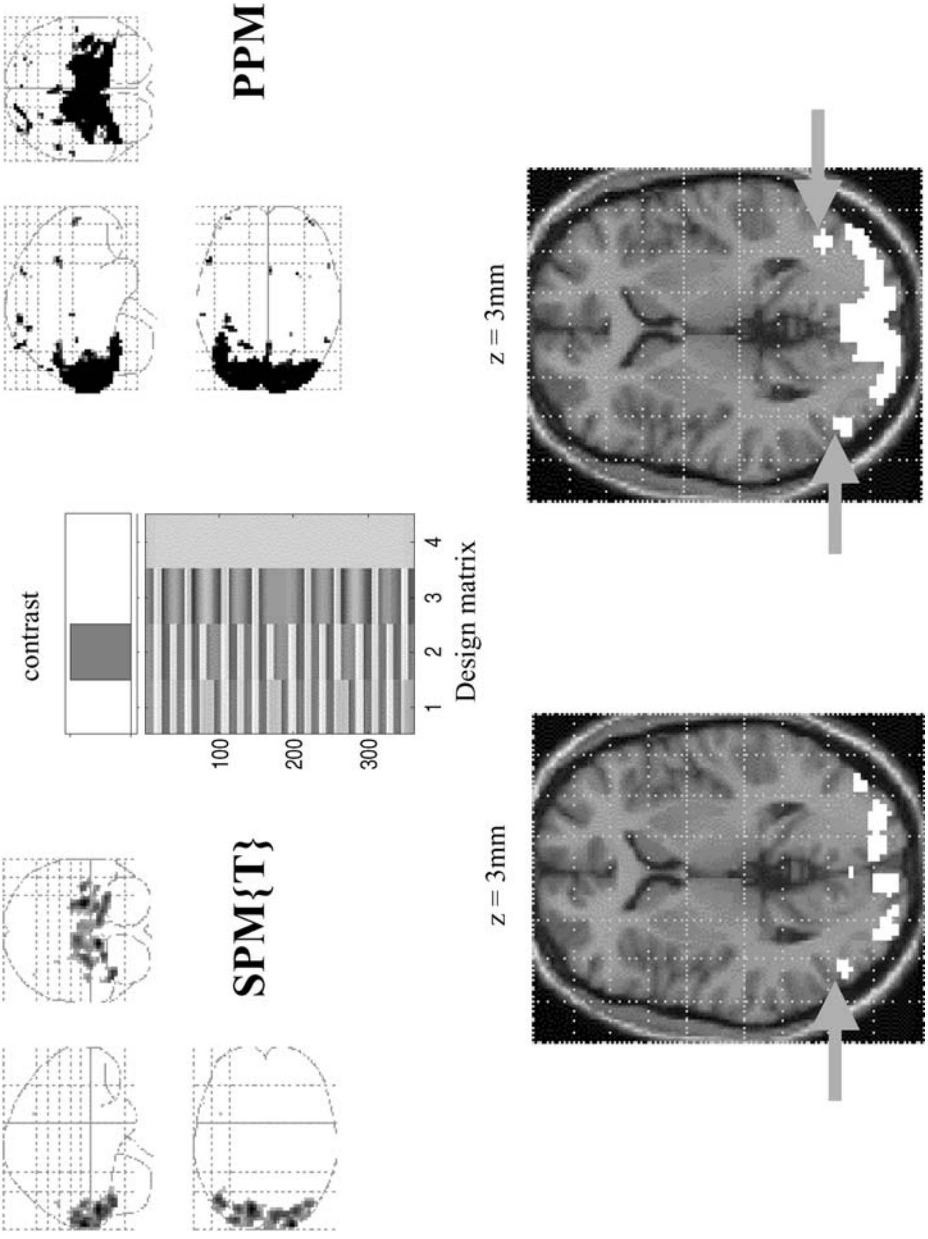
Dynamic Models

CONVOLUTION MODELS AND TEMPORAL BASIS FUNCTIONS In Friston et al. (1994), the form of the hemodynamic impulse response function (HRF) was estimated using a least squares deconvolution and a time-invariant model, where evoked neuronal responses are convolved or smoothed with an HRF to give the measured hemodynamic response (see also Boynton et al. 1996). This simple linear convolution model is the cornerstone for making statistical inferences about activations in fMRI with the GLM. An impulse response function is the response to a single impulse, measured at a series of times after the input. It characterizes the input-output behavior of the system (i.e., voxel) and places important constraints on the sorts of inputs that will excite a response.

Knowing the forms that the HRF can take is important for several reasons, not least because it allows for better statistical models of the data. The HRF may vary from voxel to voxel and this has to be accommodated in the GLM. To allow for different HRFs in different brain regions, the notion of temporal basis functions, to model evoked responses in fMRI, was introduced (Friston et al. 1995a) and applied to event-related responses in Josephs et al. (1997) (see also Lange & Zeger 1997). The basic idea behind temporal basis functions is that the hemodynamic response, induced by any given trial type, can be expressed as the linear combination of several (basis) functions of peristimulus time. The convolution model for fMRI responses takes a stimulus function encoding the neuronal responses and convolves it with an HRF to give a regressor that enters the design matrix. When using basis functions, the stimulus function is convolved with all the basis functions to give a series of regressors (in Figure 1 we used four stimulus functions and two basis functions to give eight regressors). Mathematically we can express this model as

$$\begin{matrix} y(t) = X\beta + \varepsilon \\ X_i = T_i(t) \otimes u(t) \end{matrix} \quad \Leftrightarrow \quad \begin{matrix} y(t) = u(t) \otimes h(t) + \varepsilon \\ h(t) = \beta_1 T_1(t) + \beta_2 T_2(t) + \ldots \end{matrix}, \tag{3}$$

where $\otimes$ means convolution. This equivalence illustrates how temporal basis functions allow one to take any convolution model (right) and convert it into a GLM (left). The parameter estimates β_i are the coefficients or weights that determine the


PPM
SPM{T}
contrast
Design matrix
1
2
3
4
100
200
300
z = 3mm
z = 3mm

mixture of basis functions $T_i(t)$ that best models $h(t)$, the HRF for the trial type and voxel in question. We find the most useful basis set to be a canonical HRF and its derivatives with respect to the key parameters that determine its form (see below). Temporal basis functions are important because they enable a graceful transition between conventional multilinear regression models with one stimulus function per condition and FIR models with a parameter for each time point following the onset of a condition or trial type. Figure 4 illustrates this graphically. In short, temporal basis functions offer useful constraints on the form of the estimated response that retain the flexibility of FIR models and the efficiency of single regressor models.

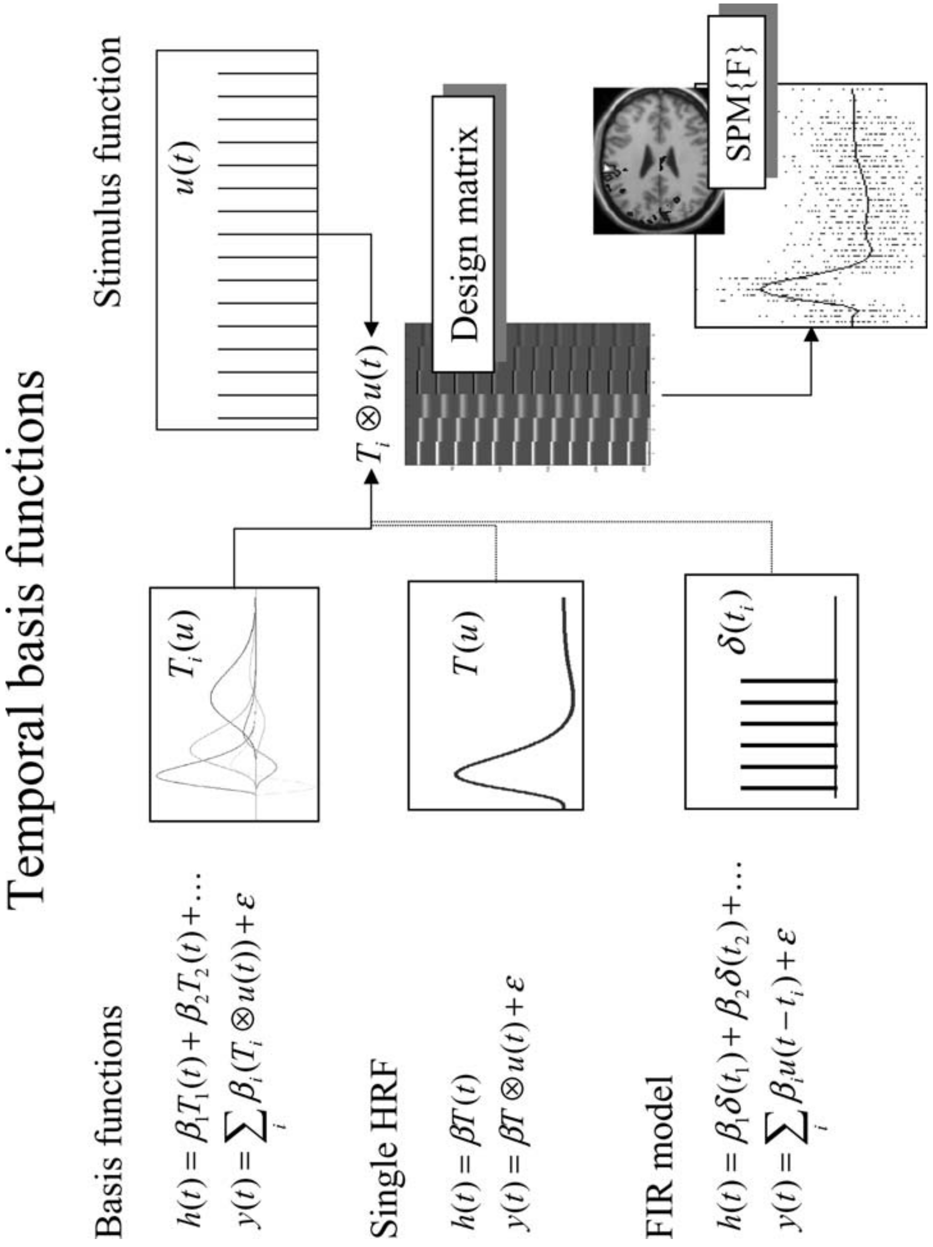
Biophysical Models

INPUT-STATE-OUTPUT SYSTEMS By adopting a convolution model for brain responses in fMRI, we are implicitly positing some underlying dynamic system that converts neuronal responses into observed hemodynamic responses. Our understanding of the biophysical and physiological mechanisms that underpin the HRF has grown considerably in the past few years (e.g., Buxton & Frank 1997, Hoge et al. 1999, Mandeville et al. 1999). Figure 5 shows some simulations based on the hemodynamic model described in Friston et al. (2000). Here, neuronal activity induces some autoregulated vasoactive signal that causes transient increases in regional cerebral blood flow (rCBF). The resulting flow increases lead to the dilation of a venous balloon, increasing its volume (v) and diluting venous blood to decrease deoxyhemoglobin content (q). The blood oxygenation level–dependent (BOLD) signal is roughly proportional to the concentration of deoxyhemoglobin (q/v) and follows the rCBF response with about a one-second delay. The model is framed in terms of differential equations, examples of which are provided in Figure 5.

Notice that we have introduced variables, like volume and deoxyhemoglobin concentrations, that are not actually observed. These are referred to as the hidden states of input-state-output models. The state and output equations of any analytic dynamical system are

←

Figure 3 Statistical parametric map (SPM) and posterior probability map (PPM) for a functional magnetic resonance imaging study of attention to visual motion (Büchel & Friston 1997). The display format in the *lower panel* uses an axial slice through extrastriate regions but the thresholds are the same as employed in maximum-intensity projections (*upper panels*). (*Upper right*) The activation threshold for the PPM was 0.7 a.u., meaning that all voxels shown had a 90% chance of an activation of 0.7% or more. (*Upper left*) The corresponding SPM using a corrected threshold at $p = 0.05$. Note the bilateral foci of motion-related responses in the PPM that are not seen in the SPM (*gray arrows*). As can be imputed from the design matrix (*upper middle panel*), the statistical model of evoked responses comprised boxcar regressors convolved with a canonical hemodynamic response function. The middle column corresponds to the presentation of moving dots and was the stimulus property tested by the contrast.


Temporal basis functions
Basis functions
$h(t) = \beta_1 T_1(t) + \beta_2 T_2(t) + \ldots$
$y(t) = \sum_i \beta_i (T_i \otimes u(t)) + \varepsilon$
Single HRF
$h(t) = \beta T(t)$
$y(t) = \beta T \otimes u(t) + \varepsilon$
FIR model
$h(t) = \beta_1 \delta(t_1) + \beta_2 \delta(t_2) + \ldots$
$y(t) = \sum_i \beta_i u(t - t_i) + \varepsilon$
$T_i(u)$
$T(u)$
$\delta(t_i)$
Stimulus function
$u(t)$
$T_i \otimes u(t)$
Design matrix
SPM{F}

$$\begin{aligned}\dot{x}(t) &= f(x, u, \theta)\\ y(t) &= g(x, u, \theta) + \varepsilon.\end{aligned} \tag{4}$$

The first line is an ordinary differential equation and expresses the rate of change of the states as a parameterized function of the states and input. Typically, the inputs $u(t)$ correspond to designed experimental effects (e.g., the stimulus function in fMRI). There is a fundamental and causal relationship (Fliess et al. 1983) between the outputs and the history of the inputs in Equation 4. This relationship conforms to a Volterra series, which expresses the output $y(t)$ as a generalized convolution of the input $u(t)$, critically without reference to the hidden states $x(t)$. This series is simply a functional Taylor expansion of the outputs with respect to the inputs (Bendat 1990). The reason it is a functional expansion is that the inputs are a function of time[1],

$$y(t) = \sum_i \int_0^t \ldots \int_0^t \kappa_i(\sigma_1, \ldots, \sigma_i) u(t - \sigma_1), \ldots, u(t - \sigma_i) d\sigma_1, \ldots, d\sigma_i$$

$$\kappa_i(\sigma_1, \ldots, \sigma_i) = \frac{\partial^i y(t)}{\partial u(t - \sigma_1), \ldots, \partial u(t - \sigma_i)}, \tag{5}$$

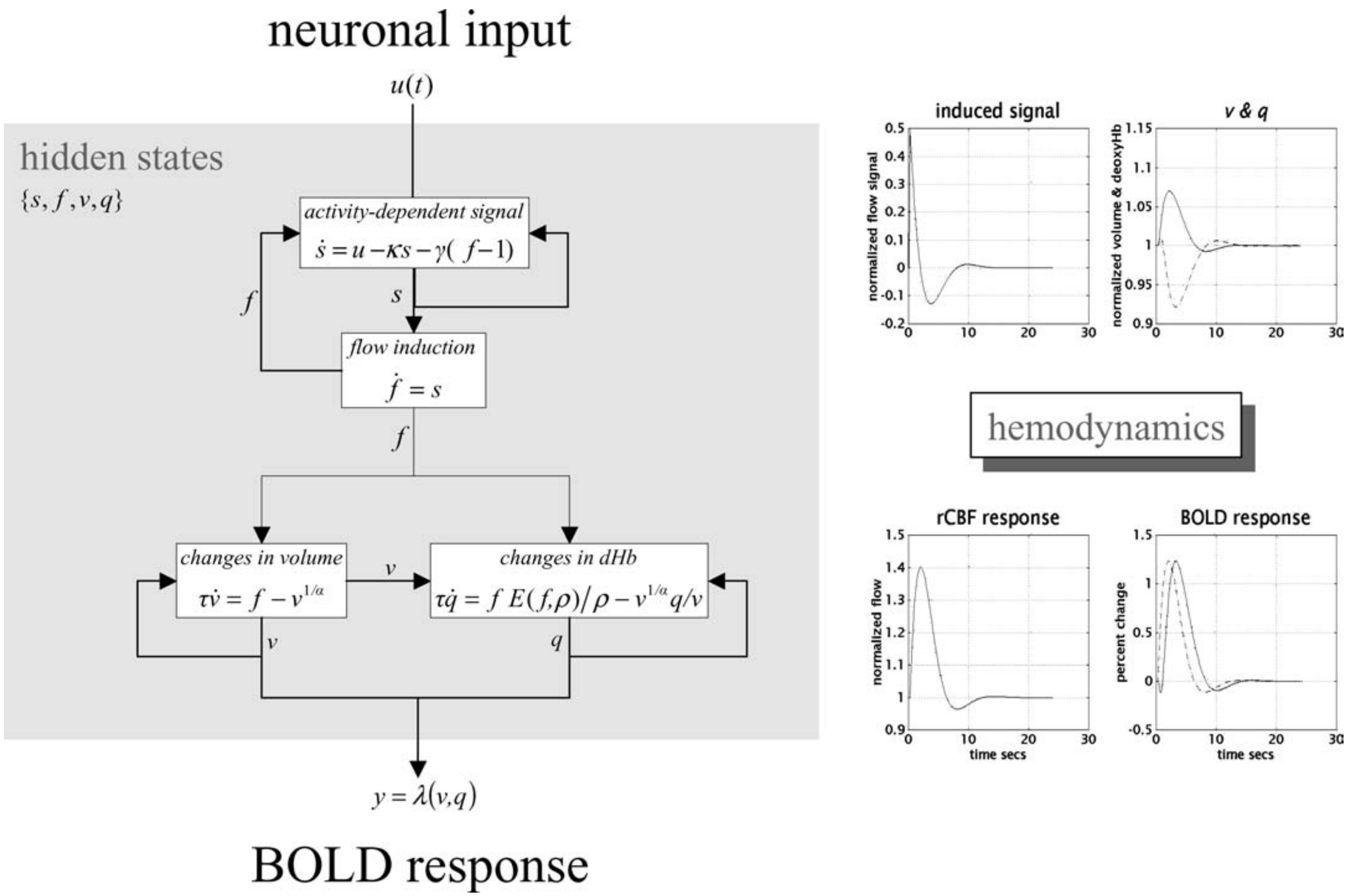
where $\kappa_i(\sigma_1, \ldots, \sigma_i)$ is the ith-order kernel. In Equation 5, the integrals are restricted to the past. This renders Equation 5 causal. The key thing here is that Equation 5 is simply a convolution and can be expressed as a GLM as in Equation 3. This means that we can take a neurophysiologically realistic model of hemodynamic responses and use it as an observation model to estimate parameters using observed data. Here the model is parameterized in terms of kernels that have a direct analytic

---

[1] For simplicity, here and in Equation 7, we deal with only one experimental input.

←

Figure 4 Temporal basis functions offer useful constraints on the form of the estimated response that retain (*a*) the flexibility of finite impulse response (FIR) models and (*b*) the efficiency of single regressor models. The specification of these constrained FIR models involves setting up stimulus functions $u(t)$ that model expected neuronal changes [e.g., boxcars of epoch-related responses or spikes (delta functions) at the onset of specific events or trials]. These stimulus functions are then convolved with a set of basis functions $T_i(t)$ of peristimulus time that, in some linear combination, model the hemodynamic impulse response function (HRF). The ensuing regressors are assembled into the design matrix. The basis functions can be as simple as a single canonical HRF (*middle*), through to a series of delayed delta functions (*bottom*). The latter case corresponds to an FIR model and the coefficients constitute estimates of the impulse response function at a finite number of discrete sampling times. Selective averaging in event-related functional magnetic resonance imaging (Dale & Buckner 1997) is mathematically equivalent to this limiting case.


neuronal input
$u(t)$
hidden states
$\{s, f, v, q\}$
activity-dependent signal
$\dot{s} = u - \kappa s - \gamma(f-1)$
s
f
flow induction
$\dot{f} = s$
f
changes in volume
$\tau\dot{v} = f - v^{1/\alpha}$
v
changes in dHb
$\tau\dot{q} = f\,E(f,\rho)/\rho - v^{1/\alpha} q/v$
v
q
$y = \lambda(v,q)$
BOLD response
induced signal
normalized flow signal
v & q
normalized volume & deoxyHb
hemodynamics
rCBF response
normalized flow
time secs
BOLD response
percent change
time secs

relation to the original parameters θ of the biophysical system (through Equation 5). The first-order kernel is simply the conventional HRF. High-order kernels correspond to high-order HRFs and can be estimated using basis functions as described above. In fact, by choosing basis functions according to

$$T(\sigma)_i = \frac{\partial \kappa(\sigma)_1}{\partial \theta_i}, \tag{6}$$

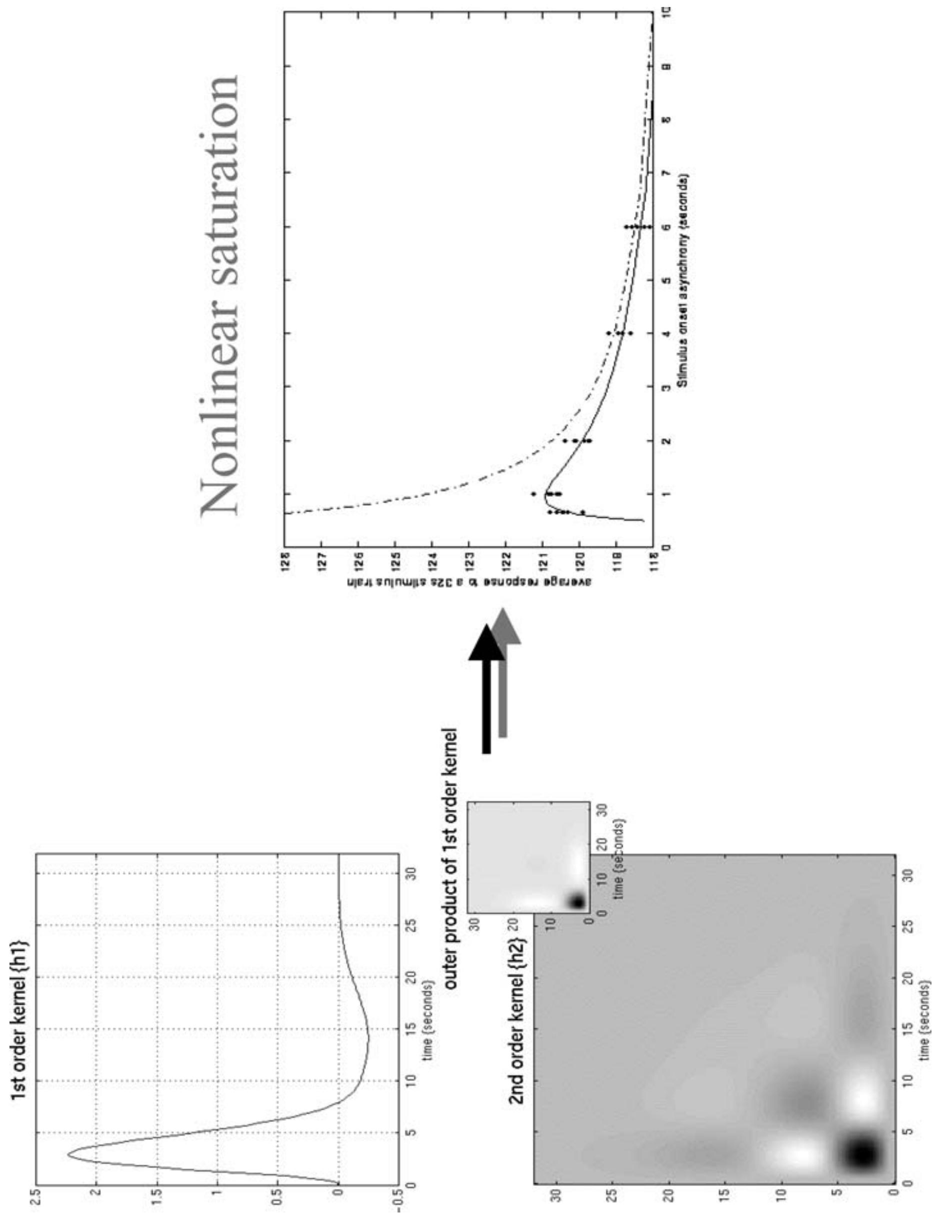
one can estimate the biophysical parameters because, to a first-order approximation, $\beta_i = \theta_i$. The critical step we have taken here is to start with a causal dynamic model of how responses are generated and construct a general linear observation model that allows us to estimate and infer things about the parameters of that model. This is in contrast to the conventional use of the GLM with design matrices that are not informed by a forward model of how data are caused. This approach to modeling brain responses has a much more direct connection with underlying physiology and rests upon an understanding of the underlying system.

NONLINEAR SYSTEM IDENTIFICATION Once a suitable causal model has been established (e.g., Figure 5), we can estimate second-order kernels. These kernels represent a nonlinear characterization of the HRF that can model interactions among stimuli in causing responses. One important manifestation of the nonlinear effects, captured by the second-order kernels, is a modulation of stimulus-specific responses by preceding stimuli that are proximate in time. This means that responses at high stimulus presentation rates saturate and, in some instances, show an inverted U behavior. This behavior appears to be specific to BOLD effects (as distinct from evoked changes in cerebral blood flow) and may represent a hemodynamic refractoriness. This effect has important implications for event-related fMRI, where one may want to present trials in quick succession.

The results of a typical nonlinear analysis are given in Figure 6. The results in the *right panel* represent the average response, integrated over a 32-second train of stimuli as a function of stimulus onset asynchrony (SOA). These responses are based on the kernel estimates (*left panels*) using data from a voxel in the left posterior temporal region of a subject obtained during the presentation of single words at different rates. The solid line represents the estimated response and shows

←

Figure 5 (*Right*) Hemodynamics elicited by an impulse of neuronal activity as predicted by a dynamical biophysical model (*left*). A burst of neuronal activity causes an increase in flow-inducing signal that decays with first-order kinetics and is downregulated by local flow. This signal increases regional cerebral blood flow (rCBF), which dilates the venous capillaries, increasing volume (v). Concurrently, venous blood is expelled from the venous pool, decreasing deoxyhemoglobin content (q). The resulting fall in deoxyhemoglobin concentration leads to a transient increase in the blood oxygenation level–dependent (BOLD) signal and a subsequent undershoot. (*Left*) Hemodynamic model on which these simulations were based (see Friston et al. 2000 for details).


1st order kernel {h1}
time (seconds)
2nd order kernel {h2}
time (seconds)
outer product of 1st order kernel
time (seconds)
Nonlinear saturation
average response to a 32s stimulus train
Stimulus onset asynchrony (seconds)

a clear maximum at just less than one second. The dots are responses based on empirical data from the same experiment. The broken line shows the expected response in the absence of nonlinear effects (i.e., that predicted by setting the second-order kernel to zero). It is clear that nonlinearities become important at around two seconds, leading to an actual diminution of the integrated response at subsecond SOAs. The implication of this sort of result is that the assumptions of the linear convolution models discussed above are violated with subsecond SOAs (see also Buckner et al. 1996, Burock et al. 1998).

In summary, we started with models of regionally specific responses, framed in terms of the general linear model, in which responses were modeled as linear mixtures of designed changes in explanatory variables. Hierarchical extensions to linear observation models enable random-effects analyses and, in particular, an empirical Bayesian approach. The mechanistic utility of these models is realized through forward models that embody causal dynamics. Simple variants of these are the linear convolution models used to construct explanatory variables in conventional analyses of fMRI data. These are a special case of generalized convolution models that are mathematically equivalent to input-state-output systems comprising hidden states. Estimation and inference with these dynamic models tells us something about how the response was caused, but only at the level of single voxels. In the next section, we adopt the same perspective on models, but in the context of distributed responses and functional integration.

MODELS OF FUNCTIONAL INTEGRATION

Functional and Effective Connectivity

Imaging neuroscience has firmly established functional specialization as a principle of brain organization in man. The integration of specialized areas has proven

---

Figure 6 (*Left panels*) Volterra kernels from a voxel in the left superior temporal gyrus at −56, −28, and 12 mm. These kernel estimates were based on a single-subject study of aural word presentation at different rates (from 0 to 90 words per minute) using a second-order approximation to a Volterra series expansion modeling the observed hemodynamic response to stimulus input (a delta function for each word). These kernels can be thought of as a characterization of the second-order hemodynamic response function. The first-order kernel κ_1 (*upper panel*) represents the (first-order) component usually presented in linear analyses. The second-order kernel (*lower panel*) is presented in image format. The color scale is arbitrary; white is positive and black is negative. The insert on the right represents $\kappa_1\kappa_1^T$, the second-order kernel predicted by a simple model that involves a linear convolution with κ_1 followed by some static nonlinearity. (*Right panel*) Integrated responses over a 32-second stimulus train as a function of stimulus onset asynchrony. Solid line: estimates based on the nonlinear convolution model parameterized by the kernels on the left. Broken line: the responses expected in the absence of second-order effects (i.e., in a truly linear system). Dots: empirical averages based on the presentation of actual stimulus trains.

more difficult to assess. Functional integration is usually inferred on the basis of correlations among measurements of neuronal activity. Functional connectivity has been defined as statistical dependencies or correlations among remote neurophysiological events. However, correlations can arise in a variety of ways; for example, in multiunit electrode recordings they can result from stimulus-locked transients evoked by a common input or reflect stimulus-induced oscillations mediated by synaptic connections (Gerstein & Perkel 1969). Integration within a distributed system is usually better understood in terms of effective connectivity: Effective connectivity refers explicitly to the influence that one neural system exerts over another, either at a synaptic (i.e., synaptic efficacy) or population level. It has been proposed that "the [electrophysiological] notion of effective connectivity should be understood as the experiment- and time-dependent, simplest possible circuit diagram that would replicate the observed timing relationships between the recorded neurons" (Aertsen & Preißl 1991). This speaks to two important points: (*a*) Effective connectivity is dynamic, i.e., activity- and time-dependent and (*b*) it depends upon a model of the interactions. The estimation procedures employed in functional neuroimaging can be divided into those based on (*a*) linear regression models (e.g., McIntosh & Gonzalez-Lima 1994) or (*b*) nonlinear dynamic causal models.

There is a necessary link between functional integration and multivariate analyses because the latter are necessary to model interactions among brain regions. Multivariate approaches can be divided into those that are inferential in nature and those that are data-led or exploratory. We first consider multivariate approaches that generally are based on functional connectivity or covariance patterns (and generally are exploratory) and then turn to models of effective connectivity (that allow for some form of inference).

EIGENIMAGE ANALYSIS AND RELATED APPROACHES In Friston et al. (1993), we introduced voxel-based principal component analysis (PCA) of neuroimaging time-series to characterize distributed brain systems implicated in sensorimotor, perceptual, or cognitive processes. These distributed systems are identified with principal components or eigenimages that correspond to spatial modes of coherent brain activity. This approach represents one of the simplest multivariate characterizations of functional neuroimaging time-series and falls into the class of exploratory analyses. Principal component or eigenimage analysis generally uses singular value decomposition to identify a set of orthogonal spatial modes that capture the greatest amount of variance expressed over time. As such, the ensuing modes embody the most prominent aspects of the variance-covariance structure of a given time-series. Noting that covariance among brain regions is equivalent to functional connectivity renders eigenimage analysis particularly interesting because it was among the first ways of addressing functional integration (i.e., connectivity) with neuroimaging data. Subsequently, eigenimage analysis has been elaborated in a number of ways. Notable among these is canonical variate analysis and multidimensional scaling (Friston et al. 1996a,b). Canonical variate analysis was introduced in the context of ManCova (multiple analysis of covariance) and uses the generalized eigenvector solution to maximize the variance that can be explained by some explanatory variables relative to error. Canonical variate analysis can be thought of as an extension of eigenimage analysis that refers explicitly to some explanatory variables and allows for statistical inference.

In fMRI, eigenimage analysis (e.g., Sychra et al. 1994) generally is used as an exploratory device to characterize coherent brain activity. These variance components may or may not be related to experimental design and endogenous coherent dynamics that have been observed in the motor system (Biswal et al. 1995). Despite its exploratory power, eigenimage analysis is fundamentally limited for two reasons. First, it offers only a linear decomposition of any set of neurophysiological measurements, and second, the particular set of eigenimages or spatial modes obtained is uniquely determined by constraints that are biologically implausible. These aspects of PCA confer inherent limitations on the interpretability and usefulness of eigenimage analysis of biological time-series and have motivated the exploration of nonlinear PCA and neural network approaches (e.g., Mørch et al. 1995).

Two other important approaches deserve mention here. The first is independent component analysis (ICA). ICA uses entropy maximization to find, using iterative schemes, spatial modes or their dynamics that are approximately independent. This is a stronger requirement than orthogonality in PCA and involves removing high-order correlations among the modes (or dynamics). It was initially introduced as spatial ICA (McKeown et al. 1998) in which the independence constraint was applied to the modes (with no constraints on their temporal expression). Approaches that are more recent use, by analogy with magneto- and electrophysiological time-series analysis, temporal ICA where the dynamics are enforced to be independent (e.g., Calhoun et al. 2001). This requires an initial dimension reduction (usually using conventional eigenimage analysis). Finally, there has been an interest in cluster analysis (Baumgartner et al. 1997). Conceptually, this can be related to eigenimage analysis through multidimensional scaling and principal coordinate analysis.

All these approaches are interesting, but hardly anyone uses them. This is largely because the approaches tell you nothing about how the brain works and don't allow one to ask specific questions. Simply demonstrating statistical dependencies among regional brain responses (i.e., demonstrating functional connectivity) does not address how these responses were caused. To address this, one needs explicit models of integration, or more precisely, effective connectivity.

Dynamic Causal Modeling

This final section examines the modeling of interactions among neuronal populations, at a cortical level, using neuroimaging time-series. The aim of these models is to estimate, and make inferences about, the coupling among brain areas and how that coupling is influenced by changes in experimental context (e.g., time or cognitive set). The basic idea is to construct a reasonably realistic neuronal model of interacting cortical regions or nodes. This model is then supplemented with a forward model of how neuronal or synaptic activity translates into a measured response (see previous section). This enables the parameters of the neuronal model (i.e., effective connectivity) to be estimated from observed data.

Intuitively, this approach regards an experiment as a designed perturbation of neuronal dynamics that are promulgated and distributed throughout a system of coupled anatomical nodes to change region-specific neuronal activity. These changes engender, through a measurement-specific forward model, responses that are used to identify the architecture and time constants of the system at a neuronal level. This represents a departure from conventional approaches (e.g., structural equation modeling and autoregression models; Büchel & Friston 1997, Harrison et al. 2003, McIntosh & Gonzalez-Lima 1994), in which one assumes the observed responses are driven by endogenous or intrinsic noise (i.e., innovations). In contradistinction, dynamic causal models assume the responses are driven by designed changes in inputs. An important conceptual aspect of dynamic causal models pertains to how the experimental inputs enter the model and cause neuronal responses. Experimental variables can illicit responses in one of two ways. First, they can elicit responses through direct influences on specific anatomical nodes. This would be appropriate, for example, in modeling sensory evoked responses in early visual cortices. The second class of input exerts its effect vicariously, through a modulation of the coupling among nodes. These sorts of experimental variables would normally be more enduring; for example, attention to a particular attribute or the maintenance of some cognitive set. These distinctions are seen most clearly in relation to particular forms of causal models used for estimation; for example, the bilinear approximation

$$\begin{aligned}\dot{x}(t) &= f(x,u)\\ &= Ax + uBx + Cu\\ y &= g(x) + \varepsilon\\ A &= \frac{\partial f}{\partial x} \quad B = \frac{\partial^2 f}{\partial x \partial u} \quad C = \frac{\partial f}{\partial u}.\end{aligned} \tag{7}$$

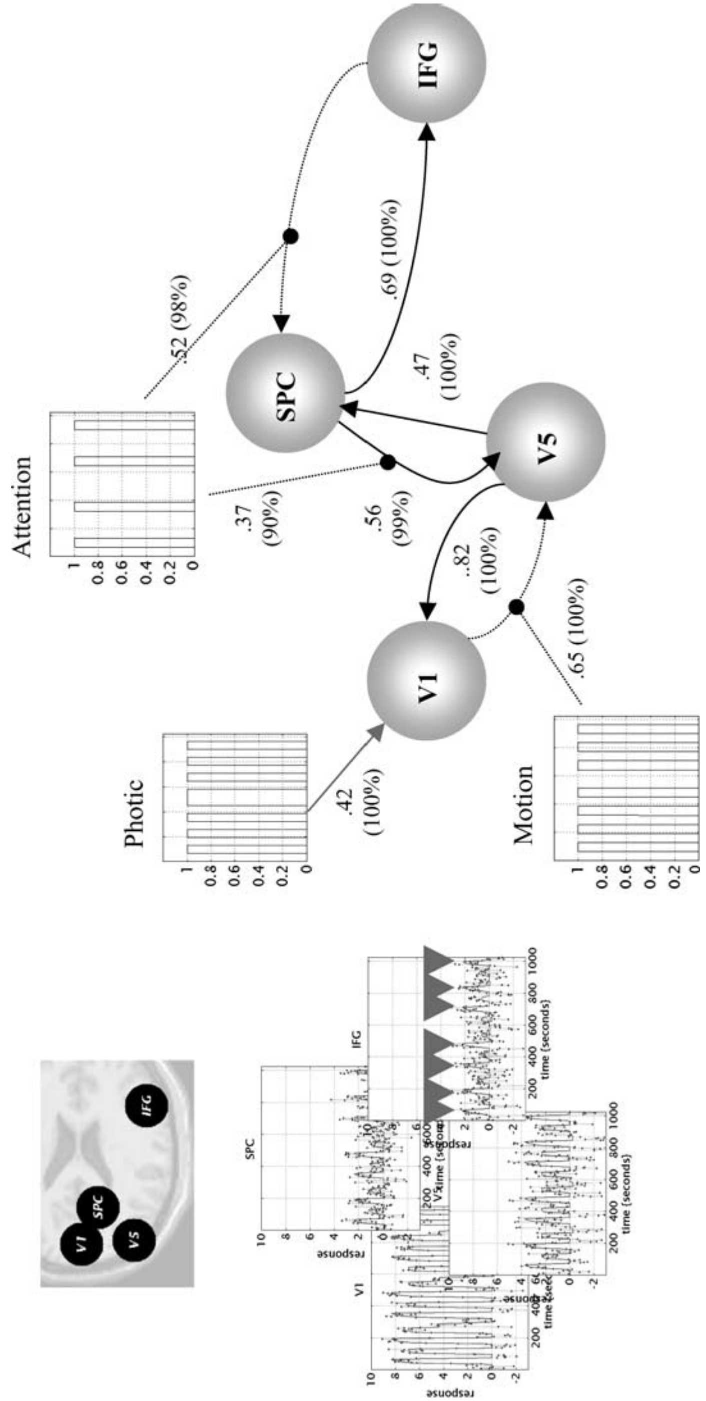
This is an approximation to any model of how changes in neuronal activity in one region, $x(t)_i$, are caused by activity in the other regions. Here the output function $g(x)$ embodies a hemodynamic model, linking neuronal activity to BOLD, for each region (e.g., that in Figure 5). The matrix A represents the connectivity among the regions in the absence of input $u(t)$. Effective connectivity is the influence that one neuronal system exerts over another in terms of inducing a response $\partial\dot{x}/\partial x$. This latent connectivity can be thought of as the intrinsic coupling in the absence of experimental perturbations. The matrix B is effectively the change in intrinsic coupling induced by the input. It encodes the input-sensitive changes in A or, equivalently, the modulation of effective connectivity by experimental manipulations. Because B is a second-order derivative, it is referred to as bilinear.

Finally, the matrix C embodies the extrinsic influences of inputs on neuronal activity. The parameters $\theta = \{A, B, C\}$ are the connectivity or coupling matrices that we wish to identify and that define the functional architecture and interactions among brain regions at a neuronal level.

Because Equation 7 has exactly the same form as Equation 4, we can express it as a GLM and estimate the parameters. Generally, estimation in the context of highly parameterized models like DCMs requires constraints in the form of priors. These priors enable conditional inference about the connectivity estimates. The sorts of questions that can be addressed with DCMs are now illustrated by looking at how attentional modulation might be mediated in sensory processing hierarchies in the brain.

DCM AND ATTENTIONAL MODULATION It has been established that the superior posterior parietal cortex (SPC) exerts a modulatory role on V5 responses using Volterra-based regression models (Friston & Büchel 2000), and that the inferior frontal gyrus (IFG) exerts a similar influence on SPC using structural equation modeling (Büchel & Friston 1997). The example here shows that DCM leads to the same conclusions but starting from a completely different construct (see Friston et al. 2003 for details). The experimental paradigm and data acquisition parameters are described in the Figure 7 legend. Figure 7 also shows the location of the regions that entered into the DCM. These regions were based on maxima from conventional SPMs testing for the effects of photic stimulation, motion, and attention. Regional time courses were taken as the first eigenvariate of 8 mm spherical volumes of interest centered on the maxima shown in the figure. The inputs, in this example, comprise one sensory perturbation and two contextual inputs. The sensory input was simply the presence of photic stimulation, and the first contextual input was presence of motion in the visual field. The second contextual input, encoding attentional set, was unity during attention to speed changes and zero otherwise. The model parameters were fitted such that the regional outputs from the model corresponded as closely as possible to the four regional eigenvariates (Figure 7, *left panel*). The intrinsic connections were constrained to conform to a hierarchical pattern in which each area was reciprocally connected to its supraordinate area. Photic stimulation entered at, and only at, V1. The effect of motion in the visual field was modeled as a bilinear modulation of the V1 to V5 connectivity and attention was allowed to modulate the backward connections from IFG and SPC.

The results of the DCM are shown in Figure 7 (*right panel*). Of primary interest is the modulatory effect of attention that is expressed in terms of the bilinear coupling parameters for this input. Analysis of the posterior densities of the bilinear parameters shows that we can be highly confident that attention modulates the backward connections from IFG to SPC and from SPC to V5. Indeed, the influences of IFG on SPC are negligible in the absence of attention (dotted connection). It is important to note that the only way attentional manipulation could effect brain responses was through this bilinear effect. Attention-related responses are seen throughout the system (attention epochs are marked with arrows in the plot of IFG


Attention
Photic
Motion
1
0.8
0.6
0.4
0.2
0
V1
V5
SPC
IFG
.42 (100%)
.82 (100%)
.65 (100%)
.56 (99%)
.37 (90%)
.47 (100%)
.69 (100%)
.52 (98%)
response
time (seconds)
200 400 600 800 1000
10 8 6 4 2 0 -2

responses in Figure 7). This attentional modulation is accounted for, sufficiently, by changing just two connections. This change is, presumably, instantiated by an instructional set at the beginning of each epoch.

The second idea this analysis illustrates is how DCM models functional segregation. Here one can regard V1 as "segregating" motion from other visual information and distributing it to the motion-sensitive area V5. This segregation is modeled as a bilinear "enabling" of V1 to V5 connections when, and only when, motion is present. Note that in the absence of motion the intrinsic V1 to V5 connection was trivially small (in fact the estimate was −0.04). The key advantage of entering motion through a bilinear effect, as opposed to a direct effect on V5, is that we can finesse the inference that V5 shows motion-selective responses with the assertion that these responses are mediated by afferents from V1. The two bilinear effects above represent two important aspects of functional integration that DCM is able characterize.

←

Figure 7 Results of a dynamic causal modeling (DCM) analysis of attention to visual motion with functional magnetic resonance imaging. (*Right panel*) Functional architecture based upon the conditional estimates shown alongside their connections, with the percent confidence that they exceeded threshold in brackets. The most interesting aspects of this architecture involve the role of motion and attention in exerting bilinear effects. Critically, the influence of motion is to enable connections from V1 to the motion-sensitive area V5. The influence of attention is to enable backward connections from the inferior frontal gyrus (IFG) to the superior parietal cortex (SPC). Furthermore, attention increases the influence of SPC on the V5. Dotted arrows connecting regions represent significant bilinear affects in the absence of a significant intrinsic coupling. (*Left panel*) Fitted responses based upon the conditional estimates and the adjusted data are shown for each region in the DCM. The insert (*upper left*) shows the location of the regions.

The fMRI data were from a study in which subjects viewed identical stimuli (visual motion subtended by radially moving dots) under different attentional manipulations of the task (detection of velocity changes) (Büchel & Friston 1997). The data were acquired from a normal subject at 2 Tesla, using a whole-body MRI system, equipped with a head volume coil. Contiguous multislice T2*-weighted fMRI images were obtained with a gradient echo-planar sequence (TE = 40 ms, TR = 3.22 s, matrix size = 64 × 64 × 32, voxel size 3 × 3 × 3 mm). Each subject had four consecutive 100-scan sessions comprising a series of 10-scan blocks under five different conditions D F A F N F A F N S. The first condition (D) was a dummy condition to allow for magnetic saturation effects. F (fixation) corresponds to a low-level baseline where the subjects viewed a fixation point at the center of a screen. In condition A (attention), subjects viewed 250 dots moving radially from the center at 4.7° per second and were asked to detect changes in radial velocity. In condition N (no attention), the subjects were asked simply to view the moving dots. In condition S (stationary), subjects viewed stationary dots. The order of A and N was swapped for the last two sessions. In all conditions, subjects fixated the center of the screen. There were no speed changes during scanning. No overt response was required in any condition.

STRUCTURAL EQUATION MODELING The central ideal behind dynamic causal modeling is to treat the brain as a deterministic nonlinear dynamic system that is subject to inputs and produces outputs. Effective connectivity is parameterized in terms of coupling among unobserved brain states (e.g., neuronal activity in different regions). The objective is to estimate these parameters by perturbing the system and measuring the response. This is in contradistinction to established methods for estimating effective connectivity from neurophysiological time-series, which include structural equation modeling (Büchel & Friston 1997, McIntosh & Gonzalez-Lima 1994) and models based on multivariate auto-regressive processes (Goebel et al. 2003, Harrison et al. 2003). In these models, there is no designed perturbation and the inputs are treated as unknown and stochastic. Furthermore, the inputs are assumed to express themselves instantaneously such that, at the point of observation, the change in states will be zero. From Equation 7, in the absence of bilinear effects, we have

$$\begin{aligned} \dot{x} &= 0 \\ &= Ax + Cu \\ x &= -A^{-1}Cu. \end{aligned} \tag{8}$$

This is the regression equation used in SEM where $A = A' - I$ and A' contains the off-diagonal connections among regions. The key point here is that A' is estimated by assuming u is some random innovation with known covariance. This is not tenable for designed experiments when u represents carefully structured experimental inputs. Although SEM and related autoregressive techniques are useful for establishing dependence among regional responses, these techniques are not a surrogate for informed causal models based on the underlying dynamics of these responses.

CONCLUSION

In this article, we reviewed the main models that underpin image analysis and touched briefly on ways of assessing specialization and integration in the brain. The key principles of functional brain architectures were used to motivate the various models considered. These can be regarded as a succession of modeling endeavors, drawing more and more on our understanding of how brain-imaging signals are generated, both in terms of biophysics and the underlying neuronal interactions. We have seen how hierarchical linear observation models encode the treatments effects elicited by experimental design. General linear models based on convolution models imply an underlying dynamic input-state-output system. The form of these systems can be used to constrain convolution models and explore some of their simpler nonlinear properties. By creating observation models based on an explicit forward model of neuronal interactions, one can now start to model and assess interactions among distributed cortical areas and make inferences about coupling at the neuronal level. The next few years will probably see an increasing realism in the dynamic causal models introduced above (see Horwitz et al. 2001). Attempts already have been made to use plausible models of neuronal ensemble to estimate network parameters of evoked responses in EEG (David & Friston 2003). These endeavors are likely to encompass fMRI signals in the near future, enabling the conjoint modeling, or fusion, of different modalities and the marriage of computational neuroscience with the modeling of brain responses.

ACKNOWLEDGMENTS

The Wellcome Trust funded this work. I would like to thank all of my colleagues at the Wellcome Department of Imaging Neuroscience for their conceptual input.

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

LITERATURE CITED

Absher JR, Benson DF. 1993. Disconnection syndromes: an overview of Geschwind's contributions. *Neurology* 43:862–67

Adler RJ. 1981. *The Geometry of Random Fields*. New York: Wiley

Aertsen A, Preißl H. 1991. Dynamics of activity and connectivity in physiological neuronal networks. In *Nonlinear Dynamics and Neuronal Networks*, ed. HG Schuster, pp. 281–302. New York: VCH Publ.

Ashburner J, Friston KJ. 2000. Voxel-based morphometry—the methods. *NeuroImage* 11:805–21

Bandettini PA, Jesmanowicz A, Wong EC, Hyde JS. 1993. Processing strategies for time course data sets in functional MRI of the human brain. *Magn. Reson. Med.* 30:161–73

Baumgartner R, Scarth G, Teichtmeister C, Somorjai R, Moser E. 1997. Fuzzy clustering of gradient-echo functional MRI in the human visual cortex. Part 1: reproducibility. *J. Magn. Reson. Imaging* 7:1094–101

Bendat JS. 1990. *Nonlinear System Analysis and Identification from Random Data.* New York: Wiley

Berry DA, Hochberg Y. 1999. Bayesian perspectives on multiple comparisons. *J. Stat. Plan. Infer.* 82:215–27

Biswal B, Yetkin FZ, Haughton VM, Hyde JS. 1995. Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn. Reson. Med.* 34:537–41

Boynton GM, Engel SA, Glover GH, Heeger DJ. 1996. Linear systems analysis of functional magnetic resonance imaging in human V1. *J. Neurosci.* 16:4207–21

Büchel C, Friston KJ. 1997. Modulation of connectivity in visual pathways by attention: cortical interactions evaluated with structural equation modeling and fMRI. *Cereb. Cortex* 7:768–78

Büchel C, Wise RJS, Mummery CJ, Poline J-B, Friston KJ. 1996. Nonlinear regression in parametric activation studies. *NeuroImage* 4:60–66

Buckner R, Bandettini P, O'Craven K, Savoy R, Petersen S, et al. 1996. Detection of cortical activation during averaged single trials of a cognitive task using functional magnetic resonance imaging. *Proc. Natl. Acad. Sci. USA* 93:14878–83

Burock MA, Buckner RL, Woldorff MG, Rosen BR, Dale AM. 1998. Randomized event-related experimental designs allow for extremely rapid presentation rates using functional MRI. *NeuroReport* 9:3735–39

Buxton RB, Frank LR. 1997. A model for the coupling between cerebral blood flow and oxygen metabolism during neural stimulation. *J. Cereb. Blood Flow Metab.* 17:64–72

Calhoun VD, Adali T, Pearlson GD, Pekar JJ. 2001. Spatial and temporal independent component analysis of functional MRI data containing a pair of task-related waveforms. *Hum. Brain Mapp.* 13:43–53

Chawla D, Buechel C, Edwards R, Howseman A, Josephs O, et al. 1999. Speed-dependent responses in V5: a replication study. *NeuroImage* 9:508–15

Dale A, Buckner R. 1997. Selective averaging of rapidly presented individual trials using fMRI. *Hum. Brain Mapp.* 5:329–40

David O, Friston KJ. 2003. A neural mass model for MEG/EEG: coupling and neuronal dynamics. *NeuroImage* 20:1743–55

Dempster AP, Laird NM, Rubin DB. 1977. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser. B* 39:1–38

Fliess M, Lamnabhi M, Lamnabhi-Lagarrigue F. 1983. An algebraic approach to nonlinear functional expansions. *IEEE Trans. Circuits Syst.* 30:554–70

Friston KJ, Büchel C. 2000. Attentional modulation of effective connectivity from V2 to V5/MT in humans. *Proc. Natl. Acad. Sci. USA* 97:7591–96

Friston KJ, Frith CD, Fletcher P, Liddle PF, Frackowiak RSJ. 1996a. Functional topography: multidimensional scaling and functional connectivity in the brain. *Cereb. Cortex* 6:156–64

Friston KJ, Frith CD, Liddle PF, Frackowiak RSJ. 1991. Comparing functional (PET) images: the assessment of significant change. *J. Cereb. Blood Flow Metab.* 11:690–99

Friston KJ, Frith CD, Liddle PF, Frackowiak RSJ. 1993. Functional connectivity: the principal component analysis of large data sets. *J. Cereb. Blood Flow Metab.* 13:5–14

Friston KJ, Frith CD, Passingham RE, Liddle PF, Frackowiak RSJ. 1992. Motor practice and neurophysiological adaptation in the cerebellum: a positron tomography study. *Proc. R. Soc. London Ser. B* 248:223–28

Friston KJ, Frith CD, Turner R, Frackowiak RSJ. 1995a. Characterizing evoked hemodynamics with fMRI. *NeuroImage* 2:157–65

Friston KJ, Harrison L, Penny W. 2003. Dynamic causal modeling. *NeuroImage* 19:1273–302

Friston KJ, Holmes AP, Worsley KJ, Poline JB, Frith CD, Frackowiak RSJ. 1995b. Statistical parametric maps in functional imaging: a general linear approach. *Hum. Brain Mapp.* 2:189–210

Friston KJ, Jezzard PJ, Turner R. 1994. Analysis of functional MRI time-series. *Hum. Brain Mapp.* 1:153–71

Friston KJ, Mechelli A, Turner R, Price CJ. 2000. Nonlinear responses in fMRI: the Balloon model, Volterra kernels, and other hemodynamics. *NeuroImage* 12:466–77

Friston KJ, Penny W. 2003. Posterior probability maps and SPMs. *NeuroImage* 19:1240–49

Friston KJ, Penny W, Phillips C, Kiebel S, Hinton G, Ashburner J. 2002. Classical and Bayesian inference in neuroimaging: theory. *NeuroImage* 16:465–83

Friston KJ, Poline J-B, Holmes AP, Frith CD, Frackowiak RSJ. 1996b. A multivariate analysis of PET activation studies. *Hum. Brain Mapp.* 4:140–51

Friston KJ, Price CJ, Fletcher P, Moore C, Frackowiak RSJ, Dolan RJ. 1996c. The trouble with cognitive subtraction. *NeuroImage* 4:97–104

Gerstein GL, Perkel DH. 1969. Simultaneously recorded trains of action potentials: analysis and functional interpretation. *Science* 164:828–30

Goebel R, Roebroeck A, Kim DS, 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

Goltz F. 1881. In *Transactions of the 7th International Medical Congress*, ed. W MacCormac, Vol. I, pp. 218–28. London: Kolkmann

Grafton S, Mazziotta J, Presty S, Friston KJ, Frackowiak RSJ, Phelps M. 1992. Functional anatomy of human procedural learning determined with regional cerebral blood flow and PET. *J. Neurosci.* 12:2542–48

Harrison LM, Penny W, Friston KJ. 2003. Multivariate autoregressive modeling of fMRI time series. *NeuroImage* 19:1477–91

Hoge RD, Atkinson J, Gill B, Crelier GR, Marrett S, Pike GB. 1999. Linear coupling between cerebral blood flow and oxygen consumption in activated human cortex. *Proc. Natl. Acad. Sci. USA* 96:9403–8

Holmes A, Ford I. 1993. A Bayesian approach to significance testing for statistic images from PET. In *Quantification of Brain Function, Tracer Kinetics and Image Analysis in Brain PET*, ed. K Uemura, NA Lassen, T Jones, I. Kanno, Ser. 1030, pp. 521–34. Amsterdam: Excerpta Medica

Horwitz B, Friston KJ, Taylor JG. 2001. Neural modeling and functional brain imaging: an overview. *Neural Netw.* 13:829–46

Josephs O, Turner R, Friston KJ. 1997. Event-related fMRI. *Hum. Brain Mapp.* 5:243–48

Lange N, Zeger SL. 1997. Non-linear Fourier time series analysis for human brain mapping by functional magnetic resonance imaging (with discussion) *J. Roy. Stat. Soc. Ser. C* 46:1–29

Lueck CJ, Zeki S, Friston KJ, Deiber MP, Cope NO, et al. 1989. The color centre in the cerebral cortex of man. *Nature* 340:386–89

Mandeville JB, Marota JJ, Ayata C, Zararchuk G, Moskowitz MA, et al. 1999. Evidence of a cerebrovascular postarteriole Windkessel with delayed compliance. *J. Cereb. Blood Flow Metab.* 19:679–89

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

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

Mørch N, Kjems U, Hansen LK, Svarer C, Law I, et al. 1995. Visualization of neural networks using saliency maps. *IEEE Int. Conf. Neural Netw.*, pp. 2085–90. Perth, Austr.

Nichols TE, Holmes AP. 2002. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Hum. Brain Mapp.* 15:1–25

Passingham RE, Stephan KE, Kotter R. 2002. The anatomical basis of functional localization in the cortex. *Nat. Rev. Neurosci.* 3:606–16

Petersen SE, Fox PT, Posner MI, Mintun M, Raichle ME. 1989. Positron emission tomographic studies of the processing of single words. *J. Cogn. Neurosci.* 1:153–70

Phillips CG, Zeki S, Barlow HB. 1984. Localisation of function in the cerebral cortex: past, present, and future. *Brain* 107:327–61

Price CJ, Friston KJ. 1997. Cognitive conjunction: a new approach to brain activation experiments. *NeuroImage* 5:261–70

Price CJ, Wise RJS, Ramsay S, Friston KJ, Howard D, et al. 1992. Regional response differences within the human auditory cortex when listening to words. *Neurosci. Lett.* 146:179–82

Sychra JJ, Bandettini PA, Bhattacharya N, Lin Q. 1994. Synthetic images by subspace transforms. I. Principal component images and related filters. *Med. Phys.* 21:193–201

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

Worsley KJ, Evans AC, Marrett S, Neelin P. 1992. A three-dimensional statistical analysis for rCBF activation studies in human brain. *J. Cereb. Blood Flow Metab.* 12:900–18

Worsley KJ, Marrett S, Neelin P, Vandal AC, Friston KJ, Evans AC. 1996. A unified statistical approach or determining significant signals in images of cerebral activation. *Hum. Brain Mapp.* 4:58–73

Zeki S. 1990. The motion pathways of the visual cortex. In *Vision: Coding and Efficiency,* ed. C Blakemore, pp. 321–45. London: Cambridge Univ. Press

Annual Review of Psychology
Volume 56, 2005

CONTENTS

Frontispiece—*Richard F. Thompson* xviii

PREFATORY

In Search of Memory Traces, *Richard F. Thompson* 1

DECISION MAKING

Indeterminacy in Brain and Behavior, *Paul W. Glimcher* 25

BRAIN IMAGING/COGNITIVE NEUROSCIENCE

Models of Brain Function in Neuroimaging, *Karl J. Friston* 57

MUSIC PERCEPTION

Brain Organization for Music Processing, *Isabelle Peretz and Robert J. Zatorre* 89

SOMESTHETIC AND VESTIBULAR SENSES

Vestibular, Proprioceptive, and Haptic Contributions to Spatial Orientation, *James R. Lackner and Paul DiZio* 115

CONCEPTS AND CATEGORIES

Human Category Learning, *F. Gregory Ashby and W. Todd Maddox* 149

ANIMAL LEARNING AND BEHAVIOR: CLASSICAL

Pavlovian Conditioning: A Functional Perspective, *Michael Domjan* 179

NEUROSCIENCE OF LEARNING

The Neuroscience of Mammalian Associative Learning, *Michael S. Fanselow and Andrew M. Poulos* 207

HUMAN DEVELOPMENT: EMOTIONAL, SOCIAL, AND PERSONALITY

Behavioral Inhibition: Linking Biology and Behavior Within a Developmental Framework, *Nathan A. Fox, Heather A. Henderson, Peter J. Marshall, Kate E. Nichols, and Melissa A. Ghera* 235

BIOLOGICAL AND GENETIC PROCESSES IN DEVELOPMENT

Human Development: Biological and Genetic Processes, *Irving I. Gottesman and Daniel R. Hanson* 263

SPECIAL TOPICS IN PSYCHOPATHOLOGY

The Psychology and Neurobiology of Suicidal Behavior, *Thomas E. Joiner Jr., Jessica S. Brown, and LaRicka R. Wingate* 287

DISORDERS OF CHILDHOOD

Autism in Infancy and Early Childhood, *Fred Volkmar, Kasia Chawarska, and Ami Klin* 315

CHILD/FAMILY THERAPY

Youth Psychotherapy Outcome Research: A Review and Critique of the Evidence Base, *John R. Weisz, Amanda Jensen Doss, and Kristin M. Hawley* 337

ALTRUISM AND AGGRESSION

Prosocial Behavior: Multilevel Perspectives, *Louis A. Penner, John F. Dovidio, Jane A. Piliavin, and David A. Schroeder* 365

INTERGROUP RELATIONS, STIGMA, STEREOTYPING, PREJUDICE, DISCRIMINATION

The Social Psychology of Stigma, *Brenda Major and Laurie T. O'Brien* 393

PERSONALITY PROCESSES

Personality Architecture: Within-Person Structures and Processes, *Daniel Cervone* 423

PERSONALITY DEVELOPMENT: STABILITY AND CHANGE

Personality Development: Stability and Change, *Avshalom Caspi, Brent W. Roberts, and Rebecca L. Shiner* 453

WORK MOTIVATION

Work Motivation Theory and Research at the Dawn of the Twenty-First Century, *Gary P. Latham and Craig C. Pinder* 485

GROUPS AND TEAMS

Teams in Organizations: From Input-Process-Output Models to IMOI Models, *Daniel R. Ilgen, John R. Hollenbeck, Michael Johnson, and Dustin Jundt* 517

LEADERSHIP

Presidential Leadership, *George R. Goethals* 545

PERSONNEL EVALUATION AND COMPENSATION

Personnel Psychology: Performance Evaluation and Pay for Performance, *Sara L. Rynes, Barry Gerhart, and Laura Parks* 571

PSYCHOPHYSIOLOGICAL DISORDERS AND PSYCHOLOGICAL EFFECTS ON MEDICAL DISORDERS

Psychological Approaches to Understanding and Treating Disease-Related Pain, *Francis J. Keefe, Amy P. Abernethy, and Lisa C. Campbell* 601

TIMELY TOPIC

Psychological Evidence at the Dawn of the Law's Scientific Age, *David L. Faigman and John Monahan* 631

INDEXES

Subject Index 661
Cumulative Index of Contributing Authors, Volumes 46–56 695
Cumulative Index of Chapter Titles, Volumes 46–56 700

ERRATA

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