Dynamic Causal Modelling

Karl J Friston

The Wellcome Dept. of Cognitive Neurology,
University College London
Queen Square, London, UK WC1N 3BG
Tel (44) 020 7833 7456
Fax (44) 020 7813 1445
email k.friston@fil.ion.ucl.ac.uk

Contents

I INTRODUCTION

In this chapter we apply the system identification techniques described in **Chapter 17** (**Classical and Bayesian Inference**) to the bilinear state-space models of effective connectivity introduced in **Chapter 19** (**Effective connectivity**). By using a bilinear approximation, to the dynamics of any system, the parameters of the implicit causal model reduce to three sets. These comprise parameters that (i) mediate the influence of extrinsic inputs on the states, (ii) mediate intrinsic coupling among the states and (iii) [bilinear] parameters that allow the inputs to modulate that coupling.

We describe this approach for the analysis of effective connectivity using experimentally designed inputs and fMRI responses. In this context, the coupling parameters correspond to effective connectivity and the bilinear parameters reflect the changes in connectivity induced by inputs. The ensuing framework allows one to characterise fMRI experiments, conceptually, as an experimental manipulation of integration among brain regions (by contextual or trial-free inputs, like time or attentional set) that is revealed using evoked responses (to perturbations or trial-bound inputs like stimuli).

As with previous analyses of effective connectivity, the focus is on experimentally induced changes in coupling (*c.f.* psychophysiologic interactions). However, unlike previous approaches in neuroimaging, the causal model ascribes responses to designed deterministic inputs, as opposed to treating inputs as unknown and stochastic.

A Background

This chapter is about modelling interactions among neuronal populations, at a cortical level, using neuroimaging (hemodynamic or electromagnetic) time series. It presents the motivation and procedures for dynamic causal modelling of evoked brain responses. The aim of this modelling 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). Dynamic causal modelling represents a fundamental departure from existing approaches to effective connectivity because it employs a more plausible generative model of measured brain responses that embraces their nonlinear and dynamic nature.

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 is transformed into a measured response. This enables the parameters of the neuronal model (*i.e.* effective connectivity) to be estimated from observed data. These supplementary models may be forward models of electromagnetic measurements or hemodynamic models of fMRI measurements. In this chapter we will focus on fMRI. Responses are evoked by known deterministic inputs that embody designed changes in stimulation or context. This is accomplished by using a dynamic input-state-output model with multiple inputs and outputs. The inputs correspond to conventional stimulus functions that encode experimental manipulations. The state variables cover both the neuronal activities and other neurophysiological or biophysical variables needed to form the outputs. The outputs are measured electromagnetic or hemodynamic responses over the brain regions considered.

Intuitively, this scheme 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 the neuronal level. This represents a departure from conventional approaches (*e.g.* structural equation modelling and autoregression models; McIntosh & Gonzalez-Lima 1994; Büchel & Friston 1997; Harrison *et al* submitted), 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, for neuroimaging, pertains to how the experimental inputs enter the model and cause neuronal responses. We have established in previous chapters that 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 modelling 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 perceptual set. These distinctions are seen most clearly in relation to existing analyses and experimental designs.

B DCM and existing approaches

The central ideal, behind dynamic causal modelling (DCM), is to treat the brain as a deterministic nonlinear dynamic system that is subject to inputs and produces outputs. Effective connectivity is parameterised 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 modelling and models based on multivariate auto-regressive processes. In these models, there is no designed perturbation and the inputs are treated as unknown and stochastic. Multivariate autoregression models and their spectral equivalents like coherence analysis, not only assume the system is driven by stochastic innovations, but are restricted to linear interactions. Structural equation modelling assumes the interactions are linear and, furthermore, instantaneous in the sense that structural equation models are not time-series models. In short, dynamic causal modelling is distinguished from alternative approaches not just by accommodating the nonlinear and dynamic aspects of neuronal interactions, but by framing the estimation problem in terms of perturbations that accommodate experimentally designed inputs. This is a critical departure from conventional

approaches to causal modelling in neuroimaging and, importantly, brings the analysis of effective connectivity much closer to the analysis of region-specific effects. Dynamic causal modelling calls upon the same experimental design principles to elicit region-specific interactions that we use in conventional experiments to elicit region-specific activations. In fact, as shown later, the convolution model, used in the standard analysis of fMRI time-series, is a special and simple case of DCM that ensues when the coupling among regions is discounted. In DCM the causal or explanatory variables that comprise the conventional design matrix become the inputs and the parameters become measures of effective connectivity. Although DCM can be framed as a generalisation of the linear models used in conventional analyses to cover bilinear models (see below), it also represents an attempt to embed more plausible forward models of how neuronal dynamics respond to inputs and produces measured responses. This reflects the growing appreciation of the role that neuronal models may have to play in understanding measured brain responses (see Horwitz *et al* 2001 for a discussion)

This chapter can be regarded as an extension of previous work on the Bayesian identification of hemodynamic models (Friston 2002) to cover multiple regions. In **Chapter 17** (**Classical and Bayesian inference**) we focussed on the biophysical parameters of a hemodynamic response in a single region. The most important parameter was the efficacy with which experimental inputs could elicit an activity-dependent vasodilatory signal. In this chapter neuronal activity is modelled explicitly, allowing for interactions among the activities of multiple regions in generating the observed hemodynamic response. The estimation procedure employed for DCM is formally identical to that described in **Chapter 17** (**Classical and Bayesian inference**).

C DCM and experimental design

DCM is used to test the specific hypothesis that motivated the experimental design. It is not an exploratory technique, as with all analyses of effective connectivity the results are specific to the tasks and stimuli employed during the experiment. In DCM designed inputs can produce responses in one of two ways. Inputs can elicit changes in the state variables (*i.e.* neuronal activity) directly. For example, sensory input could be modelled

as causing direct responses in primary visual or auditory areas. The second way in which inputs effect the system is through changing the effective connectivity or interactions. Useful examples of this sort of effect would be the attentional modulation of connections between parietal and extrastriate areas. Another ubiquitous example of this second sort of contextual input would be time. Time-dependant changes in connectivity correspond to plasticity. It is useful to regard experimental factors as inputs that belong to the class that produce evoked responses or to the class of contextual factors that induce changes in coupling (although, in principle, all inputs could do both). The first class comprises trial- or stimulus-bound perturbations whereas the second establishes a context in which effects of the first sort evoke responses. This second class is typically trial-free and induced by task instructions or other contextual changes. Measured responses in high-order cortical areas are mediated by interactions among brain areas elicited by trial-bound perturbations. These interactions can be modulated by other set-related or contextual factors that modulate the latent or intrinsic coupling among areas. Figure 1 illustrates this schematically. The important implication here, for experimental design in DCM, is that it should be multifactorial, with at least one factor controlling sensory perturbation and another factor manipulating the context in which the sensory evoked responses are promulgated throughout the system (*c.f.* psychophysiological interaction studies Friston *et al* 1997).

In this chapter we use bilinear approximations to any DCM. The bilinear approximation reduces the parameters to three sets that control three distinct things. First, the direct or extrinsic influence of inputs on brain states in any particular area. Second; the intrinsic or latent connections that couple responses in one area to the state of others and, finally, change in this intrinsic coupling induced by inputs. Although, in some instances, the relative strengths of intrinsic connections maybe of interest, most analyses of DCMs focus on the changes in connectivity embodied in the bilinear parameters. The first set of parameters are generally of little interest in the context of DCM but are the primary focus in classical analyses of regionally specific effects. In classical analyses the only way experimental effects can be expressed is though a direct or extrinsic influence on each voxel because mass-univariate models (*e.g.* SPM) preclude connections and their modulation.

Figure 1 about here

DCM is used primarily to answer questions about the modulation of effective connectivity through inferences about the bilinear parameters described above. They are bilinear in the sense that an input-dependent change in connectivity can be construed as a second-order interaction between the input and activity in a source region, when causing a response in a target region. The key role of bilinear terms reflects the fact that the more interesting applications of effective connectivity address changes in connectivity induced by cognitive set or time. In short, DCM with a bilinear approximation allows one to claim that an experimental manipulation has "activated a pathway" as opposed to a cortical region. Bilinear terms correspond to psychophysiologic interaction terms in classical regression analyses of effective connectivity (Friston *et al* 1997) and those formed by moderator variables (Kenny & Judd 1984) in structural equation modelling (Büchel & Friston 1997). This bilinear aspect speaks again to the importance of multifactorial designs that allow these interactions to be measured and the central role of the context in which region-specific responses are formed (see McIntosh 2000).

D DCM and Inference

Because DCMs are not restricted to linear or instantaneous systems they are necessarily complicated and, potentially, need a large number of free parameters. This is why they have greater biological plausibility, in relation to alternative approaches. However, this makes the estimation of the parameters more dependent upon constraints. A natural way to embody the requisite constraints is within a Bayesian framework. Consequently, dynamic causal models are estimated using Bayesian or conditional estimators and inferences about particular connections are made using the posterior or conditional density. In other words, the estimation procedure provides the probability distribution of a coupling parameter in terms of its mean and standard deviation. Having established this posterior density, the probability that the connection exceeds some specified threshold is easily computed. Bayesian inferences like this are more straightforward and

interpretable than corresponding classical inferences and furthermore eschew the multiple comparison problem. The posterior density is computed using the likelihood and prior densities. The likelihood of the data, given some parameters, is specified by the DCM (in one sense all models are simply ways of specifying the likelihood of an observation). The prior densities on the connectivity parameters offer suitable constraints to ensure robust and efficient estimation. These priors harness some natural constraints about the dynamics of coupled systems (see below) but also allow the user to specify which connections are likely to be present and those which are not. An important use of prior constraints, of this sort, is the restriction of where inputs can elicit extrinsic responses. It is interesting to reflect that conventional analyses suppose that all inputs have unconstrained access to all brain regions. This is because classical models assume activations are caused directly by experimental factors, as opposed to being mediated by afferents from other brain areas.

Additional constraints, on the intrinsic connections and their modulation by contextual inputs, can also be specified but they are not necessary. These additional constraints can be used to finesse a model by making it more parsimonious, allowing one to focus on a particular connection. We will provide examples of this below. Unlike structural equation modelling, there are no limits on the number of connections that can be modelled because the assumptions and estimations scheme used by dynamic causal modelling are completely different, relying upon known inputs.

E Overview

This chapter comprises a theoretical section and three sections demonstrating the use and validity of DCM. In the theoretical section we present the conceptual and mathematical fundaments that are used in the remaining sections. The later sections address the face, predictive and construct validity of DCM respectively. Face validity entails the estimation and inference procedure identifies what it is supposed to. The subsequent section on predictive validity uses empirical data from an fMRI study of single word processing at different rates. These data were obtained consecutively in a series of contiguous sessions. This allowed us to repeat the DCM using independent realisations

of the same paradigm. Predictive validity, over the multiple sessions, was assessed in terms of the consistency of the effective connectivity estimates and their posterior densities. The final section on construct validity revisits changes in connection strengths among parietal and extrastriate areas induced by attention to optic flow stimuli. We have established previously attentionally mediated increases in effective connectivity using both structural equation modelling and a Volterra formulation of effective connectivity (Büchel and Friston 1997, Friston and Büchel 2000). Our aim here is to show that dynamic causal modelling led us to the same conclusions. This chapter ends with a brief discussion of dynamic causal modelling, its limitations and potential applications.

II THEORY

In this section we present the theoretical motivation and operational details upon which DCM rests. In brief, DCM is a fairly standard nonlinear system identification procedure using Bayesian estimation of the parameters of deterministic input-state-output dynamic systems. In this chapter the system can be construed as a number of interacting brain regions. We will focus on a particular form for the dynamics that corresponds to a bilinear approximation to any analytic system. However, the idea behind DCM is not restricted to bilinear forms.

This section is divided into three parts. First, we describe the DCM itself, then consider the nature of priors on the parameters of the DCM and finally summarise the estimation procedure used to find the posterior distribution of these parameters. The estimation conforms to the posterior density analysis under Gaussian assumptions described in **Chapter 17** (**Classical and Bayesian inference**. In the previous chapter we were primarily concerned with estimating the efficacy with which input elicits a vasodilatory signal, presumably mediated by neuronal responses to the input. The causal models in this chapter can be regarded as a collection of hemodynamic models, one for each area, in which the experimental inputs are supplemented with neural activity from other areas. The parameters of interest now embrace not only the direct efficacy of experimental inputs but also the efficacy of neuronal input from distal regions, *i.e.* effective connectivity (see Figure 1).

The posterior density analysis finds the maximum or mode of the posterior density of the parameters (*i.e.* the most likely coupling parameters given the data) by performing a gradient assent on the log posterior. The log posterior requires both likelihood and prior terms. The likelihood obtains from Gaussian assumptions about the errors in the observation model implied by the DCM. This likelihood or forward model is described in the next subsection. By combining the ensuing likelihood with priors on the coupling and hemodynamic parameters, described in the second subsection, one can form an expression for the posterior density that is used in the estimation.

A Dynamic Causal Models

The dynamic causal model is a multiple-input multiple-output (MIMO) system that comprises m inputs and l outputs with one output per region. The m inputs correspond to designed causes (*e.g.* boxcar or stick stimulus functions). The inputs are exactly the same as those used to form design matrices in conventional analyses of fMRI and can be expanded in the usual way when necessary (*e.g.* using polynomials or temporal basis functions). In principle, each input could have direct access to every region. However, in practice the extrinsic effects of inputs are usually restricted to a single input region. Each of the l regions produces a measured output that corresponds to the observed BOLD signal. These l time-series would normally be taken as the average or first eigenvariate of key regions, selected on the basis of a conventional analysis. Each region has five state variables. Four of these are of secondary importance and correspond to the state variables of the hemodynamic model first presented in Friston *et al* (2000) and described in previous chapters. These hemodynamic states comprise a vasodilatory signal, normalised flow, normalised venous volume, and normalised deoxyhemoglobin content. These variables are required to compute the observed BOLD response and are not influenced by the states of other regions.

Central to the estimation of effective connectivity or coupling parameters are the first state variables of each region. These correspond to average neuronal or synaptic activity and are a function of the neuronal states of other brain regions. We will deal first with

the equations for the neuronal states and then briefly reprise the differential equations that constitute the hemodynamic model for each region.

1 Neuronal State Equations

Restricting ourselves to the neuronal states $z = [z_1, \ldots, z_l]^T$ one can posit any arbitrary form or model for effective connectivity

$$\dot{z} = F(z, u, \theta) \qquad 1$$

Where F is some nonlinear function describing the neurophysiological influences that activity in all l brain regions z and inputs u exert upon changes in the others. θ are the parameters of the model whose posterior density we require for inference. It is not necessary to specify the form of equation (1) because its bilinear approximation provides a natural and useful re-parameterisation in terms of effective connectivity. The bilinear form of (1) is:

$$\begin{aligned} \dot{z} &\approx Az + \sum u_j B^j z + Cu \\ &= \left(A + \sum u_j B^j\right) z + Cu \end{aligned}$$

$$\begin{aligned} A &= \frac{\partial F}{\partial z} = \frac{\partial \dot{z}}{\partial z} \\ B^j &= \frac{\partial^2 F}{\partial x \partial u_j} = \frac{\partial}{\partial u_j}\frac{\partial \dot{z}}{\partial x} \\ C &= \frac{\partial F}{\partial u} \end{aligned} \qquad 2$$

The Jacobian or connectivity matrix A represents the first-order connectivity among the regions in the absence of input. Effective connectivity is the influence that one neuronal system exerts over another in terms of inducing a response $\partial \dot{z} / \partial z$. In DCM a response is defined in terms of a change in activity with time $\dot{z}$. This latent connectivity can be thought of as the intrinsic coupling in the absence of experimental perturbations. Notice that the state, which is perturbed, depends on the experimental design (*e.g.* baseline or

control state) and therefore the intrinsic coupling is specific to each experiment. The matrices B^j are effectively the change in intrinsic coupling induced by the *j*th input. They encode the input-sensitive changes in $\partial\dot{z}/\partial z$ or, equivalently, the modulation of effective connectivity by experimental manipulations. Because B^j are second-order derivatives these terms are referred to as bilinear. Finally, the matrix C embodies the extrinsic influences of inputs on neuronal activity. The parameters $\theta^c = \{A, B^j, C\}$ are the connectivity or coupling matrices that we wish to identify and define the functional architecture and interactions among brain regions at a neuronal level. Figure 2 shows an example of a specific architecture to demonstrate the relationship between the matrix form of the bilinear model and the underlying state equations for each region. Notice that the units of connections are per unit time and therefore correspond to rates. Because we are in a dynamical setting a strong connection means an influence that is expressed quickly or with a small time constant. It is useful to appreciate this when interpreting estimates and thresholds quantitatively. This is will be illustrated below.

The neuronal activity in each region causes changes in volume and deoxyhemoglobin to engender the observed BOLD response y as described next.

Figure 2 about here

2 Hemodynamic State Equations

The remaining state variables of each region are biophysical states engendering the BOLD signal and mediate the translation of neuronal activity into hemodynamic responses. Hemodynamic states are a function of, and only of, the neuronal state of each region. The state-equations have been described in **Chapters 11** (**Hemodynamic modelling** and **Chapter 17** (**Classical and Bayesian inference**). and constitute a hemodynamic model that embeds the Balloon-Windkessel model (Buxton *et al* 1998, Mandeville *et al* 1999). A list of the biophysical parameters $\theta^h = \{\kappa, \gamma, \tau, \alpha, \rho\}$ is provided in Table 1 and a schematic of the hemodynamic model is shown in Figure 3 that contains the state-equations and output nonlinearity (*i.e.* Equation 42 in **Chapter 17**: **Classical and Bayesian inference**).

Figure 3 about here

3 The model

Combining the neuronal states with the hemodynamic states $x = \{z, s, f, v, q\}$ gives us a full forward model specified by the neuronal state equation (2) and the hemodynamic equations in Figure 3

$$\begin{aligned} \dot{x} &= f(x, u, \theta) \\ y &= \lambda(x) \end{aligned} \qquad 3$$

with parameters $\theta = \{\theta^c, \theta^h\}$. For any set of parameters and inputs, the state equation can be integrated and passed though the output nonlinearity to give the predicted response $h(u, \theta)$. This integration can be made quite expedient by capitalising on the sparsity of stimulus functions commonly employed in fMRI designs. See **Chapter 17** (**Classical and Bayesian inference**). Integrating (3) is equivalent to a generalised convolution of the inputs with the systems Volterra kernels. These kernels are easily derived from the Volterra expansion of (3) (Bendat.1990),

$$\begin{aligned} h_i(u, \theta) &= \sum_k \int_0^t \cdots \int_0^t \kappa_i^k(\sigma_1, \ldots, \sigma_k) u(t - \sigma_1), \ldots, u(t - \sigma_k) d\sigma_1, \ldots, d\sigma_k \\ \kappa_i^k(\sigma_1, \ldots, \sigma_k) &= \frac{\partial^k y_i(t)}{\partial u(t - \sigma_1), \ldots, \partial u(t - \sigma_k)} \end{aligned} \qquad 4$$

either by numerical differentiation or analytically through bilinear approximations (see Friston 2002). κ_i^k is the *k*th order kernel for region *i*. For simplicity, (4) has been written for a single input. The kernels are simply a re-parameterisation of the model. We will use these kernels to characterise regional impulse responses at neuronal and hemodynamic levels later.

The forward can be made into an observation model by adding error and confounding or nuisance effects $X(t)$ to give $y = h(u,\theta) + X\beta + \varepsilon$. Here β are the unknown coefficients of the confounds. In the examples used below, $X(t)$ comprised a low order discrete cosine set, modelling low frequency drifts and a constant term. Following the approach described in **Chapter 17** (**Classical and Bayesian inference**) we note

$$\begin{aligned} y - h(u,\eta_{\theta|y}) &\approx J\Delta\theta + X\beta + \varepsilon \\ &= \left[J, X\right]\begin{bmatrix}\Delta\theta \\ \beta\end{bmatrix} + \varepsilon \\ \Delta\theta &= \theta - \eta_{\theta|y} \end{aligned} \qquad 5$$

This local linear approximation then enters an EM scheme as described previously

Until convergence {

E-step

$$\begin{aligned} J &= \frac{\partial h(\eta_{\theta|y})}{\partial\theta} \\ \bar{y} &= \begin{bmatrix} y - h(\eta_{\theta|y}) \\ \eta_\theta - \eta_{\theta|y} \end{bmatrix}, \qquad \bar{J} = \begin{bmatrix} J & X \\ 1 & 0 \end{bmatrix}, \qquad \bar{C}_\varepsilon = \begin{bmatrix} \sum \lambda_i Q_i & 0 \\ 0 & C_\theta \end{bmatrix} \\ C_{\theta|y} &= \left(\bar{J}^T \bar{C}_\varepsilon^{-1} \bar{J}\right)^{-1} \\ \begin{bmatrix}\Delta\eta_{\theta|y} \\ \eta_{\beta|y}\end{bmatrix} &= C_{\theta|y}\left(\bar{J}^T \bar{C}_\varepsilon^{-1} \bar{y}\right) \\ \eta_{\theta|y} &\leftarrow \eta_{\theta|y} + \Delta\eta_{\theta|y} \end{aligned} \qquad 6$$

M-Step

$$P = \overline{C}_{\varepsilon}^{-1} - \overline{C}_{\varepsilon}^{-1}\overline{J}C_{\theta|y}\overline{J}^{T}\overline{C}_{\varepsilon}^{-1}$$

$$\frac{\partial F}{\partial \lambda_i} = -\tfrac{1}{2}tr\{PQ_i\} + \tfrac{1}{2}\overline{y}^{T}P^{T}Q_iP\overline{y}$$

$$\left\langle \frac{\partial^2 F}{\partial \lambda_{ij}^2} \right\rangle = -\tfrac{1}{2}tr\{PQ_iPQ_j\}$$

$$\lambda \leftarrow \lambda - \left\langle \frac{\partial^2 F}{\partial \lambda^2} \right\rangle^{-1} \frac{\partial F}{\partial \lambda}$$

)

These expressions are formally the same as equations (47) and (48) in **Chapter 17** (**Classical and Bayesian inference**) but for the addition of confounding effects in *X*. These confounds are treated as fixed effects with infinite prior variance, which does not need to appear explicitly in (6).

Note that the prediction and observations encompass the entire experiment. They are therefore large *ln x* 1 vectors whose elements run over regions and time. Although the response variable could be viewed as a multivariate times-series it is treated as a single observation vector, whose error covariance embodies both temporal and interregional correlations. $C_{\varepsilon} = V \otimes \Sigma(\lambda) = \sum \lambda_i Q_i$. This covariance is parameterised by some covariance hyperparameters λ In the examples below these correspond to region-specific error variances assuming the same temporal correlations $Q_i = V \otimes \Sigma_i$ in which Σ_i is a *l x l* sparse matrix with the *i*th leading diagonal element equal to one.

Equation (6) enables us the estimate the conditional moments of the coupling parameters (and the hemodynamics parameters) plus the hyperparameters controlling observation error. However, to proceed we need to specify the priors.

B Priors

In this context we use a fully Bayesian approach because (i) there are clear and necessary constraints on neuronal dynamics that can be used to motivate priors on the coupling parameters and (ii) empirically determined priors on the biophysical hemodynamic

parameters are relatively easy to specify. We will deal first with priors on the coupling parameters.

1 Priors on the Coupling Parameters

It is self evident that neuronal activity cannot diverge exponentially to infinite values. Therefore, we know that, in the absence of input, the dynamics must to return to a stable mode. This means the largest real component of the eigenvalues of the intrinsic coupling matrix cannot exceed zero. We use this constraint to establish a prior density on the coupling parameters A that ensures the system is dissipative.

If the largest real eigenvalue (Lyapunov exponent) is less than zero the stable mode is a point attractor. If the largest Lyapunov exponent is zero the system will converge to a periodic attractor with oscillatory dynamics. Therefore, it is sufficient to establish a probabilistic upper bound on the inter-regional coupling strengths; imposed by Gaussian priors that ensures the largest Lyapunov exponent is unlikely to exceed zero. If the prior densities of each connection are independent then the prior density can be specified in terms of a variance for the off-diagonal elements of A. This variance can then be chosen to render the probability of the principal exponent exceeding zero, less than some suitably small value.

The specification of priors on the connections can be finessed by a re-parameterisation of the coupling matrices A and B^j.

$$A \to \sigma A = \sigma \begin{bmatrix} -1 & a_{12} & \cdots \\ a_{21} & -1 & \\ \vdots & & \ddots \end{bmatrix}$$

$$B^j \to \sigma B^j = \sigma \begin{bmatrix} b^j_{11} & b^j_{12} & \cdots \\ b^j_{21} & \ddots & \\ \vdots & & \end{bmatrix} \qquad 7$$

This factorisation into a scalar and normalised coupling matrix renders the normalised couplings adimensional, such that strengths of connections among regions are relative to their self-connections. From this point on, we will deal with normalised parameters. This particular factorisation enforces the same self-connection or temporal scaling σ in all regions. This is sensible given neuronal transients are likely to decay at a similar rate in different regions (different factorisations could be employed in a different context).

Consider any set of $l(l-1)$ inter-regional connections with sum of squared values $\xi = \sum a_{ij}^2$. For any given value of ξ the largest Lyapunov exponent λ_a obtains when the connections strength are equal $a_{ij} = a$, for all $i \neq j$ in which case

$$\begin{aligned} \lambda_a &= (l-1)a - 1 \\ \xi &= l(l-1)a^2 \end{aligned} \qquad 8$$

This means that as the sum of squared connection strengths reaches $\xi = l/(l-1)$, the maximum exponent attainable, approaches zero. Consequently, if ξ is constrained to be less than this threshold, we can set an upper bound on the probability that the principal exponent exceeds zero. ξ is constrained through the priors on a_{ij}. If each connection has a prior Gaussian density with zero expectation and variance C_a, then the sum of squares has a scaled Chi-squared distribution $\xi/C_a \sim \chi^2_{l(l-1)}$ with degrees of freedom $l(l-1)$. C_a is chosen to make $p(\xi > l/(l-1))$ suitably small. *i.e.*

$$C_a = \frac{l/(l-1)}{\phi_\chi^{-1}(1-p)} \qquad 9$$

where ϕ_χ is the cumulative $\chi^2_{l(l-1)}$ distribution and *p* is the required probability. As the number of regions increases, the prior variance on connections decreases.

In addition to constraints on the normalised connections, the factorisation in (7) requires the temporal scaling parameter σ to be greater than zero. This is simply achieved through a non-central prior density specified in terms of its moments such that

$\sigma \sim N(\eta_\sigma, C_\sigma)$ where the expectation η_σ controls the characteristic time constant of the system and the variance C_σ is chosen to ensure $p(\sigma > 0)$ is small, *i.e.*

$$C_a = \left(\frac{\eta_\sigma}{\phi_N^{-1}(1-p)} \right)^2 \qquad 10$$

ϕ_N is the cumulative normal distribution and *p* the required probability.

In summary, priors on the connectivity parameters ensure that the system remains stable. The spectrum of eigenvalues of the intrinsic coupling matrix determines the time-constants of orthogonal modes or patterns of regional activity. These are scaled by σ whose prior expectation controls the characteristic time-constants (*i.e.* those observed in the absence of coupling). We will assume a value of one second. The prior variance on this scaling parameter is chosen to make the probability it is less than zero suitably small (in our case 10^{-3}). The ensuing prior density can be expressed as a function of the implicit half-life $\tau_z(\sigma) = \ln 2/\sigma$ by noting $p(\tau_z) = p(\sigma)\partial\sigma/\partial\tau_z$. See Figure 4. This portrayal of the prior density shows that we expect regional transients with time constants in the range of a few hundred milliseconds to several seconds.

The prior distribution of individual connection strengths are assumed to be identically and independently distributed with a prior expectation of zero and a variance C_a that ensures the principal exponent has a very small probability of being greater than zero (here10^{-2}). This variance decreases with the number of connections or regions. To provide an intuition about how these priors keep the system from diverging exponentially, a quantitative example is shown in Figure 5. Figure 5 shows the prior density of two connections that renders the probability of a positive exponent less than 10^{-2}. It can be seen that this density lies in a domain of parameter space encircled by regions in which the maximum Lyapunov exponent exceeds zero (bounded by dotted lines). See the Figure legend for more details.

Priors on the bilinear coupling parameters have the same form (zero mean and variance) as those for the intrinsic coupling parameters. For consistency, these parameters are also normalised by σ and are consequently adimensional. Conversely,

priors on the influences to extrinsic input are not scaled and are relatively uninformative with zero expectation and unit variance. As noted in the introduction, additional constraints can be implemented by precluding certain connections. This is achieved by setting their variance to zero.

Figure 5 about here

2 Hemodynamic Priors

The hemodynamic priors are based on those used in Friston (2002) and in **Chapter 17** (**Classical and Bayesian inference**). In brief, the mean and variance of posterior estimates of the five biophysical parameters were computed over 128 voxels using the single word presentation data presented in the next section. These means and variances (see Table 1) were used to specify Gaussian priors on the hemodynamic parameters.

Combining the prior densities on the coupling and hemodynamic parameters allows us to express the prior probability of the parameters in terms of their prior expectation η_θ and covariance C_θ

$$\theta = \begin{bmatrix} \sigma \\ a_{ij} \\ b_{ij} \\ c_{ik} \\ \theta^h \end{bmatrix}, \qquad \eta_\theta = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ \eta_\theta^h \end{bmatrix}, \qquad C_\theta = \begin{bmatrix} C_\sigma & & & & \\ & C_A & & & \\ & & C_B & & \\ & & & 1 & \\ & & & & C_h \end{bmatrix} \qquad 11$$

Where the prior covariances C_A and C_B contain leading diagonal elements C_a for all connections that are allowed to vary. Having specified the priors, we are now in a position to form the posterior and proceed with estimation using (6).

C Inference

As noted above, the estimation scheme is a posterior density analysis under Gaussian assumptions. In short, the estimation scheme provides the approximating Gaussian posterior density of the parameters $q(\theta)$ in terms of its expectation $\eta_{\theta|y}$ and covariance $C_{\theta|y}$. The expectation is also known as the posterior mode or maximum *a posteriori* (MAP) estimator. The marginal posterior probabilities are then used for inference that any particular parameter or contrast of parameters $c^T\eta_{\theta|y}$ (*e.g.* average) exceeded a specified threshold γ.

$$p = \phi_N\left(\frac{c^T\eta_{\theta|y} - \gamma}{\sqrt{c^T C_{\theta|y} c}}\right) \qquad 12$$

As above ϕ_N is the cumulative normal distribution. In this chapter, we are primarily concerned with the coupling parameters θ^c and, among these, the bilinear terms. The units of these parameters are Hz or per second (or adimensional if normalised) and the thresholds are specified as such. In dynamical modelling, strength corresponds to a fast response with a small time constant.

D Relationship to conventional analyses

It is interesting to note that conventional analyses of fMRI data using linear convolution models are a special case of dynamic causal models using a bilinear approximation. This is important because it provides a direct connection between DCM and classical models. If we allow inputs to be connected to all regions and discount interactions among regions by setting the prior variances on *A* and *B* to zero we produce a set of disconnected brain regions or voxels that respond to, and only to, extrinsic input. The free parameters of interest reduce to the values of *C*, which reflect the ability of input to excite neural activity in each voxel. By further setting the prior variances on the self connections (*i.e.* scaling parameter) and those on the hemodynamic parameters to zero we end up with a

single-input-single-output model at each and every brain region that can be reformulated as a convolution model as described in Friston (2002). For voxel *i* and input *j* the parameter c_{ij} can be estimated by simply convolving the input with $\partial\kappa_i^1/\partial c_{ij}$ where κ_i^1 is the first-order kernel meditating the influence of input *j* on output *i*. The convolved inputs are then used to form a general linear model that can be estimated using least squares in the usual way. This is precisely the approach adopted in classical analyses, in which $\partial\kappa_i^1/\partial C_{ij}$ is the hemodynamic response function. The key point here is that the general linear models used in typical data analyses are special cases of Bilinear models *that embody more assumptions*. These assumptions enter through the use of highly precise priors that discount interactions among regions and prevent any variation in biophysical responses. Having described the theoretical aspects of DCM we now turn to applications and assessing its validity.

III. FACE VALIDITY - SIMULATIONS

A Introduction

In this section we use simulated data to establish the utility of the bilinear approximation and the robustness of the estimation scheme described in the previous section. We deliberately chose an architecture that would be impossible to characterise using existing methods based on regression models (*e.g.* structural equation modelling). This architecture embodies loops and reciprocal connections and poses the problem of vicarious input; the ambiguity between the direct influences of one area and influences that are mediated through others.

1 The simulated system

The architecture is depicted in Figure 6 and has been labelled so that it is consistent with the DCM characterised empirically in the next section. The model comprises three regions; a primary (**A1**) and secondary (**A2**) auditory area and a higher-level region (**A3**). There are two inputs. The first is a sensory input u_1 encoding the presentation of epochs

of words at different frequencies. The second input u_2 is contextual in nature and is simply an exponential function of the time elapsed since the start of each epoch (with a time constant of 8 seconds). These inputs were based on a real experiment and are the same as those used in the empirical analyses of the next section. The scaling of the inputs is important for the quantitative evaluation of the bilinear and extrinsic coupling parameters. The convention adopted here is that inputs encoding events approximate delta functions such that their integral over time corresponds to the number of events that have occurred. For event-free inputs, like the maintenance of a particular instructional set, the input is scaled to a maximum of unity, so that the integral reflects the number of seconds over which the input was prevalent. The inputs were specified in time bins that were a sixteenth of the interval between scans (repetition time; TR = 1.7s).

Figure 6 about here

The auditory input is connected to the primary area; the second input has no direct effect on activity but modulates the forward connections from **A1** to **A2** so that its influence shows *adaptation* during the epoch. The second auditory area receives input from the first and sends signals to the higher area (**A3**). In addition to reciprocal backward connection, in this simple auditory hierarchy, a connection from the lowest to the highest area has been included. Finally, the first input (word presentation) modulates the self-connections of the third region. This influence has been included to show how bilinear effects can emulate nonlinear responses. A bilinear modulation of the self-connection can augment or attenuate decay of synaptic activity rendering the average response to streams of stimuli rate-dependent. This is because the bilinear effect will only be expressed if sufficient synaptic activity persists after the previous stimulus. This, in turn, depends on a sufficiently fast presentation rate. The resulting response emulates a saturation at high presentation rates or small stimulus onset asynchronies that has been observed empirically. Critically, we are in a position to disambiguate between neuronal saturation, modelled by this bilinear term, and hemodynamic saturation, modelled by nonlinearities in the hemodynamic component of this DCM. A significant bilinear self-connection implies neuronal saturation above and beyond that attributable to

hemodynamics. Figure 7 illustrates this neuronal *saturation* by plotting the simulated response of **A3** in the absence of saturation $B^1 = 0$ against the simulated response with $b^1_{3,3} = -0.4$. It is evident that there is a nonlinear sub-additive effect at high response levels. It should be noted that true neuronal saturation of this sort is mediated by second order interactions among the states (*i.e.* neuronal activity). However, as shown in Figure 7 we can emulate these effects by using the first extrinsic input as a surrogate for neuronal inputs from other areas in the bilinear component of the model.

Figure 7 about here

Using this model we simulated responses using the values for A, B^1, B^2 and C given in Figure 6 and the prior expectations for the biophysical parameters given in Table 1. The values of the coupling parameters were chosen to emulate those seen typically in practice. This ensured the simulated responses were realistic in relation to simulated noise. After down-sampling these deterministic responses every 1.7 seconds (the TR of the empirical data used in the next section) we added known noise to produce simulated data. These data comprised time-series of 256 observations with independent or serially correlated Gaussian noise based on an AR(1) process. Unless otherwise stated, the noise had 0.5 standard deviation and was *i.i.d.* (independently and identically distributed). The drift terms were formed from the first six components of a discrete cosine set mixed linearly with normal random coefficients, scaled by one over the order. This emulates a $1/f^2$ plus white noise spectrum for the noise and drifts. See the lower panel of Figure 6 for an exemplar data simulation with i.i.d. noise of unit variance.

2. Exemplar analysis

The analysis described in the previous section was applied to the data shown in Figure 6. The priors on coupling parameters were augmented by setting the variance of the off-diagonal elements of B^1 (saturation) and all but two connections in B^2 (adaptation) to zero. These two connections were the first and second forward connections of this cortical hierarchy. The first had simulated adaptation, whereas the second did not.

Extrinsic input was restricted to the primary area **A1** by setting the variances of all but c_{11} to zero. We placed no further constraints on the intrinsic coupling parameters. This is equivalent to allowing full connectivity. This would be impossible with structural equation modelling. The results are presented in Figure 8 in terms of the MAP or conditional expectations of the coupling parameters (upper panels) and the associated posterior probabilities (lower panels) using Eq(14). It can be seen that the intrinsic coupling parameters are estimated reasonably accurately with a slight overestimation of the backward connection from **A3** to **A2**. The bilinear coupling parameters modelling adaptation are shown in the lower panels and the estimators have correctly identified the first forward connection as the locus of greatest adaptation. The posterior probabilities suggest inferences about the coupling parameters would lead us to the veridical architecture if we considered only connections whose half life exceeded 4 seconds with 90% confidence or more.

Figures 8 and 9 about here

The MAP estimates allow us to compute the MAP kernels associated with each region both in terms of neuronal output and hemodynamics response using Eq (6). The neuronal and hemodynamic kernels for the three regions are shown in Figure 9 (upper panels). It is interesting to note that the regional variation in the form of the neuronal kernels is sufficient to induce differential onset and peak latencies, in the order of a second or so, in the hemodynamic kernels despite the fact that neuronal onset latencies are the same. This difference in form is due to the network dynamics as activity is promulgated up the system and is recursively re-entered into lower levels Notice also that the neuronal kernels have quite protracted dynamics compared to the characteristic neuronal time constants of each area (about a second). This enduring activity, particularly in the higher two areas is a product of the network dynamics. The MAP estimates also enable us to compute the predicted response (lower left panel) in each region and compare it to the true response without observation noise (lower right panel). This comparison shows that the actual and predicted responses are very similar.

In Friston *et al* (2002) we repeated this estimation procedure to explore the face validity of the estimation scheme over a range of hyperparameters like noise levels, slice timing artifacts, extreme values of the biophysical parameters *etc*. In general the scheme proved to be robust to most violations assessed. Here we will just look at the effects of error variance on estimation because this speaks to some important features of Bayesian estimation in this context and the noise levels that can be tolerated.

B. The effects of noise

In this sub-section we investigate the sensitivity and specificity of posterior density estimates to the level of observation noise. Data were simulated as described above and mixed with various levels of white noise. For each noise level the posterior densities of the coupling parameters were estimated and plotted against the noise hyperparameter (expressed as its standard deviation) in terms of the posterior mean and 90% confidence intervals. Figure 10 shows some key coupling parameters that include both zero and non-zero connection strengths. The solid lines represent the posterior expectation or MAP estimator and the broken lines indicate the true value. The grey areas encompass the 90% confidence regions. Characteristic behaviours of the estimation are apparent from these results. As one might intuit, increasing the level of noise increases the uncertainty in the posterior estimates as reflected by an increase in the conditional variance and a widening of the confidence intervals. This widening is, however, bounded by the prior variances to which the conditional variances asymptote, at very high levels of noise. Concomitant with this effect is "shrinkage" of some posterior means to their prior expectation of zero. Put simply, when the data become very noisy the estimation relies more heavily upon priors and the prior expectation is given more weight. This is why priors of the sort used here are referred to as "shrinkage priors". These simulations suggest that for this level of evoked response, noise levels between 0-2 permit the connection strengths can be identified with a fair degree of precision and accuracy. Noise levels in typical fMRI experiments are about 0.5-1.5. The units of signal and noise are adimensional and correspond to percentage whole brain mean. Pleasingly, noise did not lead to false inferences in the sense that the posterior densities always encompassed the true values even at high levels of noise (Figure 10).

Figure 10 about here

IV PREDICTIVE VALIDITY – AN ANALYSIS OF SINGLE WORD PROCESSING

A Introduction

In this section we illustrate the predictive validity of DCM by showing that reproducible results can be obtained from independent data. The data set we used was especially designed for these sorts of analyses, comprising over 1,200 scans with a relatively short TR of 1.7 seconds. This necessitated a limited field of coverage but provided relatively high temporal acuity. The paradigm was a passive listening task, using epochs of single words presented at different rates. These data have been used previously to characterise nonlinear aspects of hemodynamics (e.g. Friston *et al* 1998, 2000, and 2002). Details of the experimental paradigm and acquisition parameters are provided in the legend to Figure 11. These data were acquired in consecutive sessions of a 120 scans enabling us to analyse the entire time-series or each session independently. We first present the results obtained by concatenating all the sessions into a single data sequence. We then revisit the data, analysing each session independently to provide 10 independent conditional estimates of the coupling parameters to assess reproducibility and mutual predictability.

Figure 11 about here

B Analysis of the complete time-series

Three regions were selected using maxima of the SPM{F} following a conventional SPM analysis (see Figure 11). The three maxima were those that were closest to the primary and secondary auditory areas and Wernicke's area in accord with the anatomic designations provided in the atlas of Talairach and Tournoux (1988). Region-specific time-series comprised the first eigenvariate of all voxels within a 4mm-radius sphere centred on each location. The anatomical locations are shown in Figure 11. As in the

simulations there were two inputs corresponding to a delta function for the occurrence of an aurally presented word and a parametric input modelling within-epoch adaptation. The outputs of the system were the three eigenvariate time-series from each region. As in the previous section we allowed for a fully connected system. In other words, each region was potentially connected to every other region. Generally, one would impose constraints on highly unlikely or implausible connections by setting their prior variance to zero. However, we wanted to demonstrate that dynamic causal modelling can be applied to connectivity graphs that would be impossible to analyse with structural equation modelling. The auditory input was connected to **A1**. In addition, auditory input entered bilinearly to emulate saturation, as in the simulations. The contextual input, modelling putative adaptation, was allowed to exert influences over all intrinsic connections. From a neurobiological perspective an interesting question is whether plasticity can be demonstrated in forward connections or backward connections. Plasticity, in this instance, entails a time-dependent increase or decrease in effective connectivity and would be inferred by significant bilinear coupling parameters associated with the second input.

Figures 12 and 13 about here

The inputs, outputs and priors on the DCM parameters were entered into the Bayesian estimation procedure as described above. Drifts were modelled with the first 40 components of a discrete cosine set, corresponding to X in Eq(6). The results of this analysis, in terms of the posterior densities and ensuing Bayesian inference are presented in Figures 12 and 13. Bayesian inferences were based upon the probability that the coupling parameters exceeded 0.0866. This corresponds to a half-life of 8 seconds. Intuitively, this means that we only consider the influences, of one region on another, to be meaningfully large if this influence is expressed within a time frame of 8 seconds or less. The results show that the most probable architecture, given the inputs and data, conforms to a simple hierarchy of forward connections where **A1** influences **A2** and **WA**, whereas **A2** sends connections just to **WA** (Figure 12). Although backward connections between **WA** and **A2** were estimated to be greater than our threshold with 82%

confidence they are not shown in Figure 12 (which is restricted to posterior probabilities of 90% or more). Saturation could be inferred in **A1** and **WA** with a high degree of confidence with b_{11}^{1} and b_{33}^{1} being greater than .5. Significant plasticity or time-dependent changes were expressed predominantly in the forward connections, particularly that between **A1** and **A3** *i.e.* $b_{13}^{2} = 0.37$. The conditional estimates are shown in more detail in Figure 13 along with the conditional fitted responses and associated kernels. A full posterior density analysis for a particular contrast of effects is shown in Figure 13a (lower panel). This contrast tested for the average plasticity over all forward and backward connections and demonstrates that we can be virtually certain plasticity was greater than zero.

This analysis illustrates three things. First, the DCM has defined a hierarchical architecture that is a sufficient explanation for the data and is indeed, the most likely given the data. This hierarchical structure was not part of the prior constraints because we allowed for a fully connected system. Second, the significant bilinear effects of auditory stimulation suggest there is measurable neuronal saturation above and beyond that attributable to hemodynamic nonlinearities. This is quite significant because such disambiguation is usually impossible given just hemodynamic responses. Finally, we were able to show time-dependent decreases in effective connectivity in forward connections from **A1**. Although this experiment was not designed to test for plasticity, the usefulness of DCM, in studies of learning and priming, should be self-evident.

C Reproducibility

The analysis above was repeated identically for each and every 120-scan session to provide 10 sets of Bayesian estimators. Drifts were modelled with the first 4 components of a discrete cosine set. The estimators are presented graphically in Figure 14 and demonstrate extremely consistent results. In the upper panels the intrinsic connections are shown to be very similar in their profile again reflecting a hierarchical connectivity architecture. The conditional means and 90% confidence regions for two connections are shown in Figure 14a. These connections included the forward connection from **A1** to **A2** that is consistently estimated to be very strong. The backward connection from **WA** to **A2** was weaker but was certainly greater than zero in every analysis. Equivalent results

were obtained for the modulatory effects or bilinear terms, although the profile was less consistent (Figure 14b). However, the posterior density of the contrast testing for average time-dependent adaptation or plasticity is relatively consistent and again almost certainly greater than zero, in each analysis.

To illustrate the stability of hyperparameter estimates, over the 10 sessions, the standard deviations of observation error are presented for each session over the three areas in Figure 15. As typical of studies at this field strength the standard deviation of noise is about 0.8-1% whole brain mean. It is pleasing to note that the session to session variability in hyperparameter estimates was relatively small, in relation to region to region differences.

In summary, independent analyses of data acquired under identical stimulus conditions, on the same subject, in the same scanning session, yield remarkably similar results. These results are biologically plausible and speak to the interesting notion that time-dependent changes, following the onset of a stream of words, are prominent in forward connections among auditory areas.

Figure 14 and 15 about here

V. CONSTRUCT VALIDITY – AN ANALYSIS OF ATTENTIONAL EFFECTS ON CONNECTIONS

A Introduction

In this final section we address the face validity of DCM. In previous chapters we have seen that attention positively modulates the backward connections in a distributed system of cortical regions mediating attention to radial motion. We use the same data in this section. In brief, subjects viewed optic flow stimuli comprising radially moving dots at a fixed velocity. In some epochs, subjects were asked to detect changes in velocity (that did not actually occur). This attentional manipulation was validated *post-hoc* using psychophysics and the motion after-effect. Analyses using structural equation modelling (Büchel & Friston 1997) and a Volterra formulation of effective connectivity (Friston & Büchel 2000) have established a hierarchical backwards modulation of effective

connectivity where a higher area increases the effective connectivity among two subordinate areas. These analyses have been extended using variable parameter regression and Kalman filtering (Büchel & Friston 1998) to look at the effective of attention directly on interactions between **V5** and the posterior parietal complex. In this context, the Volterra formulation can be regarded as a highly finessed regression model that embodies nonlinear terms and some dynamic aspects of fMRI time-series. However, even simple analyses, such as those employing psychophysiological interactions, point to the same conclusion that attention generally increases the effective connectivity among extrastriate and parietal areas. In short, we have established that the superior posterior parietal cortex (**SPC**) exerts a modulatory role on **V5** responses using Volterra-based regression models (Friston and Büchel 2000) and that the inferior frontal gyrus (**IFG**) exerts a similar influence on **SPC** using structural equation modelling (Büchel and Friston 1997). The aim of this section was to show that DCM leads one to the same conclusions but using a completely different approach.

1 Analysis

The experimental paradigm and data acquisition parameters are described in the legend to Figure 16. This Figure also shows the location of the regions that entered into the DCM (Figure 16b - insert). These regions were based on maxima from conventional SPMs testing for the effects of photic stimulation, motion and attention. As in the previous section, regional time courses were taken as the first eigenvariate of spherical volumes of interest centred 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 one 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 outputs corresponded to the four regional eigenvariates in (Figure 16b). 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 modelled as a bilinear modulation of the **V1** to **V5**

connectivity and attention was allowed to modulate the backward connections from **IFG** and **SPC**.

Figure 16 about here

The results of the DCM are shown in Figure 16a. Of primary interest here is the modulatory effect of attention that is expressed in terms of the bilinear coupling parameters for this third input. As hoped, 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 in Figure 16a). It is important to note that the only way that attentional manipulation can 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** responses in Figure 24b). This attentional modulation is accounted for, sufficiently, by changing just two connections. This change is, presumably, instantiated by instructional set at the beginning of each epoch. The second thing, this analysis illustrates, is the how functional segregation is modelled in DCM. Here one can regard **V1** as a 'segregating' motion from other visual information and distributing it to the motion-sensitive area **V5**. This segregation is modelled 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 MAP 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 was designed to characterise.

VI CONCLUSION

In this chapter we have presented dynamic causal modelling. DCM is a causal modelling procedure for dynamical systems in which causality is inherent in the differential equations that specify the model. The basic idea is to treat the system of interest, in this case the brain, as an input-state-output system. By perturbing the system with known inputs, measured responses are used to estimate various parameters that govern the evolution of brain states. Although there are no restrictions on the parameterisation of the model, a bilinear approximation affords a simple re-parameterisation in terms of effective connectivity. This effective connectivity can be latent or intrinsic or, through bilinear terms, model input-dependent changes in effective connectivity. Parameter estimation proceeds using fairly standard approaches to system identification that rest upon Bayesian inference.

Dynamic causal modelling represents a fundamental departure from conventional approaches to modelling effective connectivity in neuroscience. The critical distinction between DCM and other approaches, such as structural equation modelling or multivariate autoregressive techniques is that the input is treated as known, as opposed to stochastic. In this sense DCM is much closer to conventional analyses of neuroimaging time series because the causal or explanatory variables enter as known fixed quantities. The use of designed and known inputs in characterising neuroimaging data with the general linear model or DCM is a more natural way to analyse data from designed experiments. Given that the vast majority of imaging neuroscience relies upon designed experiments we consider DCM a potentially useful complement to existing techniques. In the remainder of this section we consider two potential limitations of DCM and comment upon extensions. We develop this point and the relationship of DCM to other approaches in the final chapter of this section.

References

JS Bendat. (1990) Nonlinear System Analysis and Identification from Random Data. John Wiley and Sons, New York USA

Büchel C, Friston KJ (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 Friston KJ Dynamic Changes in Effective Connectivity Characterised by Variable Parameter Regression and Kalman Filtering. Human Brain Mapping 1998; 6:403-408

RB Buxton, EC Wong, and LR Frank. Dynamics of blood flow and oxygenation changes during brain activation: The Balloon model. (1998) MRM 39, 855-864

Friston KJ Büchel C Fink GR Morris J Rolls E and Dolan RJ (1997) Psychophysiological and modulatory interactions in neuroimaging. NeuroImage 6:218-229

KJ Friston, O Josephs, G Rees, and R Turner. (1998) Nonlinear event-related responses in fMRI. MRM 39, 41-52

Friston KJ and Büchel C (2000). Attentional modulation of effective connectivity from V2 to V5/MT in humans. Pro Natl Acad. Sci USA 97:7591-7596

Friston KJ Mechelli A Turner R & Price CJ. Nonlinear responses in fMRI: The Balloon model, Volterra kernels and other hemodynamics. NeuroImage 2000; 12:466-477

Friston KJ (2002) Bayesian estimation of dynamical systems: An application to fMRI. NeuroImage – 16:513-530

Friston KJ, Penny W, Phillips C, Kiebel S, Hinton G and Ashburner J. (2002) Classical and Bayesian inference in neuroimaging: Theory. NeuroImage. 16:465-483

Friston KJ, Harrison L and Penny W. (2002) Dynamic Causal Modelling. NeuroImage. *under revision*

Gerstein GL and Perkel DH. (1969) Simultaneously recorded trains of action potentials: Analysis and functional interpretation. Science 164: 828-830

Grubb RL, Rachael ME, Euchring JO, and Ter-Pogossian MM. (1974) The effects of changes in PCO2 on cerebral blood volume, blood flow and vascular mean transit time. Stroke 5, 630-639

Harrison LM, Penny W and Friston K.J (2003) Multivariate Autoregressive Modelling of fMRI
time series. NeuroImage. Submitted

Harville DA (1977) Maximum likelihood approaches to variance component estimation and to related problems. J. Am. Stat.. Assoc. 72:320-338

Horwitz B Friston KJ & Taylor JG Neural modeling and functional brain imaging: an overview. Neural Networks 2001; 13:829-846

Kenny DA, Judd CM (1984) Estimating nonlinear and interactive effects of latent variables. Psychol Bull 96:201-210.

Mandeville JB, Marota JJ, Ayata C, Zararchuk G, Moskowitz MA, B Rosen B, and Weisskoff RM (1999) Evidence of a cerebrovascular postarteriole Windkessel with delayed compliance. J. Cereb. Blood Flow Metab. 19, 679-689

McIntosh AR, Gonzalez-Lima F (1994) Structural equation modelling and its application to network analysis in functional brain imaging. Human Brain Mapping 2:2-22.

McIntosh AR (2000) Towards a network theory of cognition. Neural Networks 13:861-870

Neal R.M., and Hinton G.E. (1998) A view of the EM algorithm that justifies incremental, sparse and other variants. in Learning in Graphical models, MI Jordan Ed. Kluwer Academic Press, pp 355-368

J Talairach, and P Tournoux. (1988) A Co-planar stereotaxic atlas of a human brain. Thieme, Stuttgart.

Figure Legends

FIGURE 1

This is a schematic illustrating the concepts underlying dynamic causal modelling. In particular it highlights the two distinct ways in which inputs or perturbations can illicit responses in the regions or nodes that comprise the model. In this example there are five nodes, including visual areas **V1** and **V4** in the fusiform gyrus, areas 39 and 37 and the superior temporal gyrus **STG**. Stimulus-bound perturbations designated u_1 act as extrinsic inputs to the primary visual area **V1**. Stimulus-free or contextual inputs u_2 mediate their effects by modulating the coupling between **V4** and **BA39** and between **BA37** and **V4**. For example, the responses in the angular gyrus (**BA39**) are caused by inputs to **V1** that are transformed by **V4**, where the influences exerted by **V4** are sensitive to the second input. The dark square boxes represent the components of the DCM that transform the state variables z_i in each region (neuronal activity) into a measured (hemodynamic) response y_i.

FIGURE 2

This schematic (upper panel) recapitulates the architecture in Figure 1 in terms of the differential equations implied by a bilinear approximation. The equations in each of the white areas describe the change neuronal activity z_i in terms of linearly separable components that reflect the influence of other regional state variables. Note particularly, how the second contextual inputs enter these equations. They effectively increase the intrinsic coupling parameters (a_{ij}) in proportion to the bilinear coupling parameters (b_{ij}^k). In this diagram the hemodynamic component of the DCM illustrates how the neuronal states enter a region-specific hemodynamic model to produce the outputs y_i that are a function of the region's biophysical states reflecting deoxyhemoglobin content and venous volume (q_i and v_i). The lower panel reformulates the differential equations in the upper panel into a matrix format. These equations can be summarised more compactly in terms of coupling parameter matrices A, B^j and C. This form of

expression is used in the main text and shows how it relates to the underlying differential equations that describe the state dynamics.

FIGURE 3

This schematic shows the architecture of the hemodynamic model for a single region (regional subscripts have been dropped for clarity). Neuronal activity induces a vasodilatory and activity-dependent signal *s* that increases the flow *f*. Flow causes changes in volume and deoxyhemoglobin (*v* and *q*). These two hemodynamic states enter the output nonlinearity Eq(4) to give the observed BOLD response *y*. This transformation from neuronal states z_i to hemodynamic response y_i is encoded graphically by the dark-grey boxes in the previous figure and in the insert above.

FIGURE 4

Prior probability density functions for the temporal scaling parameter or self-connection σ. This has a Gaussian form (left panel) that translates into a skewed distribution, when expressed in terms of the characteristic half-life of neural transients τ_z in any particular region (right panel). This prior distribution implies that neuronal activity will decay with a half-life of roughly 500 milliseconds, falling in the range of 300 ms to 2s.

FIGURE 5

Prior probability density on the intrinsic coupling parameters for a specific intrinsic coupling matrix *A*. The left-hand panel shows the real value of the largest eigenvalue of *A* (the principal Lyapunov exponent) as a function of the connection from the first to the second region and the reciprocal connection from the second to the first. The remaining connections were held constant at 0.5. This density can be thought of as a slice through a multidimensional spherical distribution over all connections. The right panel shows the prior probability density function and the boundaries at which the largest real eigenvalue exceeds zero (dotted lines). The variance or dispersion of this probability distribution is chosen to ensure that the probability of excursion into unstable domains of parameter space is suitably small. These domains are the upper right and lower left bounded regions.

FIGURE 6

This is a schematic of the architecture used to generate simulated data. Non-zero intrinsic connections are shown as directed black arrows with the strength or true parameter alongside. Here, the perturbing input is the presentation of words (sensory inputs) and acts as an intrinsic influence on **A1**. In addition, this input modulates the self-connection of **A3** to emulate saturation like-effects (see main text and Figure 7). The contextual input is a decaying exponential of within-epoch time and positively modulates the forward connection from **A1** to **A2**. The lower panel shows how responses were simulated by mixing the output of the system described above with drifts and noise as described in the main text.

FIGURE 7

This is a plot of the simulated response with saturation against the equivalent response with no saturation. These simulated responses were obtained by setting the bilinear coupling parameter b_{33}^{1} labelled "neuronal saturation" in the previous figure to -0.4 and zero respectively. The key thing to observe is a saturation of responses at high levels. The broken line depicts the response expected in the absence of saturation. This illustrates how bilinear effects can introduce nonlinearities into the response.

FIGURE 8

Results summarising the conditional estimation based upon the simulated data of Figure 6. The upper panels show the conditional estimates and posterior probabilities pertaining to the intrinsic coupling parameters. The lower panels show the equivalent results for bilinear coupling parameters mediating the effect of within-epoch time. Conditional or MAP estimates of the parameters are shown in image format with arbitrary scaling. The posterior probabilities that these parameters exceeded a threshold of ln(2)/4 per sec. are shown as three-dimensional bar charts. True values and probabilities are shown on the left whereas the estimated values and posterior probabilities are shown on the right. This illustrates that the conditional estimates are a reasonable approximation to the true values

and, in particular, the posterior probabilities conform to the true probabilities, if we consider values of 90% or more.

FIGURE 9

These results are based upon the conditional or MAP estimates of the previous Figure. The upper panels show the implied first-order kernels for neuronal responses (upper-left) and equivalent hemodynamic responses (upper-right) as a function of peri-stimulus time for each of the three regions. The lower panels show the predicted response based upon the MAP estimators and a comparison of this response to the true response. The agreement is self-evident.

FIGURE 10

Posterior densities as a function of noise levels: The analysis, summarised in the previous two figures, was repeated for simulated data sequences at different levels of noise ranging from 0 to 2 units of standard deviation. Each graph shows the conditional expectation or MAP estimate of a coupling parameter (solid line) and the 90% confidence region (grey region). The true value for each parameter is also shown (broken line). The top row shows the temporal scaling parameter and the extrinsic connection between the first input and the first area. The middle row shows some intrinsic coupling parameters and the bottom row bilinear parameters. As anticipated the conditional variance of these estimators increases with noise, as reflected by a divergence of the confidence region with increasing standard deviation of the error.

FIGURE 11

Region selection for the empirical word processing example: Statistical Parametric Maps of the F ratio, based upon a conventional SPM analysis, are shown in the left panels and the spatial locations of the selected regions are shown on the right. These are superimposed on a T1-weighted reference image. The regional activities shown in the next Figure correspond to the first eigenvariates of a 4mm-radius sphere centred on the following coordinates in the standard anatomical space of Talairach and Tournoux. Primary auditory area **A1**; -50, -26, 8mm. Secondary auditory area **A2;** -64, -18, 2mm

and Wernicke's area **WA**; -56, -48, 6mm. In brief, we obtained fMRI time-series from a single subject at 2 Tesla using a Magnetom VISION (Siemens, Erlangen) whole body MRI system, equipped with a head volume coil. Contiguous multi-slice T2*-weighted fMRI images were obtained with a gradient echo-planar sequence using an axial slice orientation (TE = 40ms, TR = 1.7 seconds, 64x64x16 voxels). After discarding initial scans (to allow for magnetic saturation effects) each time-series comprised 1,200 volume images with 3mm isotropic voxels. The subject listened to monosyllabic or bisyllabic concrete nouns (i.e. 'dog', 'radio', 'mountain', 'gate') presented at 5 different rates (10 15 30 60 and 90 words per minute) for epochs of 34 seconds, intercalated with periods of rest. The 5 presentation rates were successively repeated according to a Latin Square design. The data were processed within SPM99 (Wellcome Department of Cognitive Neurology, http://www.fil.ion.ucl.ac.uk/spm). The time-series were realigned, corrected for movement-related effects and spatially normalised. The data were smoothed with a 5mm isotropic Gaussian kernel. The SPM{F} above was based on a standard regression model using word presentation rate as the stimulus function and convolving it with a canonical hemodynamic response and its temporal derivative to form regressors.

FIGURE 12

Results of a DCM analysis applied to the data described in the previous Figure. The display format follows that of Figure 6. The coupling parameters are shown alongside the corresponding connections. The values in brackets are the percentage confidence that these values exceed a threshold of ln(2)/8 per sec..

FIGURE 13

This Figure provides a more detailed characterisation of the conditional estimates. The images in the top-row are the MAP estimates for the intrinsic and bilinear coupling parameters, pertaining to saturation and adaptation. The middle panel shows the posterior density of a contrast of all bilinear terms mediating adaptation, namely the modulation of intrinsic connections by the second time-dependent experimental effect. The predicted responses based upon the conditional estimators are shown for each of the three regions on the lower left (solid lines) with the original data (dots) after removal of

confounds. A re-parameterisation of the conditional estimates, in terms of the first-order kernels, is shown on the lower right. The hemodynamic (left) and neuronal (right) kernels should be compared with the equivalent kernels for the simulated data in Figure 9.

FIGURE 14

Results of the reproducibility analyses: a) Results for the intrinsic parameters. The profile of conditional estimates for the 10 independent analyses described in the main text are shown in image format, all scaled to the maximum. The posterior densities, upon which these estimates are based, are shown for two selected connections in the lower two graphs. These densities are displayed in terms of their expectation and 90% confidence intervals (grey bars) for the forward connection from **A1** to **A2**. The equivalent densities are shown for the backward connection from **WA** to **A2**. Although the posterior probability that the latter connections exceeded the specified threshold was less than 90%, it can be seen that this connection is almost certainly greater than zero. b) Equivalent results for the bilinear coupling matrices mediating adaptation. The lower panels here refer to the posterior densities of a contrast testing for the mean of all bilinear parameters (left) and the extrinsic connection to **A1** (right).

FIGURE 15

ReML hyperparameter variance estimates for each region and analysis: These estimates provide an anecdotal characterisation of the within- and between-area variability, in hyperparameter estimates, and show that they generally lie between 0.8 and 1 (adimensional units corresponding to % whole brain mean).

FIGURE 16

Results of the empirical analysis of the attention study. a) Functional architecture based upon the conditional estimates displayed using the same format as Figure 12. The most interesting aspects of this architecture involved 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 latent influence of **SPC** on the **V5**. Dotted arrows connecting regions represent significant bilinear affects in the absence of a significant intrinsic coupling. b) Fitted responses based upon the conditional estimates and the adjusted data are shown using the same format as in Figure 13. The insert shows the location of the regions, again adopting the same format in previous Figures. The location of these regions centred on the primary visual cortex **V1**; 6, -84, -6mm: motion sensitive area **V5;** 45, -81, 5mm. Superior parietal cortex, **SPC**; 18, -57, 66mm. Inferior frontal gyrus, **IFG**, 54, 18, 30mm. The volumes from which the first eigenvariates were calculated corresponded to 8mm radius spheres centred on these locations.

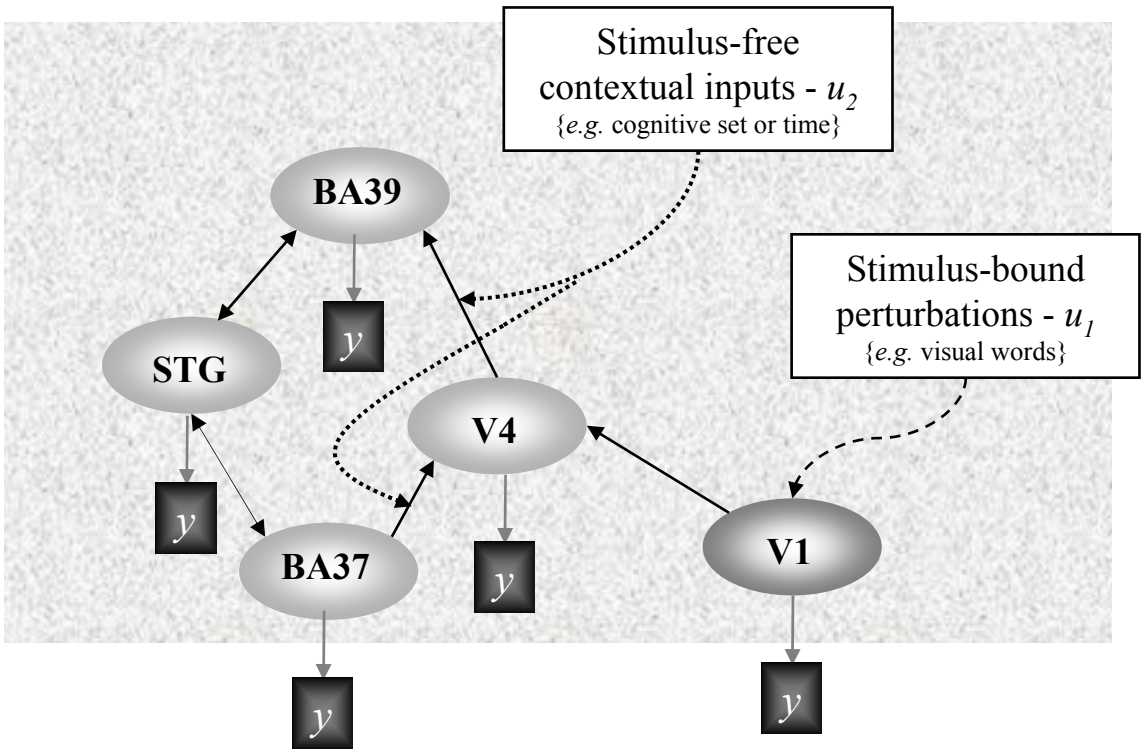
Subjects were studied with fMRI under identical stimulus conditions (visual motion subtended by radially moving dots) whilst manipulating the attentional component of the task (detection of velocity changes). The data were acquired from normal subjects at 2 Tesla using a Magnetom VISION (Siemens, Erlangen) whole body MRI system, equipped with a head volume coil. Here we analyse data from the first subject. Contiguous multi-slice T2*-weighted fMRI images were obtained with a gradient echo-planar sequence (TE = 40ms, TR = 3.22 seconds, matrix size = 64x64x32, voxel size 3x3x3mm). Each subject had 4 consecutive 100-scan sessions comprising a series of 10-scan blocks under 5 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 centre of a screen. In condition A (Attention) subjects viewed 250 dots moving radially from the centre at 4.7 degrees 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 centre of the screen. In a pre-scanning session the subjects were given 5 trials with 5 speed changes (reducing to 1%). During scanning there were no speed changes. No overt response was required in any condition.

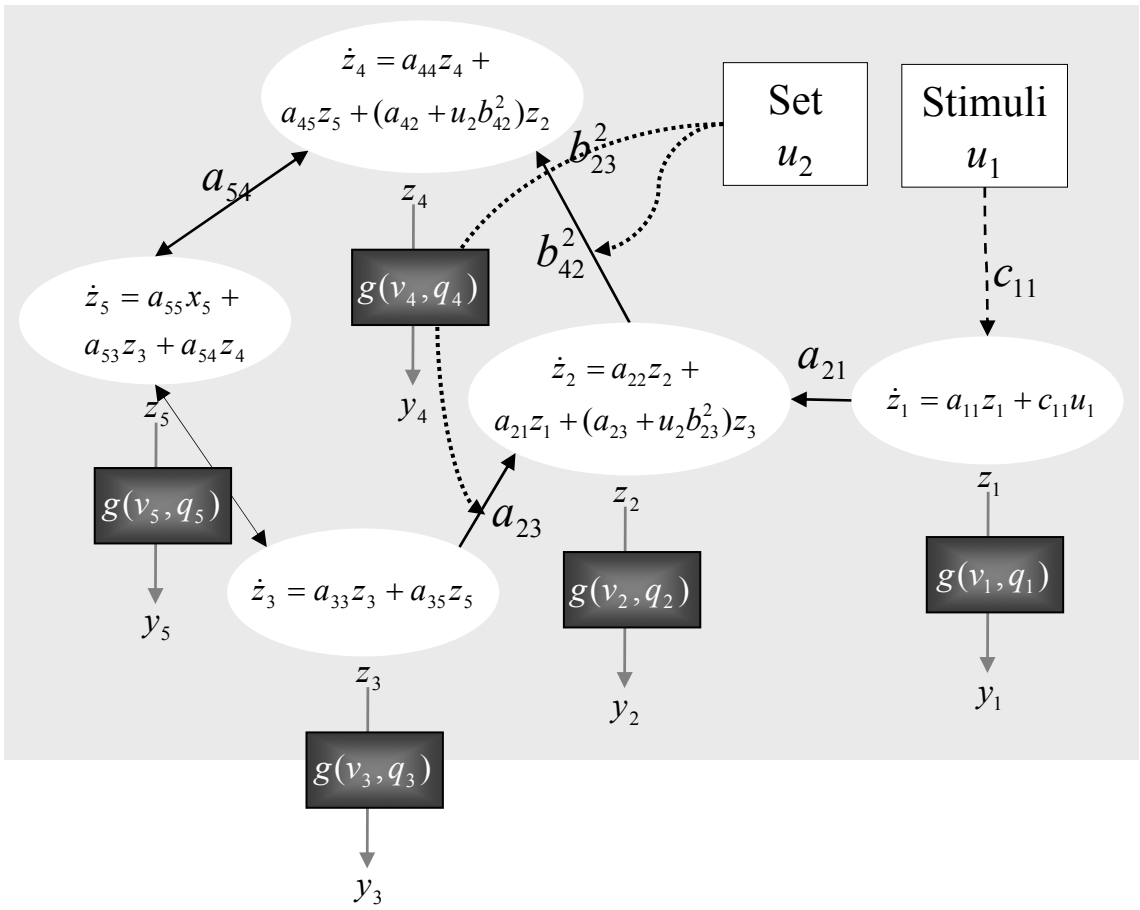
Table 1

Priors on biophysical parameters

Parameter	**Description**	**Prior mean** η_θ	**Prior variance** C_θ
κ	rate of signal decay	0.65 per sec	0.015
γ	rate of flow-dependent elimination	0.41 per sec	0.002
τ	hemodynamic transit time	0.98 sec	0.0568
α	Grubb's exponent	0.32	0.0015
ρ	resting oxygen extraction fraction	0.34	0.0024

Functional integration and the modulation of specific pathways





latent connectivity induced connectivity

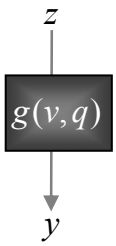
$$\begin{bmatrix} \dot{z}_1 \\ \vdots \\ \dot{z}_5 \end{bmatrix} = \left\{ \begin{bmatrix} a_{11} & \cdots & 0 \\ a_{21} \; a_{22} \; a_{23} & & \\ \vdots \quad\; a_{33} & & a_{35} \\ \quad a_{42} & a_{44} & a_{45} \\ 0 \quad \cdots \; a_{53} & a_{54} & a_{55} \end{bmatrix} + u_2 \begin{bmatrix} 0 & \cdots & 0 \\ & b_{23}^2 & \\ \vdots & \ddots & \vdots \\ b_{42}^2 & & \\ 0 & \cdots & 0 \end{bmatrix} \right\} \begin{bmatrix} z_1 \\ \vdots \\ z_5 \end{bmatrix} + \begin{bmatrix} c_{11} & 0 \\ \vdots & \vdots \\ 0 & 0 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

Forward, backward & self

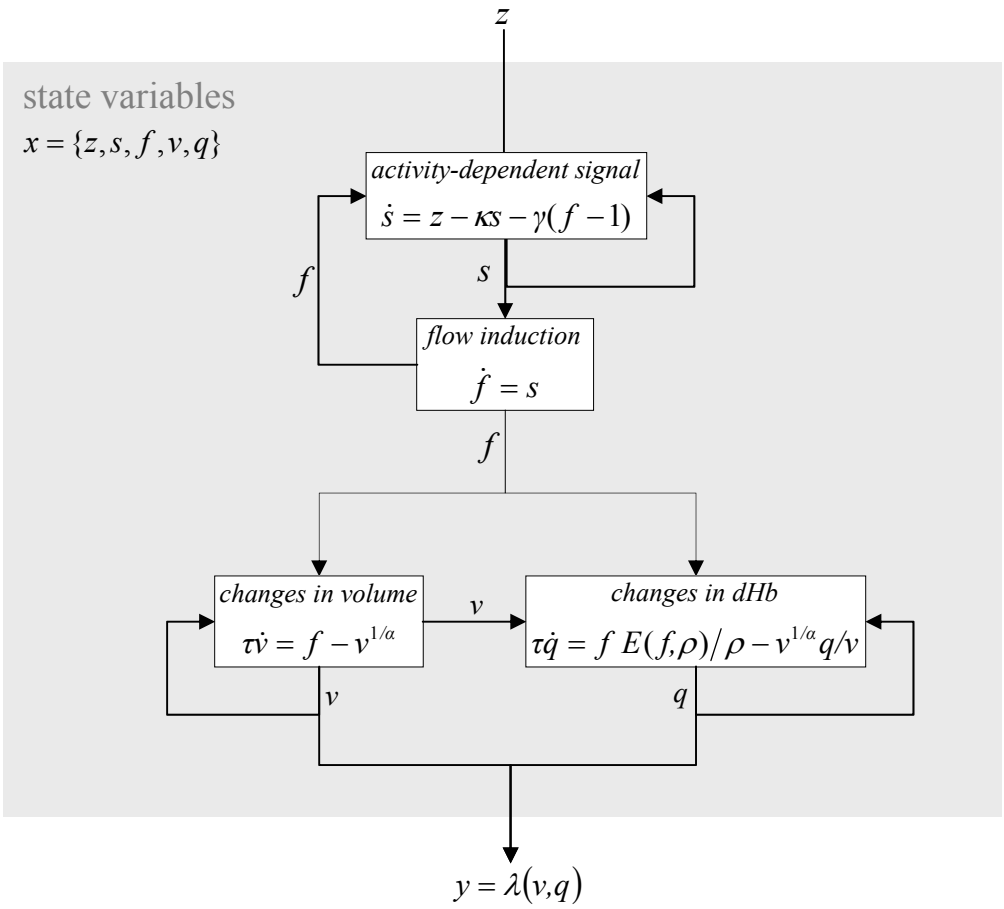
$$\dot{z} = (A + \sum_j u_j B^j)z + Cu$$

The bilinear model

The hemodynamic model

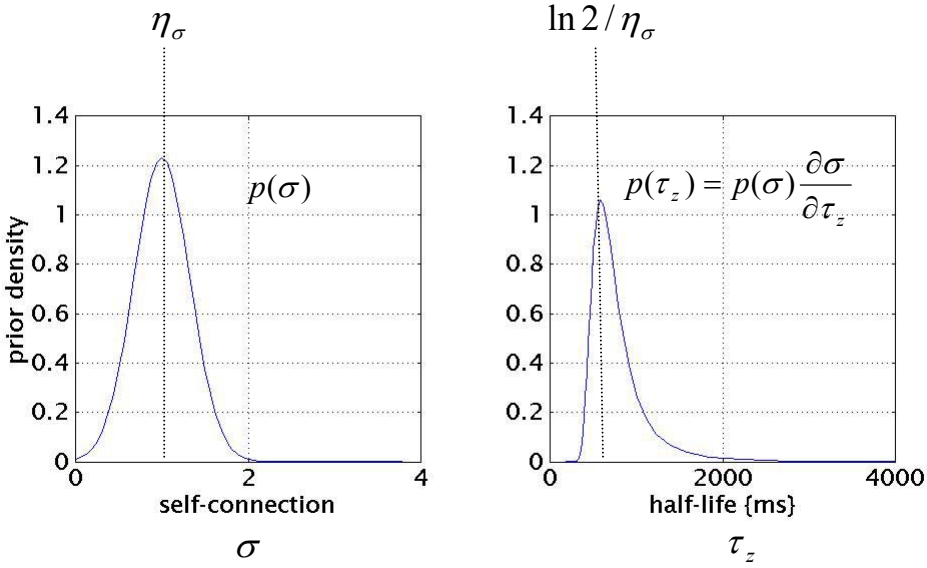


neuronal input

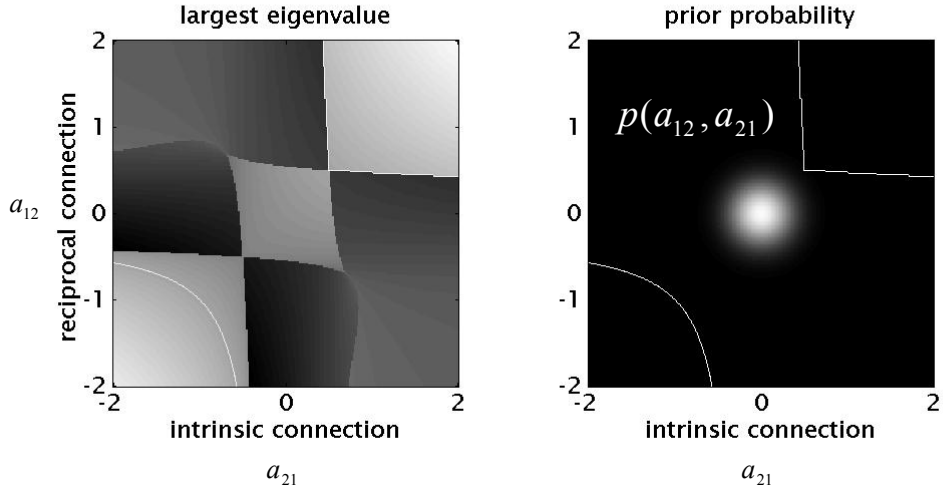


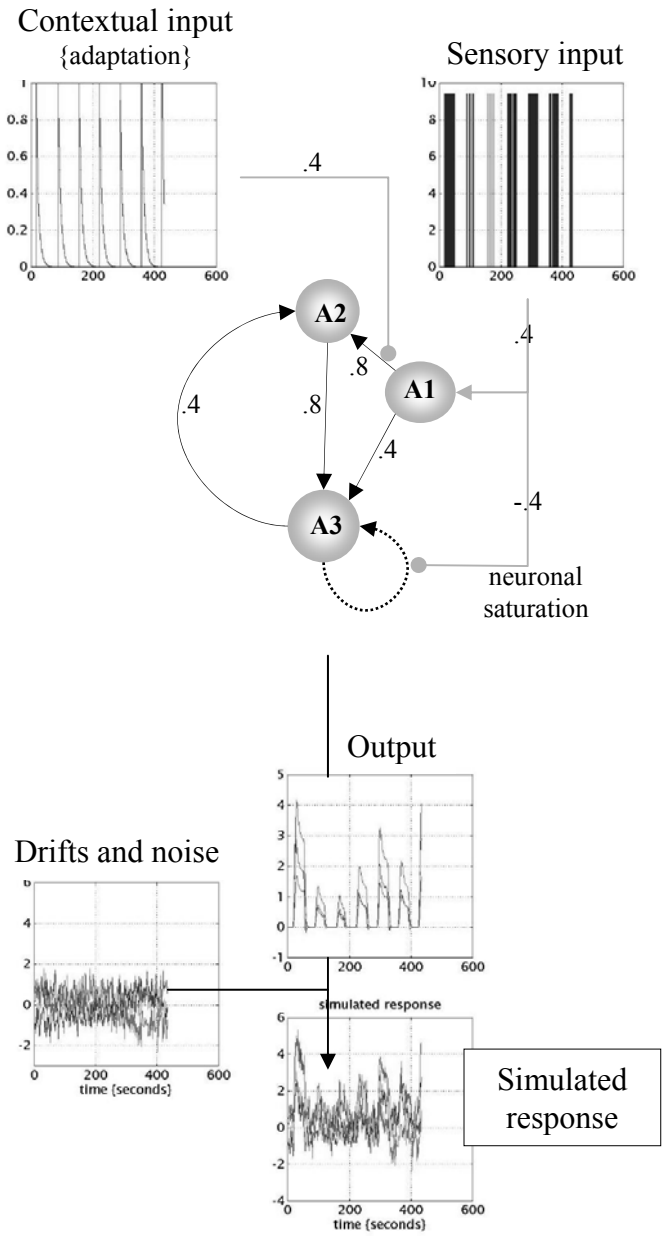
hemodynamic response

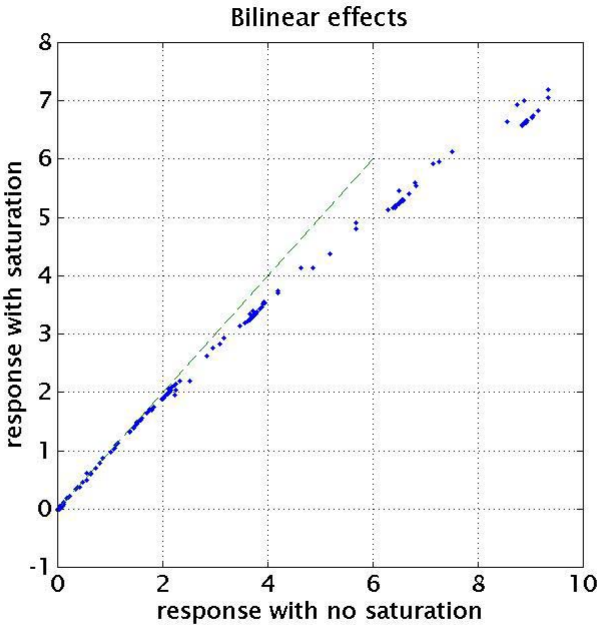
Self connections and temporal scaling



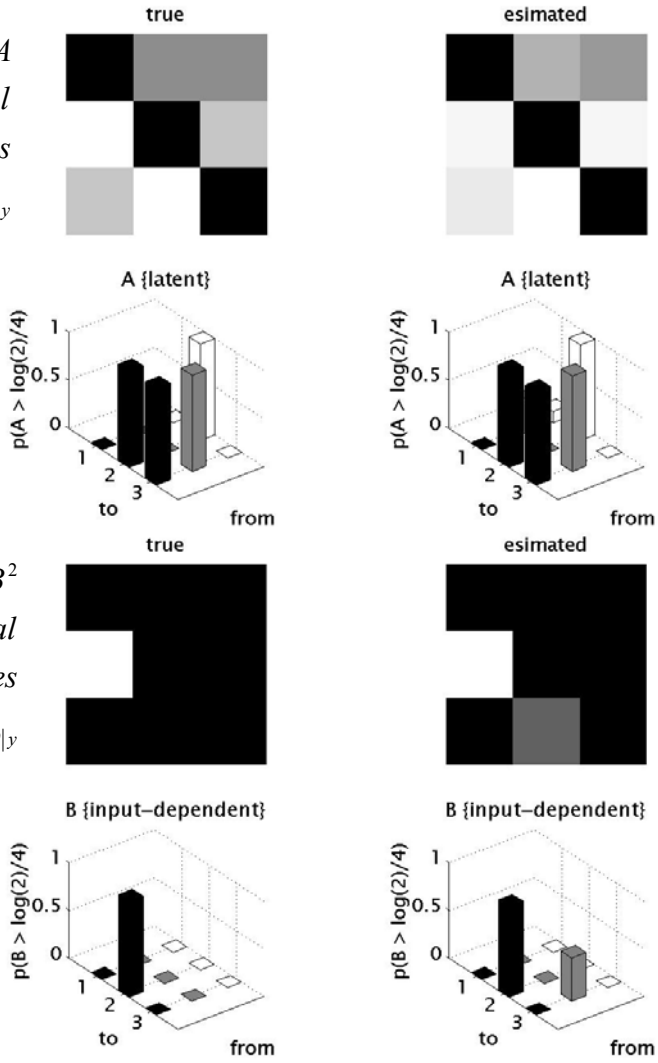
$$A = \begin{bmatrix} -1 & a_{12} & \frac{1}{2} \\ a_{21} & -1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -1 \end{bmatrix}$$



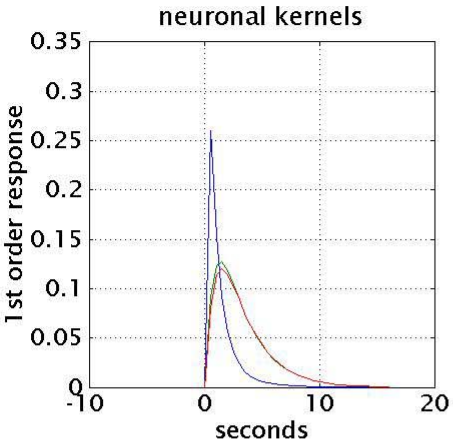

Contextual input
{adaptation}
Sensory input
.4
A2
.8
A1
.4
.8
.4
.4
A3
-.4
neuronal
saturation
Output
Drifts and noise
simulated response
time {seconds}
time {seconds}
Simulated
response

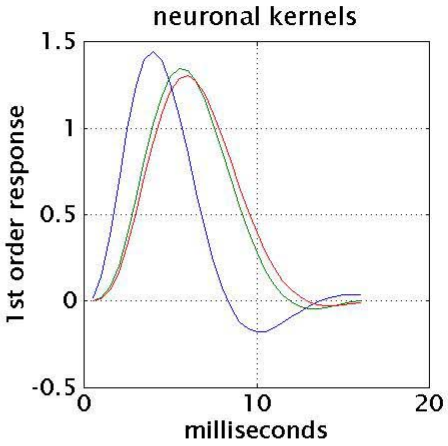

Bilinear effects
response with saturation
response with no saturation
8
7
6
5
4
3
2
1
0
-1
0
2
4
6
8
10

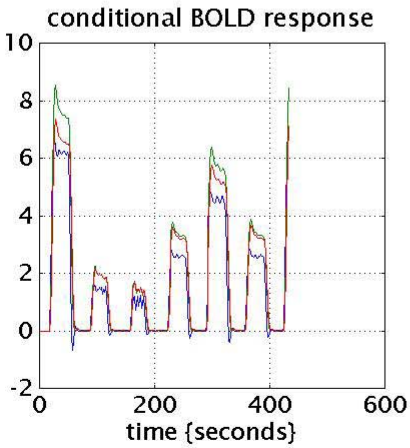
Values of A conditional estimates $\eta_{\theta|y}$

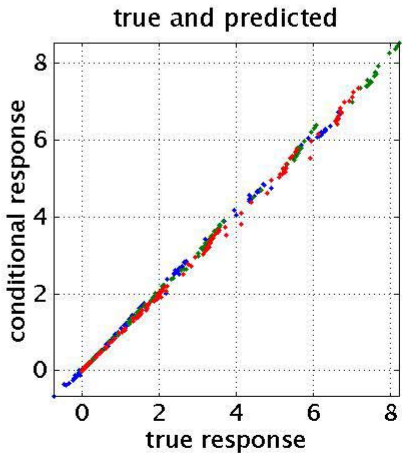


Values of B^2 conditional estimates $\eta_{\theta|y}$

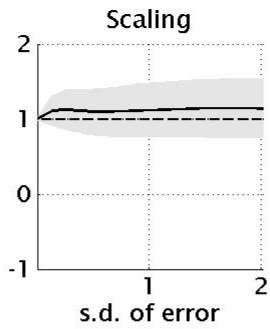

neuronal kernels
1st order response
seconds

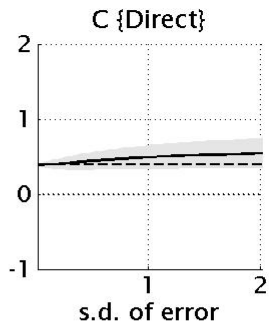

neuronal kernels
1st order response
milliseconds

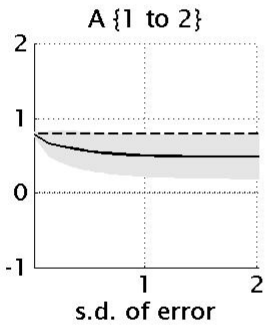

conditional BOLD response
time {seconds}

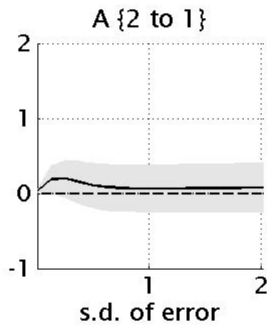

true and predicted
conditional response
true response

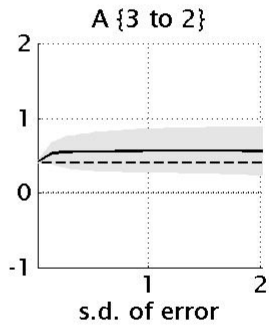
The effects of noise

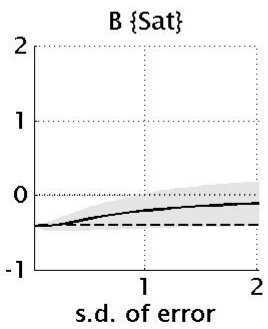


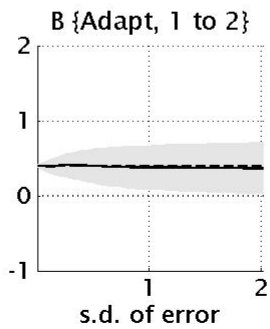


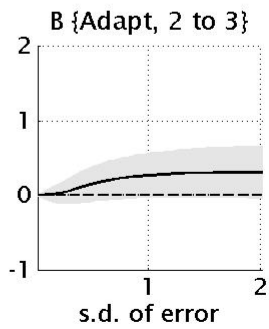




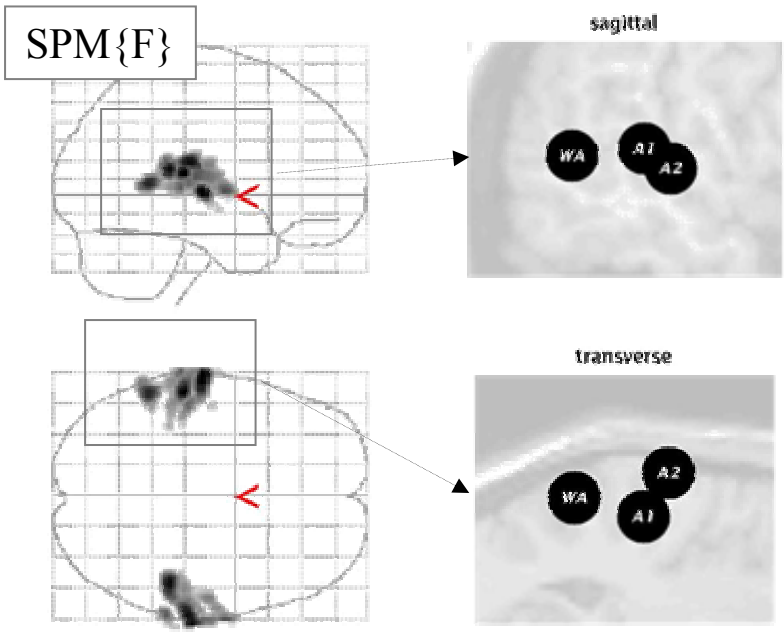




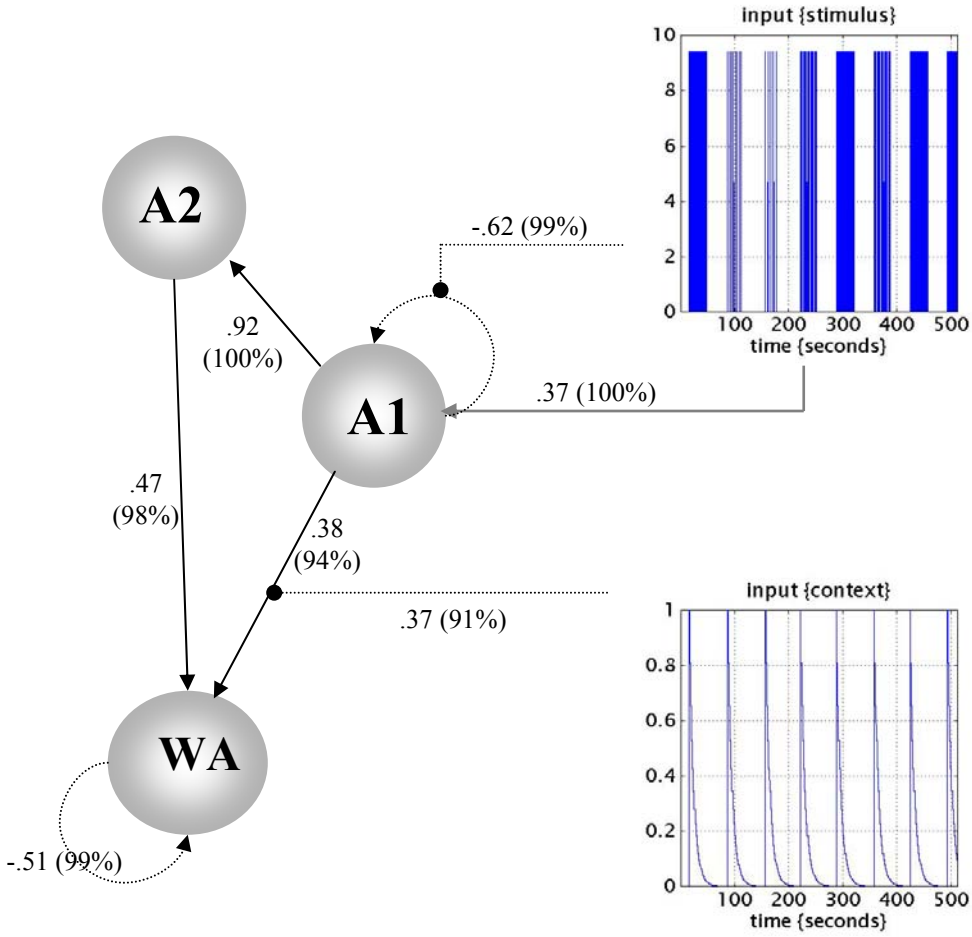




An empirical example:
single word processing at different rates

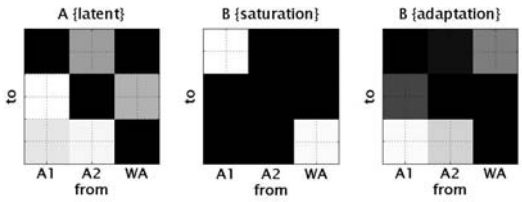


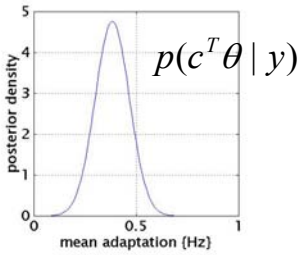
Estimated architecture



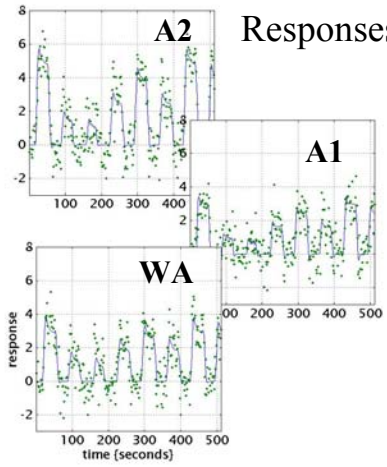
a

MAP estimates $\eta_{\theta|y}$



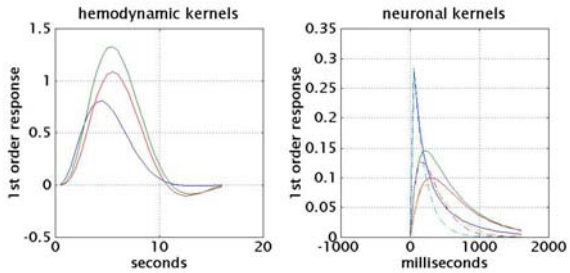


b



Responses

Kernels



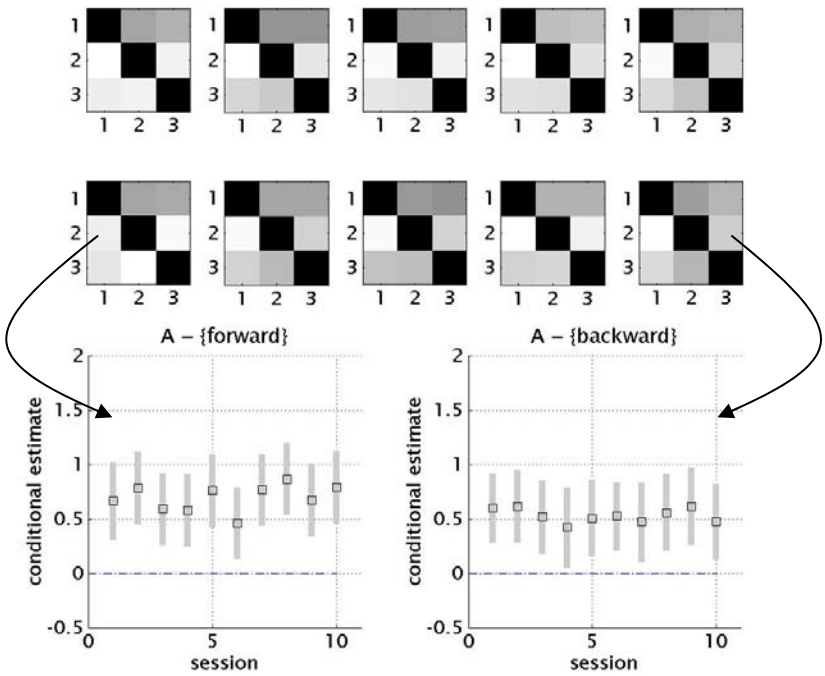
Reproducibility

a

Intrinsic connections (A)

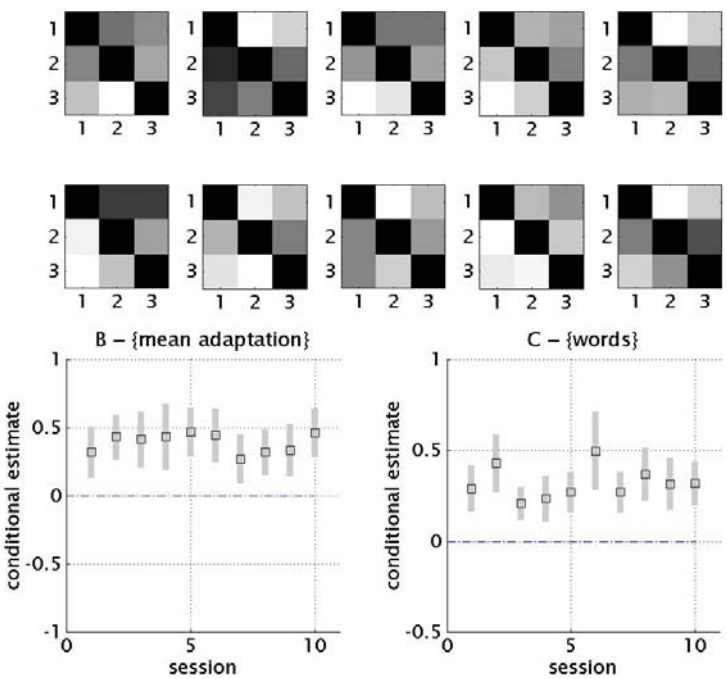
conditional estimates $\eta_{\theta|y}$

contrasts $c^T \eta_{\theta|y}$

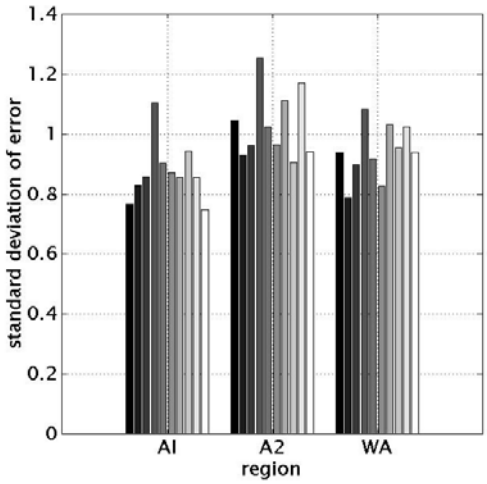


b

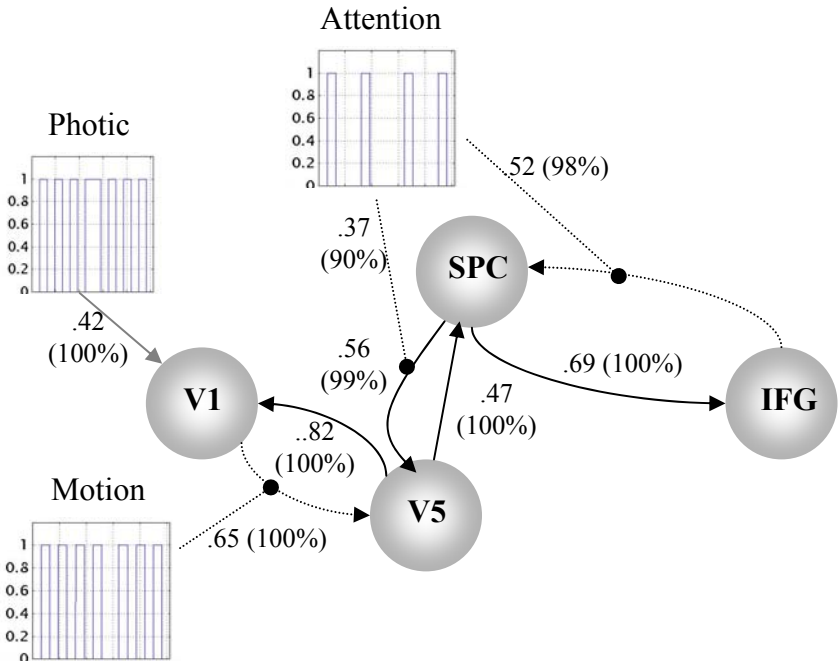
Adaptation (B^2)



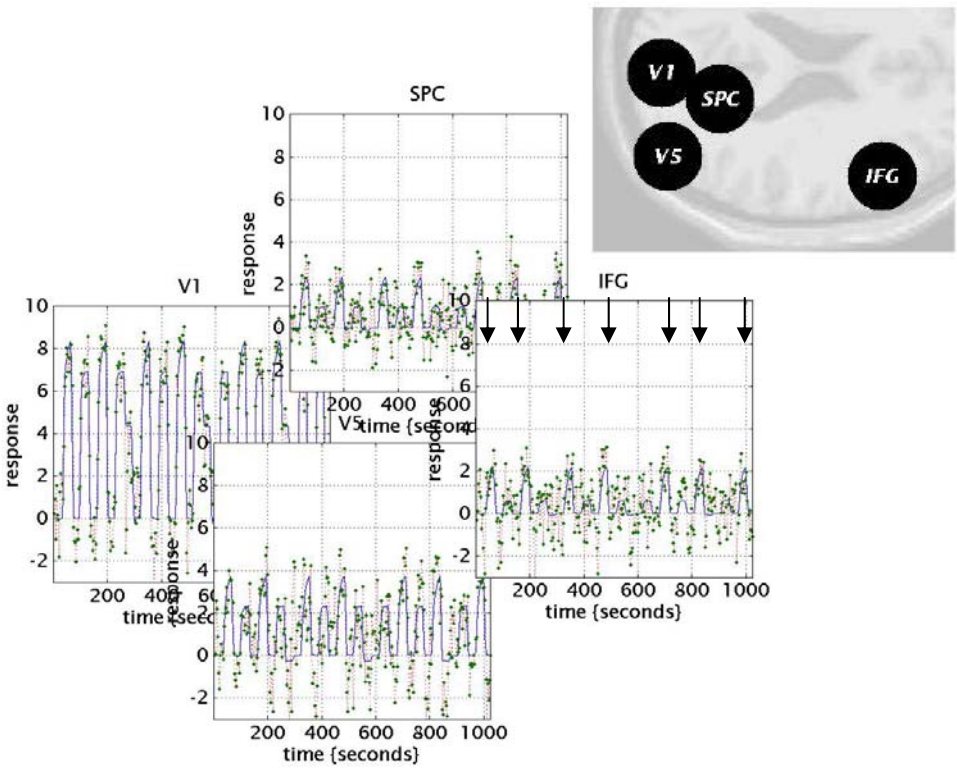
ReML error
variance estimates



a


Attention
Photic
Motion
.52 (98%)
.37
(90%)
SPC
.42
(100%)
.56
(99%)
.69 (100%)
V1
.47
(100%)
IFG
..82
(100%)
V5
.65 (100%)

b


V1
SPC
V5
IFG
response
time {seconds}