Dynamic causal modeling of evoked responses in EEG and MEG.

Olivier David, Stefan J. Kiebel, Lee M. Harrison, Jérémie Mattout, James M. Kilner, Karl J. Friston

Dynamic Causal Modeling of Evoked Responses in EEG and MEG

Olivier David,* Stefan J. Kiebel, Lee M Harrison, Jérémie Mattout, James M. Kilner, Karl J. Friston

Wellcome Department of Imaging Neuroscience, Functional Imaging Laboratory, 12 Queen Square, London WC1N 3BG, UK

*Present address: INSERM U594 Neuroimagerie Fonctionnelle et Métabolique, Université Joseph Fourier, CHU – Pavillon B – BP 217, 38043 Grenoble Cedex 09, France

Corresponding author:

Stefan Kiebel

The Wellcome Dept. of Imaging Neuroscience,

Institute of Neurology, UCL

12 Queen Square, London, UK WC1N 3BG

Tel (44) 207 833 7478

Fax (44) 207 813 1420

Email skiebel@fil.ion.ucl.ac.uk

Abstract

Neuronally plausible, generative or forward models are essential for understanding how event-related fields (ERFs) and potentials (ERPs) are generated. In this paper we present a new approach to modeling event-related responses measured with EEG or MEG. This approach uses a biologically informed model to make inferences about the underlying neuronal networks generating responses. The approach can be regarded as a neurobiologically constrained source reconstruction scheme, in which the parameters of the reconstruction have an explicit neuronal interpretation. Specifically, these parameters encode, among other things, the coupling among sources and how that coupling depends upon stimulus attributes or experimental context. The basic idea is to supplement conventional electromagnetic forward models, of how sources are expressed in measurement space, with a model of how source activity is generated by neuronal dynamics. A single inversion of this extended forward model enables inference about both the spatial deployment of sources and the underlying neuronal architecture generating them. Critically, this inference covers long-range connections among well-defined neuronal subpopulations.

In a previous paper, we simulated ERPs using a hierarchical neural-mass model that embodied bottom-up, top-down and lateral connections among remote regions. In this paper, we describe a Bayesian procedure to estimate the parameters of this model using empirical data. We demonstrate this procedure by characterizing the role of changes in cortico-cortical coupling, in the genesis of ERPs. In the first experiment, ERPs recorded during the perception of faces and houses were modeled as distinct cortical sources in the ventral visual pathway. Category-selectivity, as indexed by the face-selective N170, could be explained by category-specific differences in forward connections from sensory to higher areas in the ventral stream. We were able to quantify and make inferences about these effects using conditional estimates of connectivity. This allowed us to identify where, in the processing stream, category-selectivity emerged.

In the second experiment we used an auditory oddball paradigm to show the mismatch negativity can be explained by changes in connectivity. Specifically, using Bayesian model selection, we assessed changes in backward connections, above and beyond changes in forward connections. In accord with theoretical predictions, there was strong evidence for learning-related changes in both forward and backward coupling. These examples show that category- or context-specific coupling among cortical regions can be assessed explicitly, within a mechanistic, biologically motivated inference framework.

Keywords: Electroencephalography, magnetoencephalography, neural networks, nonlinear dynamics, causal modeling, and Bayesian inference.

Introduction

Event-related fields (ERFs) and potentials (ERPs) have been used for decades as putative magneto- and electrophysiological correlates of perceptual and cognitive operations. However, the exact neurobiological mechanisms underlying their generation are largely unknown. Previous studies have shown that ERP-like responses can be reproduced by brief perturbations of model cortical networks (David *et al*., 2005; Jansen and Rit, 1995; Rennie *et al*., 2002). The goal of this paper was to demonstrate that biologically plausible dynamic causal models (DCMs) can explain empirical ERP phenomena. In particular, we show that changes in connectivity, among distinct cortical sources, are sufficient to explain stimulus- or set-specific ERP differences. Adopting explicit neuronal models, as an explanation of observed data, may afford a better understanding of the processes underlying event-related responses in magnetoencephalography (MEG) and electroencephalography (EEG).

Functional vs. effective connectivity

The aim of dynamic causal modeling (Friston *et al*., 2003) is to make inferences about the coupling among brain regions or sources and how that coupling is influenced by experimental factors. DCM uses the notion of *effective connectivity*, defined as the influence one neuronal system exerts over another. DCM represents a fundamental departure from existing approaches to connectivity because it employs an explicit generative model of measured brain responses that embraces their nonlinear causal architecture. The alternative to causal modeling is to simply establish statistical dependencies between activity in one brain region and another. This is referred to as *functional connectivity*. Functional connectivity is useful because it rests on an operational definition and eschews any arguments about how dependencies are caused. Most approaches in the EEG and MEG literature address functional connectivity, with a focus on dependencies that are expressed at a particular frequency of oscillations (*i.e.* coherence). See Schnitzler and Gross (2005) for a nice review. Recent advances have looked at nonlinear or generalized synchronization in the context of chaotic oscillators (*e.g.* Rosenblum *et al* 2002) and stimulus-locked responses of coupled oscillators (see Tass 2004). These characterizations often refer to phase-synchronization as a useful measure of nonlinear dependency. Another exciting development is the reformulation of coherence in terms of autoregressive models. A compelling example is reported in Brovelli *et al* (2004) who were able show that "synchronized beta oscillations bind multiple sensorimotor areas into a large-scale network during motor maintenance behavior and carry Granger causal influences from primary somatosensory and inferior posterior parietal cortices to motor cortex." Similar developments have been seen in functional neuroimaging with fMRI (*e.g.* Harrison *et al.,* 2003; *Roebroeck et al.*, 2005).

These approaches generally entail a two-stage procedure. First an electromagnetic forward model is inverted to estimate the activity of sources in the brain. Then, a *post-hoc* analysis is used to establish statistical dependencies (*i.e.* functional connectivity) using coherence, phase-synchronization, Granger influences or related analyses such as (linear} directed transfer functions and (nonlinear) generalized synchrony. DCM takes a very different approach and uses a forward model that explicitly includes long-range connections among neuronal sub-populations underlying measured sources. A single Bayesian inversion allows one to infer on parameters of the model (*i.e.* effective connectivity) that mediate functional

connectivity. This is like performing a biological informed source reconstruction with the added constraint that the activity in one source has to be caused by activity in other, in a biologically plausible fashion. This approach is much closer in sprit to the work of Robinson *et al* (2004) who show that "model-based electroencephalographic (EEG) methods can quantify neurophysiologic parameters that underlie EEG generation in ways that are complementary to and consistent with standard physiologic techniques." DCM also speaks to the interest in neuronal modeling of ERPs in specific systems. See for example Melcher and Kiang (1996), who evaluate a detailed cellular model of brainstem auditory evoked potentials (BAEP) and conclude "it should now be possible to relate activity in specific cell populations to psychophysical performance since the BAEP can be recorded in behaving humans and animals." See also Dau *et al* (2003). Although the models presented in this paper are more generic than those invoked to explain the BAEP, they share the same ambition of understanding the mechanisms of response generation and move away from phenomenological or descriptive quantitative EEG measures.

Dynamic causal modeling

The central idea behind DCM is to treat the brain as a deterministic nonlinear dynamical system that is subject to inputs, and produces outputs. Effective connectivity is parameterized in terms of coupling among unobserved brain states, *i.e.* neuronal activity in different regions. Coupling is estimated by perturbing the system and measuring the response. This is in contradistinction to established methods for estimating effective connectivity from neurophysiological time series, which include structural equation modeling and models based on multivariate autoregressive processes (Büchel and Friston, 1997; Harrison *et al.,* 2003; McIntosh and Gonzalez-Lima, 1994). In these models, there is no designed perturbation and the inputs are treated as unknown and stochastic. Although the principal aim of DCM is to explain responses in terms of context-dependent coupling, it can also be viewed as a biologically informed inverse solution to the source reconstruction problem. This is because estimating the parameters of a DCM rests on estimating the hidden states of the modeled system. In ERP studies, these states correspond to the activity of the sources that comprise the model. In addition to biophysical and coupling parameters the DCMs parameters cover the spatial expression of sources at the sensor level. This means that inverting the DCM entails a simultaneous reconstruction of the source configuration and their dynamics.

Because DCMs are not restricted to linear or instantaneous systems they generally depend on a large number of free parameters. However, because it is biologically grounded, parameter estimation is constrained. A natural way to embody these constraints is within a Bayesian framework. Consequently, DCMs are estimated using Bayesian inversion and inferences about particular connections are made using their posterior or conditional density. DCM has been previously validated with functional magnetic resonance imaging (fMRI) time series (Friston *et al.,* 2003; Riera *et al.,* 2004). fMRI responses depend on hemodynamic processes that effectively low-pass filter neuronal dynamics. However, with ERPs this is not the case and there is sufficient information, in the temporal structure of evoked responses, to enable precise conditional identification of quite complicated DCMs. In this study, we use a model described recently (David *et al*., 2005) that embeds cortical sources, with several source-specific neuronal subpopulations, into hierarchical cortico-cortical networks.

This paper is structured as follows. In the theory section we review the neural mass model used to generate MEG/EEG-like evoked responses. This section summarizes David *et al*. (2005) in which more details about the generative model and associated dynamics can be found. The next section provides a brief review of Bayesian estimation, conditional inference and model comparison that are illustrated in the subsequent section. An empirical section then demonstrates the use of DCM for ERPs by looking at changes in connectivity that were induced, either by category-selective activation of different pathways in the visual system, or by sensory learning in an auditory oddball paradigm. This section concludes with simulations that demonstrate the face validity of the particular DCMs employed. Details about how the empirical data were acquired and processed will be found in an appendix.

THEORY

Intuitively, the DCM scheme regards an experiment as a designed perturbation of neuronal dynamics that are promulgated and distributed throughout a system of coupled anatomical nodes or sources to produce region-specific responses. This system is modeled using a dynamic input–state–output system with multiple inputs and outputs. Responses are evoked by deterministic inputs that correspond to experimental manipulations (*i.e.* presentation of stimuli). Experimental factors (*i.e.* stimulus attributes or context) can also change the parameters or causal architecture of the system producing these responses. The state variables cover both the neuronal activities and other neurophysiological or biophysical variables needed to form the outputs. In our case, outputs are those components of neuronal responses that can be detected by MEG/EEG sensors.

In neuroimaging, DCM starts with a reasonably realistic neuronal model of interacting cortical regions. This model is then supplemented with a forward model of how neuronal activity is transformed into measured responses; here MEG/EEG scalp averaged responses. This enables the parameters of the neuronal model (*i.e.*, effective connectivity) to be estimated from observed data. For MEG/EEG data, the supplementary model is a forward model of electromagnetic measurements that accounts for volume conduction effects (Mosher *et al*., 1999). We first review the neuronal component of the forward model and then turn to the modality-specific measurement model.

A neural mass model

The majority of neural mass models of MEG/EEG dynamics have been designed to generate spontaneous rhythms (David and Friston, 2003; Jansen and Rit, 1995; Lopes da Silva *et al*., 1974; Robinson *et al.,* 2001; Stam *et al.*, 1999) and epileptic activity (Wendling *et al.,* 2002). These models use a small number of state variables to represent the expected state of large neuronal populations, *i.e.* the neural mass. To date, event-related responses of neural mass models have received less attention (David *et al*., 2005; Jansen and Rit, 1995; Rennie *et al*., 2002). Only recent models have embedded basic anatomical principles that underlie extrinsic connections among neuronal populations:

The cortex has a hierarchical organization (Crick and Koch, 1998; Felleman and Van Essen, 1991), comprising forward, backward and lateral processes that can be understood from an anatomical and cognitive perspective (Engel *et al*., 2001). The direction of an anatomical projection is usually inferred from the laminar pattern of its origin and termination.

We have developed a hierarchical cortical model to study the genesis of ERFs/ERPs (David *et al.,* 2005). This model is used here as a DCM. The neuronal part of the DCM comprises a network or graph of sources. In brief, each source is modeled with three neuronal subpopulations. These subpopulations are interconnected with intrinsic connections within each source. The sources are interconnected by extrinsic connections among specific subpopulations. The specific source and target subpopulations define the connection as forward, backward or lateral. The model is now reviewed in terms of the differential equations that embody its causal architecture.

Neuronal state equations

The model (David *et al*., 2005) embodies directed extrinsic connections among a number of sources, each based on the Jansen model (Jansen and Rit, 1995), using the connectivity rules described in (Felleman and Van Essen, 1991). These rules, which rest on a tri-partitioning of the cortical sheet into supra-, infra-granular layers and granular layer 4, have been derived from experimental studies of monkey visual cortex. We assume these rules generalize to other cortical regions (but see Smith and Poplin, 2001 for a comparison of primary visual and auditory cortices). Under these simplifying assumptions, directed connections can be classified as: (i) Bottom-up or forward connections that originate in agranular layers and terminate in layer 4. (ii) Top-down or backward connections that connect agranular layers. (iii) Lateral connections that originate in agranular layers and target all layers. These long-range or extrinsic cortico-cortical connections are excitatory and comprise the axonal processes of pyramidal cells.

The Jansen model (Jansen and Rit, 1995) emulates the MEG/EEG activity of a cortical source using three neuronal subpopulations. A population of excitatory pyramidal (output) cells receives inputs from inhibitory and excitatory populations of interneurons, via intrinsic connections (intrinsic connections are confined to the cortical sheet). Within this model, excitatory interneurons can be regarded as spiny stellate cells found predominantly in layer 4. These cells receive forward connections. Excitatory pyramidal cells and inhibitory interneurons occupy agranular layers and receive backward and lateral inputs. Using these connection rules, it is straightforward to construct any hierarchical cortico-cortical network model of cortical sources. See Figure 1.

Figure 1 about here

The ensuing DCM is specified in terms of its state equations and an observer or output equation

$$\begin{aligned} \dot{x} &= f(x,u,\theta) \\ h &= g(x,\theta) \end{aligned} \qquad 1$$

where *x* are the neuronal states of cortical areas, *u* are exogenous inputs and *h* is the output of the system. θ are quantities that parameterize the state and observer equations (see also below under 'Prior assumptions'). The state equations $f(x,u,\theta)$ (Jansen and Rit 1995; David and Friston 2003; David *et al* 2005) for the neuronal states of multiple areas are[1]

$$
\begin{aligned}
\dot{x}_7 &= x_8 \\
\dot{x}_8 &= \frac{H_e}{\tau_e}((C^B + C^L + \gamma_3 I)S(x_0)) - \frac{2x_8}{\tau_e} - \frac{x_7}{\tau_e^2} \\
\dot{x}_1 &= x_4 \\
\dot{x}_4 &= \frac{H_e}{\tau_e}((C^F + C^L + \gamma_1 I)S(x_0) + C^U u) - \frac{2x_4}{\tau_e} - \frac{x_1}{\tau_e^2} \\
\dot{x}_0 &= x_5 - x_6 \\
\dot{x}_2 &= x_5 \\
\dot{x}_5 &= \frac{H_e}{\tau_e}((C^B + C^L)S(x_0) + \gamma_2 S(x_1)) - \frac{2x_5}{\tau_e} - \frac{x_2}{\tau_e^2} \\
\dot{x}_3 &= x_6 \\
\dot{x}_6 &= \frac{H_i}{\tau_i}\gamma_4 S(x_7) - \frac{2x_6}{\tau_i} - \frac{x_3}{\tau_i^2}
\end{aligned}
\qquad 2
$$

where $x_j = [x_j^{(1)}, x_j^{(2)}, \ldots]^T$ . . The states $x_0^{(i)}, \ldots, x_8^{(i)}$ represent the mean transmembrane potentials and currents of the three subpopulations in the *i*-th source. The state equations specify the rate of change of voltage as a function of current and specify how currents change as a function of voltage and current. Figure 1 depicts the states by assigning each subpopulation to a cortical layer. For schematic reasons we have lumped superficial and deep pyramidal units together, in the infra-granular layer. The matrices C^F, C^B, C^L encode forward, backward and lateral extrinsic connections respectively. From Eq.2 and Figure 1 it can be seen that the state equations embody the connection rules above. For example, extrinsic connections mediating changes in mean excitatory [depolarizing] current x_8, in the supragranular layer, are restricted to backward and lateral connections. The depolarisation of pyramidal cells $x_0 = x_2 - x_3$ represents a mixture of potentials induced by excitatory and inhibitory [depolarizing and hyperpolarizing] currents respectively. This pyramidal potential is the presumed source of observed MEG/EEG signals.

The remaining constants in the state equation pertain to two operators, on which the dynamics rest. The first transforms the average density of pre-synaptic inputs into the average postsynaptic membrane potential. This transformation is equivalent to a convolution with an impulse response or kernel,

[1] Propagation delays Δ on the connections have been omitted for clarity, here and in Figure 1. See Appendix A.1 for details of how delays are incorporated.

$$p(t)_e = \begin{cases} \frac{H_e}{\tau_e} t \exp(-t/\tau_e) & t \geq 0 \\ 0 & t < 0 \end{cases} \quad (3)$$

where subscript "e" stands for "excitatory". Similarly, the subscript "i" is used for inhibitory synapses. H controls the maximum post-synaptic potential and τ represents a lumped rate constant. The second operator *S* transforms the potential of each subpopulation into firing rate, which is the input to other subpopulations. This operator is assumed to be an instantaneous sigmoid nonlinearity

$$S(x) = \frac{1}{1+\exp(-rx)} - \frac{1}{2} \quad (4)$$

where $r = .56$ determines its form. Figure 2 shows examples of these synaptic kernels and sigmoid functions. Interactions, among the subpopulations, depend on the constants $\gamma_{1,2,3,4}$, which control the strength of intrinsic connections and reflect the total number of synapses expressed by each subpopulation. A DCM, at the neuronal level, obtains by coupling sources with extrinsic connections as described above. A typical three-source DCM is shown in Figure 3. See David and Friston 2003 and David *et al* 2005 for further details.

Figure 2 about here

Event-related input and ERP-specific effects

To model event-related responses, the network receives inputs via input connections C^U. These connections are exactly the same as forward connections and deliver inputs *u* to the spiny stellate cells in layer 4. In the present context, inputs *u* model afferent activity relayed by subcortical structures and is modeled with two components.

$$u(t) = b(t,\eta_1,\eta_2) + \sum \theta_i^c \cos(2\pi(i-1)t) \quad (5)$$

The first is a gamma density function $b(t,\eta_1,\eta_2) = \eta_2^{\eta_1} t^{\eta_1 - 1} \exp(-\eta_2 t)/\Gamma(\eta_1)$ with shape and scale constants η_1 and η_2 (see Table 1). This models an event-related burst of input that is delayed (by η_1/η_2 sec) with respect to stimulus onset and dispersed by subcortical synapses and axonal conduction. Being a density function, this component integrates to unity over peristimulus time. The second component is a discrete cosine set modeling systematic fluctuations in input, as a function of peristimulus time. In our implementation peristimulus time is treated as a state variable, allowing the input to be computed explicitly during integration.

Critically, the event-related input is exactly the same for all ERPs. This means the effects of experimental factors are mediated through ERP-specific changes in connection strengths. This models experimental effects in terms of differences in forward, backward or lateral connections that confer a selective sensitivity on each source, in terms of its response to others. The experimental or ERP-specific effects are modeled by coupling gains

$$\begin{aligned} C^F_{ijk} &= C^F_{ij} G_{ijk} \\ C^B_{ijk} &= C^B_{ij} G_{ijk} \\ C^L_{ijk} &= C^L_{ij} G_{ijk} \end{aligned} \qquad 6$$

Here, C_{ij} encodes the strength of the latent connection to the *i*-th source from the *j*-th and G_{ijk} encodes its *k*-th ERP-specific gain. By convention, we set the gain of the first ERP to unity, so that subsequent ERP-specific effects are relative to the first[2]. The reason we model experimental effects in terms of gain, as opposed to additive effects, is that by construction, connections are always positive. This is assured; provided the gain is also positive.

The important point here is that we are explaining experimental effects, not in terms of differences in neuronal responses, but in terms of the neuronal architecture or coupling generating those responses. This is a fundamental departure from classical approaches, which characterize experimental effects descriptively, at the level of the states (*e.g.* a face-selective difference in ERP amplitude around 170ms). DCM estimates these response differentials but only as an intermediate step in the estimation of their underlying cause; namely changes in coupling.

Eq.2 defines the neuronal component of the DCM. These ordinary differential equations can be integrated using standard techniques (see Appendix A.2 and Kloeden and Platen, 1999) to generate pyramidal depolarisations, which enters the observer function to generate the predicted MEG/EEG signal.

Figure 3 about here

Observation equations

The dendritic signal of the pyramidal subpopulation of the *i*-th source $x_0^{(i)}$ is detected remotely on the scalp surface in MEG/EEG. The relationship between scalp data and pyramidal activity is linear

$$h = g(x,\theta) = LKx_0 \qquad 7$$

[2] In fact, in our implementation, the coupling gain is a function of any set of explanatory variables encoded in a design matrix, which can contain indicator variables or parametric variables. For simplicity, we limit this paper to categorical (ERP-specific) effects.

where *L* is a lead-field matrix (*i.e.,* forward electromagnetic model), which accounts for passive conduction of the electromagnetic field (Mosher *et al*., 1999). If the spatial properties (orientation and position) of the source are known, then the lead-field matrix *L* is also known. In this case, $K = diag(\theta^K)$ is a leading diagonal matrix, which controls the contribution θ_i^K of pyramidal depolarisation to the *i*-th source density. If the orientation is not known then $L = [L_x, L_y, L_z]$ encodes sensor responses to orthogonal dipoles and the source orientation can be derived from the contribution to these orthogonal components encoded by $K = [diag(\theta_x^K), diag(\theta_y^K), diag(\theta_z^K)]^T$. In this paper, we assume a fixed orientation for multiple dipoles for each source (see Appendix A.3) but allow the orientation to be parallel or anti-parallel (*i.e.* θ^K can be positive or negative). The rationale for this is that the direction of current flow induced by pyramidal cell depolarisation depends on the relative density of synapses proximate and distal to the cell body.

Dimension reduction

For computational reasons, it is sometimes expedient to reduce the dimensionality of the sensor data, while retaining the maximum amount of information. This is assured by projecting the data onto a subspace defined by its principal eigenvectors *E*

$$\begin{aligned} y &\leftarrow Ey \\ L &\leftarrow EL \\ \varepsilon &\leftarrow E\varepsilon \end{aligned} \qquad 8$$

Because this projection is orthonormal, the independence of the projected errors is preserved and the form of the error covariance components of the observation model remains unchanged. In this paper we reduce the sensor space to three dimensions (see appendix A.4).

The observation model

In summary, our DCM comprises a state equation that is based on neurobiological heuristics and an observer based on an electromagnetic forward model. By integrating the state equation and passing the ensuing states through the observer we generate a predicted measurement. This corresponds to a generalized convolution of the inputs to generate an output $h(\theta)$. This generalized convolution furnishes an observation model for the vectorised data[3] *y* and the associated likelihood

$$\begin{aligned} y &= vec(h(\theta) + X\theta^X) + \varepsilon \\ p(y|\theta,\lambda) &= N(vec(h(\theta) + X\theta^X), diag(\lambda) \otimes V) \end{aligned} \qquad 9$$

[3] Concatenated column vectors of data from each channel

Measurement noise ε is assumed to be zero mean and independent over channels, *i.e.* $Cov(\varepsilon) = diag(\lambda) \otimes V$, where λ is an unknown vector of channel-specific variances. *V* represents the errors temporal autocorrelation matrix, which we assume is the identity matrix. This is tenable because we down-sample the data to about 8-ms. Low frequency noise or drift components are modeled by *X*, which is a block diagonal matrix with a low-order discrete cosine set for each ERP and channel. The order of this set can be determined by Bayesian model selection (see below). In this paper we used three components for the first study and four for the second. The first component of a discrete cosine set is simply a constant.

This model is fitted to data using Variational Bayes (see below). This involves maximizing the variational free energy with respect to the conditional moments of the free parameters θ. These parameters specify the constants in the state and observation equations above. The parameters are constrained by a prior specification of the range they are likely to lie in (Friston *et al*., 2003). These constraints, which take the form of a prior density $p(\theta)$, are combined with the likelihood $p(y \mid \theta, \lambda)$, to form a posterior density $p(\theta \mid y, \lambda) \propto p(y \mid \theta, \lambda) p(\theta)$ according to Bayes rule. It is this posterior or conditional density we want to approximate. Gaussian assumptions about the errors in Eq.9 enable us to compute the likelihood from the prediction error. The only outstanding quantities we require are the priors, which are described next.

Prior assumptions

Here we describe how the constant terms, defining the connectivity architecture and dynamical behavior of the DCM, are parameterized and our prior assumptions about these parameters. Priors have a dramatic impact on the landscape of the objective function to be extremised: precise prior distributions ensure that the objective function has a global minimum that can be attained robustly. Under Gaussian assumptions, the prior distribution $p(\theta_i)$ of the *I*-th parameter is defined by its mean and variance. The mean corresponds to the prior expectation. The variance reflects the amount of prior information about the parameter. A tight distribution (small variance) corresponds to precise prior knowledge.

Critically, nearly all the constants in the DCM are positive. To ensure positivity we estimate the log of these constants under Gaussian priors. This is equivalent to adopting a log-normal prior on the constants *per se*. For example, the forward connections are re-parameterized as $C_{ij}^F = \exp(\theta_{ij}^F)$, where $p(\theta_{ij}^F) = N(\mu, v)$. μ and v are the prior expectation and variance of $\ln C_{ij}^F = \theta_{ij}^F$. A relatively tight or informative log-normal prior obtains when $v \approx \frac{1}{16}$. This allows for a scaling around the prior expectation of up to a factor of two. Relatively flat priors, allowing for an order of magnitude scaling, correspond to $v \approx \frac{1}{2}$. The ensuing lognormal densities are shown in Figure 4 for a prior expectation of unity (*i.e.* $\mu = 0$).

Table 1 about here

The parameters of the state equation can be divided into five subsets: (i) *extrinsic connection* parameters, which specify the coupling strengths among areas and (ii) *intrinsic* connection parameters, which reflect our knowledge about canonical micro-circuitry within an area. (iii) *Conduction* delays. (iv) *Synaptic* parameters controlling the dynamics within an area and (v) *input* parameters, which control the subcortical delay and dispersion of event-related responses. Table 1 shows how the constants of the state equation are re-parameterized in terms of θ. It can be seen that we have adopted relatively uninformative priors on the extrinsic coupling $v = \frac{1}{2}$ and tight priors for the remaining constants $v = \frac{1}{16}$. Some parameters (intrinsic connections and inhibitory synaptic parameters) have infinitely tight priors and are fixed at their prior expectation. This is because changes in these parameters and the excitatory synaptic parameters are almost redundant, in terms of system responses. The priors in Table 1 conform to the principle that the parameters we want to make inferences about, namely extrinsic connectivity, should have relatively flat priors. This ensures that the evidence in the data constrains the posterior or conditional density in an informative and useful way. In what follows we review briefly our choice of prior expectations (see David *et al.,* 2005 for details).

Figure 4 about here

Prior expectations

Extrinsic parameters comprise the matrices $\{\theta^F, \theta^B, \theta^L, \theta^G, \theta^U\}$ that control the strength of connections and their gain. The prior expectations for forward, backward, and lateral; $\ln 32$, $\ln 16$ and $\ln 4$ respectively, embody our prior assumption that forward connections exert stronger effects than backward or lateral connections. The prior expectation of θ^G_{ijk} is zero, reflecting the assumption that, in the absence of evidence to the contrary, experimental effects are negligible and the trial-specific gain is $e^0 = 1$. In practice, DCMs seldom have a full connectivity and many connections are disabled by setting their prior to $N(-\infty, 0)$. This is particularly important for the input connections parameterized by θ^U_i, which generally restrict inputs to one or two cortical sources.

We fixed the values of intrinsic coupling parameters as described in (Jansen and Rit, 1995). Inter-laminar conduction delays were fixed at 2 ms and inter-regional delays had a prior expectation of 16 ms. The priors on the synaptic parameters for the *I*-th area $\{\theta^\tau_i, \theta^H_i\}$ constrain the lumped time-constant and relative postsynaptic density of excitatory synapses respectively. The prior expectation for the lumped time constant was 8-ms. This may seem a little long but it has to accommodate, not only dynamics within dendritic spines, but integration throughout the dendritic tree.

Priors on the input parameters $\{\theta^\eta_1, \theta^\eta_2, \theta^c_1, \ldots, \theta^c_8\}$ were chosen to give an event-related burst, with a dispersion of about 32 ms, 96 ms after trial onset. The input fluctuations were relatively constrained with a prior on their coefficients of $p(\theta^c_i) = N(0,1)$. We used the same prior on the contribution of depolarisation

to source dipoles $p(\theta_i^K) = N(0,1)$. This precludes large values explaining away ERP differences in terms of small differences at the cortical level (Grave de Peralta Menendez and Gonzalez-Andino, 1998). Finally, the coefficients of the noise fluctuations were unconstrained, with flat priors $p(\theta^X) = N(0,\infty)$.

Summary

In summary, a DCM is specified in through its priors. These are used to specify (i) how regions are interconnected, (ii) which regions receive subcortical inputs, and (iii) which cortico-cortical connections change with the levels of experimental factors. Usually, the most interesting questions pertain to changes in cortico-cortical coupling that explain differences in ERPs. These rest on inferences about the coupling gains θ_{ijk}^G. This section has covered the likelihood and prior densities necessary for conditional estimation. For each model, we require the conditional densities of two synaptic parameters per source $\{\theta_i^\tau, \theta_i^H\}$, ten input parameters $\{\theta_1^\eta, \theta_2^\eta, \theta_1^c, \ldots, \theta_8^c\}$ and the extrinsic coupling parameters, gains and delays $\{\theta^F, \theta^B, \theta^L, \theta^G, \theta^U, \theta^\Delta\}$. The next section reviews conditional estimation of these parameters, inference and model selection.

BAYESIAN INFERENCE AND MODEL COMPARISON

Estimation and inference

For a given DCM, say model *m*; parameter estimation corresponds to approximating the moments of the posterior distribution given by Bayes rule

$$p(\theta \mid y, m) = \frac{p(y \mid \theta, m)\, p(\theta, m)}{p(y \mid m)} \qquad 10$$

The estimation procedure employed in DCM is described in (Friston, 2002). The posterior moments (conditional mean η and covariance Σ) are updated iteratively using Variational Bayes under a fixed-form Laplace (i.e. Gaussian) approximation to the conditional density $q(\theta) = N(\eta, \Sigma)$. This can be regarded as an Expectation-Maximization (**EM**) algorithm that employs a local linear approximation of Eq.9 about the current conditional expectation. The **E**-step conforms to a Fisher-scoring scheme (Press *et al*., 1992) that performs a descent on the variational free energy $F(q, \lambda, m)$ with respect to the conditional moments. In the **M**-Step, the error variances λ are updated in exactly the same way. The estimation scheme can be summarized as follows:

Repeat until convergence

$$\begin{aligned}\mathbf{E}-\text{Step} \quad & q \leftarrow \min_{q} F(q,\lambda,m) \\ \mathbf{M}-\text{Step} \quad & \lambda \leftarrow \min_{\lambda} F(q,\lambda,m) = \max_{\lambda} L(\lambda,m)\end{aligned}$$

$$\begin{aligned}F(q,\lambda,m) &= \left\langle \ln q(\theta) - \ln p(y \mid \theta,\lambda,m) - \ln p(\theta \mid m) \right\rangle_q \\ &= D(q \,||\, p(\theta \mid y,\lambda,m) - L(\lambda,m)\end{aligned} \qquad 11$$

$$L(\lambda,m) = \ln p(y \mid \lambda,m)$$

Note that the free energy is simply a function of the log-likelihood and the log-prior for a particular DCM and $q(\theta)$. $q(\theta)$ is the approximation to the posterior density $p(\theta \mid y,\lambda,m)$ we require. The **E**-step updates the moments of $q(\theta)$ (these are the variational parameters η and Σ) by minimizing the variational free energy. The free energy is the divergence between the real and approximate conditional density minus the log-likelihood. This means that the conditional moments or variational parameters maximize the log-likelihood $L(\lambda,m)$ while minimising the discrepancy between the true and approximate conditional density. Because the divergence does not depend on the covariance parameters, minimizing the free energy in the **M**-step is equivalent to finding the maximum likelihood estimates of the covariance parameters. This scheme is identical to that employed by DCM for fMRI, the details of which can be found in Friston (2002) and Friston *et al* (2003).

Conditional inference

Inference on the parameters of a particular model proceeds using the approximate conditional or posterior density $q(\theta)$. Usually, this involves specifying a parameter or compound of parameters as a contrast $c^T\eta$. Inferences about this contrast are made using its conditional covariance $c^T\Sigma c$. For example, one can compute the probability that any contrast is greater than zero or some meaningful threshold, given the data. This inference is conditioned on the particular model specified. In other words, given the data and model, inference is based the probability that a particular contrast is bigger than a specified threshold. In some situations one may want to compare different models. This entails Bayesian model comparison.

Model comparison and selection

Different models are compared using their evidence (Penny *et al.,* 2004). The model evidence is

$$p(y \mid m) = \int p(y \mid \theta,m)\, p(\theta \mid m) d\theta \,. \qquad 12$$

The evidence can be decomposed into two components: an accuracy term, which quantifies the data fit, and a complexity term, which penalizes models with a large number of parameters. Therefore, the evidence

embodies the two conflicting requirements of a good model, that it explains the data and is as simple as possible. In the following, we approximate the model evidence for model *m*, with the free energy after convergence. This rests on the assumption that λ has a point mass at its maximum likelihood estimate (equivalent to its conditional estimate under flat priors); *i.e.*, $\ln p(y \mid m) = \ln\langle p(y \mid \lambda, m)\rangle_{\lambda} = L(\lambda, m)$. After convergence the divergence is minimized and

$$\ln p(y \mid m) = L(\lambda, m) \approx -F(q, \lambda, m) \qquad 13$$

See Eq.11. The most likely model is the one with the largest log-evidence. This enables Bayesian model selection. Model comparison rests on the likelihood ratio of the evidence for two models. This ratio is the Bayes factor B_{ij}. For models *i* and *j*

$$\ln B_{ij} = \ln p(y \mid m = i) - \ln p(y \mid m = j) \qquad 14$$

Conventionally, strong evidence in favor of one model requires the difference in log-evidence to be three or more. We have now covered the specification, estimation and comparison of DCMs. In the next section we will illustrate their application to real data using two important examples of how changes in coupling can explain ERP differences.

EMPIRICAL STUDIES

In this section we illustrate the use of DCM by looking at changes in connectivity induced in two different ways. In the first experiment we recorded ERPs during the perception of faces and houses. It is well-known that the N170 is a specific ERP correlate of face perception (Allison *et al.*, 1999). The N170 generators are thought to be located close to the lateral fusiform gyrus, or Fusiform Face Area (**FFA**). Furthermore, the perception of houses has been shown to activate the Parahippocampal Place Area (**PPA**) using fMRI (Aguirre *et al*., 1998; Epstein and Kanwisher, 1998; Haxby *et al*., 2001; Vuilleumier *et al*., 2001). In this example, differences in coupling define the category-selectivity of pathways that are accessed by different categories of stimuli. A category-selective increase in coupling implies that the region receiving the connection is selectively more sensitive to input elicited by the stimulus category in question. This can be attributed to a functional specialization of receptive field properties and processing dynamics of the region receiving the connection. In the second example, we used an auditory oddball paradigm, which produces mismatch negativity (MMN) or P300 components in response to rare stimuli, relative to frequent (Debener *et al.,* 2002; Linden *et al.,* 1999). In this paradigm, we attribute changes in coupling to plasticity underlying the perceptual learning of frequent or standard stimuli.

In the category-selectivity paradigm there are no necessary changes in connection strength; pre-existing differences in responsiveness are simply disclosed by presenting different stimuli. This can be modeled by differences in forward connections. However, in the oddball paradigm, the effect only emerges once

standard stimuli have been learned. This implies some form of perceptual or sensory learning. We have presented a quite detailed analysis of perceptual learning in the context of empirical Bayes (Friston 2003). We concluded that the late components of oddball responses could be construed as a failure to suppress prediction error, after learning the standard stimuli. Critically, this theory predicts that learning-related plasticity should occur in backward connections generating the prediction, which are then mirrored in forward connections. In short, we predicted changes in forward and backward connections when comparing ERPs for standard and oddball stimuli.

In the first example, we are interested in where category-selective differences in responsiveness arise in a forward processing stream. Backward connections are probably important in mediating this selectivity but exhibit no learning-related changes *per se*. We use inferences based on the conditional density of coupling-gain, when comparing face and house ERPs, to address this question. In the second example, our question is more categorical in nature; namely, are changes in backward and lateral connections necessary to explain ERPs differences between standards and oddballs, relative to changes in forward connections alone? We illustrate the use of Bayesian model comparison to answer this question. See Appendices A.3 and A.4 for a description of the data acquisition, lead-field specification and preprocessing.

Category-selectivity: effective connectivity in the ventral visual pathway

ERPs elicited by brief presentation of faces and houses were obtained by averaging trials over three successive 20-minute sessions. Each session comprised 30 blocks of faces or houses only. Each block contained 12 stimuli presented every 2.6s for 400ms. The stimuli comprised 18 neutral faces and 18 houses, presented in grayscale. To maintain attentional set, the subject was asked to perform a one-back task, *i.e.* indicate, using a button press, whether or not the current stimulus was identical to the previous.

As reported classically, we observed a stronger N170 component during face perception in the posterior temporal electrodes. However, we also found other components, associated with house perception, which were difficult to interpret on the basis of scalp data. It is generally thought that face perception is mediated by a hierarchical system of bilateral regions (Haxby *et al.,* 2002). (i) A core system, of occipito-temporal regions in extrastriate visual cortex (inferior occipital gyrus, **IOG**; lateral fusiform gyrus or face area, **FFA**; superior temporal sulcus, **STS**), that mediates the visual analysis of faces, and (ii) an extended system for cognitive functions. This system (intra-parietal sulcus; auditory cortex; amygdala; insula; limbic system) acts in concert with the core system to extract meaning from faces. House perception has been shown to activate the Parahippocampal Place Area (**PPA**) (Aguirre *et al*., 1998; Epstein and Kanwisher, 1998; Haxby *et al*., 2001; Vuilleumier *et al*., 2001). In addition, the Retrosplenial Cortex (**RS**) and the lateral occipital gyrus are more activated by houses, compared to faces (Vuilleumier *et al.,* 2001). Most of these regions belong to the ventral visual pathway. It has been argued that the functional architecture of the ventral visual pathway is not a mosaic of category-specifics modules, but rather embodies a continuous representation of information about object attributes (Ishai *et al.,* 1999).

Figure 5 about here

DCM specification

We tested whether differential propagation of neuronal activity through the ventral pathway is sufficient to explain the differences in measured ERPs. On the basis of a conventional source localization (see Appendix A.3) and previous studies (Allison *et al*., 1999; Haxby *et al.,* 2001; Haxby *et al*., 2002; Ishai *et al*., 1999; Vuilleumier *et al*., 2001), we specified the following DCM (see Figure 5): bilateral occipital regions close to the calcarine sulcus (**V1**) received subcortical visual inputs. From **V1** onwards, the pathway for house perception was considered to be bilateral and to hierarchically connect **RS** and **PPA** using forward and backward connections. The pathway engaged by face perception was restricted to the right hemisphere and comprised connections from **V1** to **IOG**, which projects to **STS** and **FFA**. In addition, bilateral connections were included, between **STS** and **FFA**, as suggested in (Haxby *et al*., 2002). These connections constituted our DCM mediating ERPs to houses and faces. This DCM is constrained anatomically by the number and location of regional sources that accounted for most of the variance in sensor-space (see Appendix A.4). Face- or house-specific ERP components were hypothesized to arise from category-selective, stimulus-bound, activation of forward pathways. To identify these category-selective streams we allowed the forward connections, in the right hemisphere, to change with category. Our hope was that these changes would render **PPA** more responsive to houses while the **FFA** and **STS** would express face-selective responses.

Figure 6 about here

Conditional inference

The results are shown in Figure 6, in terms of predicted cortical responses and coupling parameters. Using this DCM we were able to replicate the functional anatomy, disclosed by the above fMRI studies: the response in **PPA** was more marked when processing houses versus faces. This was explained, in the model, by an increase of forward connectivity in the medial ventral pathway from **RS** to **PPA**. This difference corresponded to a coupling-gain of over five-fold. Conversely, the model exhibited a much stronger response in **FFA** and **STS** during face perception, as suggested by the Haxby model (Haxby *et al*., 2002). This selectivity was due to an increase in coupling from **IOG** to **FFA** and from **IOG** to **STS**. The face-selectivity of **STS** responses was smaller than in the **FFA**, the latter mediated by an enormous gain of about nine-fold $(1/0.11 = 9.09)$ in sensitivity to inputs from **IOG**. The probability, conditional on the data and model, that changes in forward connections to the **PPA**, **STS** and **FFA**, were greater than zero, was essentially 100% in all cases. The connections from **V1** to **IOG** showed no selectivity. This suggests that category-selectivity emerges downstream from **IOG**, at a fairly high level. Somewhat contrary to expectations (see Vuilleumier *et al*., 2001), the coupling from **V1** to **RS** showed a mild face-selective bias, with an increase of about 80% $(1/0.55 = 1.82)$.

Note how the ERPs of each source are successively transformed and delayed from area to area. This reflects the intrinsic transformations within each source, the reciprocal exchange of signals between areas and the ensuing conduction delays. These transformations are mediated by intrinsic and extrinsic connections and are the dynamic expression of category selectivity in this DCM.

The conditional estimate of the subcortical input is also shown in Figure 6. The event-related response input was expressed about 96 ms after stimulus onset. The accuracy of the model is evident in the left panel of Figure 6, which shows the measured and predicted responses in sensor space, after projection onto their three principal eigenvectors.

Auditory oddball: effective connectivity and sensory learning

Auditory stimuli, 1000 or 2000 Hz tones with 5 ms rise and fall times and 80 ms duration, were presented binaurally for 15 minutes, every 2 seconds in a pseudo-random sequence. 2000-Hz tones (oddballs) occurred 20% of the time (120 trials) and 1000-Hz tones (standards) 80% of the time (480 trials). The subject was instructed to keep a mental record of the number of 2000-Hz tones.

Late components, characteristic of rare events, were seen in most frontal electrodes, centered on 250 ms to 350 ms post-stimulus. As reported classically, early components (*i.e.* the N100) were almost identical for rare and frequent stimuli. Using an conventional reconstruction algorithm (see Appendix A.3) cortical sources were localized symmetrically along the medial part of the upper bank of the Sylvian fissure, in the right middle temporal gyrus, left medial and posterior cingulate, and bilateral orbitofrontal cortex (see insert in Figure 7). These locations are in good agreement with the literature: Sources along the upper bank of the Sylvian fissure can be regarded as early auditory cortex, although they are generally located in the lower bank of the Sylvian fissure (Heschls gyrus). Primary auditory cortex has major inter-hemispheric connections through the corpus callosum. In addition, these areas project to temporal and frontal lobes following different streams (Kaas and Hackett, 2000; Romanski *et al.,* 1999). Finally, cingulate activations are often found in relation to oddball tasks, either auditory or visual (Linden *et al.,* 1999).

Figure 7 about here

DCM specification

Using these sources and prior knowledge about the functional anatomy of the auditory system, we constructed the following DCM (Figure 7): an extrinsic (thalamic) input entered bilateral primary auditory cortex (**A1**) which was connected to ipsilateral orbitofrontal cortex (**OF**). In the right hemisphere, an indirect forward pathway was specified from **A1** to **OF** through the superior temporal gyrus (**STG**). All these connections were reciprocal. At the highest level in the hierarchy, **OF** and left posterior cingulate cortex (**PC**) was laterally and reciprocally connected.

Model comparison

Given these nodes and their connections, we created four DCMs that differed in terms of which connections could show putative learning-related changes. The baseline model precluded any differences between standard and oddball trials. The remaining four models allowed changes in forward **F**, backward **B**, forward and backward **FB** and all connections **FBL**, with the primary auditory sources. The results of a Bayesian model comparison (Penny *et al.* 2004) are shown in Figure 7, in terms of the respective log-evidences (referred to the baseline model with no coupling changes). There is very strong evidence for conjoint

changes in backward and lateral connections, above and beyond changes in forward or backward connections alone. The **FB** model supervenes over the **FBL** model that was augmented with plasticity in lateral connections between **A1**. This is interesting because the **FBL** model had more parameters, enabling a more accurate modeling of the data. However, the improvement in accuracy did not meet the cost of increasing the model complexity and the log-evidence fell by 4.224. This means there is strong evidence for the **FB** model, in relation to the **FBL** model. Put more directly, the data are $e^{4.224} = 68.3$ times more likely to have been generated by the **FB** model than the **FBL** model. The results of this Bayesian model comparison suggest the theoretical predictions were correct.

Other theoretical perspectives suggest that the MMN can be explained simply by an adaptation to standard stimuli that may only involve intrinsic connections (see, for example, Ulanovsky *et al* 2003). This hypothesis could be tested using stimulus-specific changes in intrinsic connections and model selection to assess whether the data are explained better by changes in intrinsic connectivity, extrinsic connectivity or both. We will pursue this in a future communication.

Figure 8 about here

Conditional inference

The conditional estimates and posterior confidences for the **FB** model are shown in Figure 8 and reveal a profound increase, for rare events, in all connections. We can be over 95% confident these connections increased. As above, these confidences are based on the conditional density of the coupling-gains. The conditional density of a contrast, averaging over all gains in backward connections, is shown in Figure 9. We can be 99.9% confident this contrast is greater than zero. The average is about one, reflecting a gain of about $e^1 \approx 2.7$, *i.e.,* more than a doubling of effective connectivity.

These changes produce a rather symmetrical series of late components, expressed to a greater degree, but with greater latency, at hierarchically higher levels. In comparison with the visual paradigm above, the subcortical input appeared to arrive earlier, around 64 ms after stimulus onset. The remarkable agreement between predicted and observed channel responses is seen in the left panel, again shown as three principal eigenvariates.

In summary, this analysis suggests that a sufficient explanation for mismatch responses is an increase in forward and backward connections with primary auditory cortex. This results in the appearance of exuberant responses after the N100 in **A1** to unexpected stimuli. This could represent a failure to suppress prediction error, relative to predictable or learned stimuli, which can be predicted more efficiently.

Figure 9 about here

Simulations

In this introductory paper we have focussed on the motivation and use of DCM for ERPs. We hope to have established its construct validity in relation to other neurobiological constructs (*i.e.,* the functional anatomy of category-selectivity as measured by fMRI and predictive coding models of perceptual learning). There are many other aspects of validity that could be addressed and will be in future communications. Here, we briefly establish face validity (the procedure estimates what it is supposed to) of the particular DCM described above. This was achieved by integrating the DCM and adding noise to simulate responses of a system with known parameters. Face validity requires the true values to lie within the 90% confidence intervals of the conditional density. We performed two sets of simulations. The first involved changing one of the parameters (the gain in the right **A1** to **STG** connection) and comparing the true values with the conditional densities. The second used the same parameters but different levels of noise (*i.e.,* different variance parameters). In short, we reproduced our empirical study but with known changes in connectivity. We then asked whether the estimation scheme could recover the true values, under exactly the same conditions entailed by the empirical studies above.

The first simulations used the conditional estimates from the **FB** model of the auditory oddball paradigm. The gain on the right **A1** to **STG** connection was varied from one half to two, i.e. θ_{62}^{G} was increased from $-\ln 2$ to $\ln 2$ in 16 steps. The models were integrated to generate responses to the estimated subcortical input and Gaussian noise was added using the empirical ReML variance estimates. The conditional densities of the parameters were estimated from these simulated data in exactly the same way as for the empirical data. Note that this is a more stringent test of face validity than simply estimating connection strengths: we simulated an entire paradigm and tried to recover the changes or gain in coupling subtending the oddball effect. The results of these simulations are shown in Figure 10 for the connection that changed (right **A1** to **STG**: upper panel) and for one that did not (right **OF** to **A1**: lower panel). In both cases the true value fell within the 90% confidence intervals. This speaks to the sensitivity (upper panel) and specificity (lower panel) of conditional inferences based on this model.

Figure 10 about here

The results of the second simulations are shown in Figure 11. Here we repeated the above procedure but changed the variance parameters, as opposed to a coupling parameter. We simply scaled all the error variances by a factor that ranged from a half to two, in 16 steps. Figure 11 shows that the true value (of the right **A1** to **STG** connection) again fell well within the 90% conditional confidence intervals, even for high levels of noise. These results also speak to the characteristic shrinkage of conditional estimators: Note that the conditional expectation is smaller than the true value at higher noise levels. The heuristic, behind this effect, is that noise or error induces a greater dependency on the priors and a consequent shrinkage of the conditional expectation to the prior expectation of zero. Having said this, the effect of doubling error variance in this context is unremarkable.

Figure 11 about here

Discussion

We have described a Bayesian inference procedure in the context of DCM for ERPs. DCMs are used in the analysis of effective connectivity to provide posterior or conditional distributions. These densities can then be used to assess changes in effective connectivity caused by experimental manipulations. These inferences, however, are contingent on assumptions about the architecture of the model, *i.e.,* which regions are connected and which connections are modulated by experimental factors. Bayesian model comparison can be used to adjudicate among competing models, or hypotheses, as demonstrated above. In short, DCMs can be used to test hypotheses about the functional organization of macroscopic brain activity. In neuroimaging, DCMs have been applied to fMRI data (Friston *et al.,* 2003; Penny *et al.,* 2004; Riera *et al.,* 2004). We have shown that MEG/EEG event-related responses can also be subject to DCM.

The approach can be regarded as a neurobiologically constrained source reconstruction scheme, in which the parameters of the reconstruction have an explicit neuronal interpretation, or as a characterization of the causal architecture of the neuronal system generating responses. We hope to have shown that it is possible to test mechanistic hypotheses in a more constrained way than classical approaches because the prior assumptions are physiologically informed.

Our DCMs use a neural mass model that embodies long-range cortico-cortical connections by considering forward, backward and lateral connections among remote areas (David *et al.,* 2005). This allows us to embed neural mechanisms generating MEG/EEG signals that are located in well-defined regions. This may make the comparison with fMRI activations easier than alternative models based on continuous cortical fields (see Liley *et al.,* 2002; Robinson *et al.,* 2001). However, it would be interesting to apply DCM to cortical field models because of the compelling work with these models.

Frequently asked questions

In presenting this work to our colleagues we encountered a number of recurrent questions. We use these questions to frame our discussion of DCM for ERPs.

- *How do the results change with small changes in the priors?*

Conditional inferences are relatively insensitive to changes in the priors. This is because we use relatively uninformative priors on the parameters about which inferences are made. Therefore, confident inferences about coupling imply a high conditional precision. This means that most of the conditional precision is based on the data (because the prior precision is very small). Changing the prior precision will have a limited effect on the conditional density and the ensuing inference.

- *What are effects of wrong network specification (e.g. including an irrelevant source or not including a relevant source or the wrong specification of connections)?*

This is difficult to answer because the effects will depend on the particular data set and model employed. However, there is a principled way in which questions of this sort can be answered. This uses Bayesian model comparison: If the contribution of a particular source, or connection is in question, one can compute the log-evidence for two models that do and do not contain the source or connection. If it was important the

differences in log-evidence will be significant. Operationally, the effects of changing the architecture are reformulated in terms of changing the model. Because the data does not change these effects can be evaluated quantitatively in terms of the log-evidence (*i.e.* likelihood of the data given the models in question).

- *How sensitive is the model to small changes in the parameters?*

This is quantified by the curvature of the free energy with respect to parameters. This sensitivity is in fact the conditional precision or certainty. If the free energy changes quickly as one leaves the maximum (*i.e.* conditional mode or expectation), then the conditional precision is high. Conversely, if the maximum is relatively flat, changes in the parameter will have a smaller effect and conditional uncertainty is higher. Conditional uncertainly is a measure of the information, about the parameter, in the data.

- *What is the role of source localization in DCM?*

It has no role. Source localization refers to inversion of an electromagnetic forward model. Because this is only a part of the DCM, Bayesian inversion of the DCM implicitly performs the source localization. Having said this, in practice priors on the location or orientation (*i.e.* spatial parameters) can be derived from classical source reconstruction techniques. In this paper we used a distributed source reconstruction to furnish spatial priors on the DCM. However, these priors do not necessarily have to come from a classical inverse solution. Our current evaluations of DCM, using somatosensory evoked potentials (whose spatial characteristics are well known) suggest that the conditional precision of the orientation is much greater than the location. This means that one could prescribe tight priors on the location (from source reconstruction, from fMRI analyses, or from the literature) and let DCM estimate the conditional density of the orientation. We will report these and related issues in Kiebel *et al* (in preparation).

- *How do you select the sources for the DCM?*

DCM is an inference framework that allows one to answer questions about a well-specified model of functional anatomy. The sources specify that model. Conditional inferences are then conditional on that model. Questions about which is the best model use Bayesian model selection as described above. In principle, it is possible to compare an ensemble of models with all permutations of sources and simply select the model that has the greatest log-evidence. We will illustrate this in a forthcoming multi-subject study of the MMN in normal subjects.

- *How do you assess the generalisability of a DCM?*

In relation to a particular data set, the conditional density of the parameters implicitly maximizes generalisability. This is because the free energy can be reformulated in terms of an accuracy term that is maximized and a complexity term that is minimized (Penny *et al* 2004). Minimizing complexity ensures generalization. This aspect of variational learning means that we do not have to use *ad-hoc* measures of generalization (*e.g.* splitting the data into training and test sets). Generalization is an implicit part of the estimation. In relation to generalization over different data sets one has to consider the random effects entailed by different subjects or sessions. In this context, generalization and reproducibility are a more empirical issue. We will report an analysis of the MMN in a large cohort of normal subjects (Garrido *et al*; in preparation).

- *How can you be sure that a change in connectivity is not due to a wrong model?*

There is no such thing as a wrong model. Models can only be better or worse than other models. We quantify this in terms of the likelihood of each model (*i.e.,* the log-evidence) and select the best model. We then usually make conditional inferences about the parameters, conditional on the best model. One could of course argue that all possible models have not been tested, but at least one has a framework that can accommodate any alternative model.

- *What is the basis for the claim that the neural mass models and DCMs are biologically grounded?*

This is based largely on the use of the Jansen and Rit model (1995) as an elemental model for each source. We deliberately chose an established model from the EEG literature for which a degree of predictive and face validity had already been established. This model has been evaluated in a range of different contexts and its ability to emulate and predict biological phenomena has been comprehensively assessed (David *et al* 2003, Jansen and Rit 1995 and references therein). The biological plausibility of the extrinsic connections has been motivated at length in David *et al* (2003), where we show that a network of Jansen and Rit sources can reproduce a variety of EEG phenomena.

- *Why did we exclude thalamus from our models?*

Because it was not necessary to answer the question we wanted to ask. In the models reported in this paper the effects of subcortical transformations are embodied in the parameters of the input function. If one thought that cortico-subcortical interactions were important it would be a simple matter to include a thalamic source that was invisible to measurement space (*i.e.* set the lead field's priors to zero). One could then use Bayesian model comparison to assess whether modeled cortico-thalamic interactions were supported by the data.

- *Does DCM deal with neuronal noise?*

No. In principle DCM could deal with noise at the level of neuronal states by replacing the ordinary differential equations with stochastic differential equations. However, this would call for a very different estimation scheme in which there was conditional uncertainty about the [hidden] neuronal states. Conventionally, these sorts of systems are estimated using a recurrent Bayesian update scheme such as Kalman or Particle filtering. We are working on an alternative (Dynamic Expectation Maximization) but it will be some time before it will be applied to DCM.

Conclusion

We have focused, in this paper, on the changes in connectivity, between levels of an experimental factor, to explain differences in the form of ERFs/ERPs. We have illustrated this through the analysis of real ERPs recorded in two classical paradigms: ERPs recorded during the perception of faces versus houses and the auditory oddball paradigm. We were able to differentiate two streams within the ventral visual pathway corresponding to face and house processing, leading to preferential responses in the fusiform face area and parahippocampal place area respectively. These results concur with fMRI studies (Haxby *et al.,* 2001; Vuilleumier *et al.,* 2001). We have shown how different hypotheses about the genesis of the MMN could be tested, such as learning-related changes in forward or backward connections. Our results suggest that

bottom-up processes have a key role, even in late components such as the P300. This finding is particularly interesting as top-down processes are usually invoked to account for late responses.

The long-term agenda of our modeling program is to establish the validity of neuronal network models so that they can be used as forward models to explain MEG/EEG and fMRI data. As shown in this study the key advantage, afforded by neuronally plausible models in comparison to conventional analyses, is the ability to pinpoint specific neuronal mechanisms underlying normal or pathological responses. By integrating knowledge from various fields dealing with the study of the brain, *i.e.* cognitive and computational neuroanatomy, neurobiology and functional imaging, it may be possible in the near future to construct ever more realistic and constrained models that will allow us to test functionally specific hypotheses. The goal of this paper was to demonstrate the feasibility of this approach in non-invasive electrophysiology.

Software Note

The analyses presented in this paper will be available as a toolbox, distributed with the next release [SPM5] of the SPM software (http://www.fil.ion.ucl.ac.uk/spm)

Acknowledgements

The Wellcome Trust supported this work. We thank Antoine Ducorps (MEG Center, Paris) for his help for the EEG/MRI co-registration, the SHFJ CEA-Orsay for providing the BrainVisa toolbox (cortical mesh), and Sylvain Baillet (CNRS UPR640, Paris) for developing the Brainstorm toolbox (source lead-fields).

Abbreviations:

DCM: Dynamic Causal Model(ing)
EEG: ElectroEncephaloGraphy
ERF: Event-Related Field
ERP: Event-Related Potential
MEG: MagnetoEncephaloGraphy
MMN: MisMatch Negativity

Appendices

A.1 Integrating delay differential equations

Here we describe integration of delay differential equations of the form

$$\dot{x}_i(t) = f_i(x_1(t-\tau_{i1}),\ldots,x_n(t-\tau_{in})) \qquad \text{A.1}$$

for *n* states $x = [x_1(t),\ldots,x_n(t)]^T$, where state *j* causes changes in state *i* with delay τ_{ij}. By taking a Taylor expansion about $\tau = 0$ we get, to first order

$$\begin{aligned}\dot{x}_i(t) &= f_i(x(t)) - \sum_j \tau_{ij}\, \partial f_i / \partial \tau_{ij} \\ &= f_i(x(t)) - \sum_j \tau_{ij} J_{ij} \dot{x}(t)_j\end{aligned} \qquad \text{A.2}$$

where $J = \partial f / \partial x$ is the systems Jacobian. A.2 can be expressed in matrix form as

$$\dot{x}(t) = f - \tau \circ J\dot{x}(t) \qquad \text{A.3}$$

where $\circ$ denotes the Hadamard or element-by-element product. On rearranging A.3 we obtain an ordinary differential equation that can be integrated in the usual way (see appendix A.2).

$$\begin{aligned}\dot{x}_i(t) &= D^{-1} f(x(t)) \\ D &= I + \tau \circ J\end{aligned} \qquad \text{A.4}$$

A.2 Integration

In this work, integration of the ordinary differential equations,

$$\dot{x}(t) = f(x,u) \qquad \text{A.7}$$

proceeded using the Taylor expansion of the change in states

$$\begin{aligned}\Delta x(\tau) &= x(t+\tau) - x(t) \\ &= \tau\,\partial\Delta/\partial\tau + \tfrac{1}{2}\tau^2\,\partial^2\Delta/\partial\tau^2 + \ldots \\ &= Uf(x)\end{aligned} \qquad \text{A.8}$$

$$\begin{aligned}U &= (\tau + \tfrac{1}{2}\tau^2 J + \ldots) \\ &= (\exp(\tau J) - I)J^{-1}\end{aligned}$$

$J = \partial f(x,u)/\partial x$. To avoid matrix inversion, *U* can be computed efficiently with the following pseudo-code

$$\begin{aligned}&U = Q = \tau I \\ &for\ i = 1:256 \\ &\quad Q = \tfrac{1}{i}\tau QJ \\ &\quad U = U + Q \\ &end\end{aligned} \qquad \text{A.9}$$

Critically, *U* is only re-evaluated whenever the input *u* changes. This provides a very efficient integration scheme for systems with sparse inputs (*e.g.* ERP models). However, this efficiency is at the cost of inaccuracies due to ignoring changes in the Jacobian with states (*i.e*. nonlinearities in A.7). These inaccuracies are limited because the nonlinear state equation is evaluated fully at each update $\Delta x = Uf(x,u)$.

A.3 Data acquisition and source reconstruction

Both data sets were acquired from the same subject, in the same session, using 128 EEG electrodes and 2048 Hz sampling. Before averaging, data were referenced to mean activity and band-pass filtered between 1 and 20 Hz. Trials showing ocular artifacts (~30%) and 11 bad channels were removed from further analysis.

EEG electrodes were co-registered with subject's structural MRI and meshes of the scalp and of the white-gray matter interface were extracted (Mangin *et al.,* 1995). 7204 current dipoles were then distributed over and normal to the cortical surface. For each dipole, the EEG scalp topography was computed using a single shell spherical model (Mosher *et al.,* 1999). Regions of interest (patches) were selected as follows: (i) About 0.5% of dipoles were selected by retaining the most significant dipoles from the initial set (David *et al.,* 2002). (ii) The dipoles at the center of mass of the ensuing clusters were selected and neighboring dipoles were added isotropically, to create patches corresponding to cortical patches of about 1-2 cm^2. (iii) The lead-field of each source (columns of the lead field matrix *L*) was then computed by averaging the lead-field of each dipole in the corresponding patch. This assumes a uniform current density within each cortical patch.

This is quite an involved procedure. It should be noted that a lead field could be computed for a small set of equivalent current dipoles, or 'virtual electrodes' placed at the maxima of distributed source reconstructions. The DCM models electrical responses of discrete sources that are defined anatomically by the lead fields. In our example these sources were the cortical patches above. It is important to note that reconstruction procedure is only necessary to define the lead field of the forward model to provide anatomical priors on the model. The analysis *per se* uses the original data in measurement space (or some projection) and, in principle, could proceed without lead fields (*i.e.,* without any anatomical constraints).

A.4 Data preprocessing

To reduce the dimensionality of the data they were projected onto the first three spatial modes following a singular value decomposition of the scalp data, between 0 and 500ms. This was for computational expediency. Reduction using principal eigenvariates preserves the most information in the data, in this case about 70%. This selection of channels or modes that should enter a DCM will be the subject of a technical note (Kiebel *et al* in preparation). Finally, the data were down-sampled in time to 8-ms time bins. Again this was for computational reasons (equivalent results were obtained with 4ms bins; however, the integration scheme became unstable with 16ms bins).

References

Aguirre GK, Zarahn E, D'Esposito M (1998). An area within human ventral cortex sensitive to "building" stimuli: evidence and implications. *Neuron* **21**: 373-383.

Allison T, Puce A, Spencer DD, McCarthy G (1999) Electrophysiological studies of human face perception. I: Potentials generated in occipitotemporal cortex by face and non-face stimuli. *Cereb Cortex* **9**: 415-430.

Büchel C, Friston KJ (1997) Modulation of connectivity in visual pathways by attention: cortical interactions evaluated with structural equation modeling and fMRI. *Cereb Cortex* **7**: 768-778.

Brovelli A, Ding M, Ledberg A, Chen Y, Nakamura R, Bressler SL. (2004) Beta oscillations in a large-scale sensorimotor cortical network: directional influences revealed by Granger causality. *Proc Natl Acad Sci USA*. **101**: 9849-54.

Crick F, Koch C (1998) Constraints on cortical and thalamic projections: the no-strong-loops hypothesis. *Nature* **391**: 245-250.

Dau T (2003). The importance of cochlear processing for the formation of auditory brainstem and frequency following responses. *J Acoust Soc. Am.* **113**: 936-50

David O, Garnero L, Cosmelli D, Varela FJ (2002) Estimation of neural dynamics from MEG/EEG cortical current density maps: application to the reconstruction of large-scale cortical synchrony. *IEEE Trans Biomed. Eng.* **49**: 975-987.

David O, Friston KJ (2003) A neural mass model for MEG/EEG: coupling and neuronal dynamics. *NeuroImage* **20**: 1743-1755.

David O, Harrison L, Friston KJ (2005) Modelling event-related responses in the brain. *NeuroImage* **25**: 756-770.

Debener S, Kranczioch C, Herrmann CS, Engel AK (2002) Auditory novelty oddball allows reliable distinction of top-down and bottom-up processes of attention. *Int. J Psychophysiol*. **46**: 77-84.

Engel AK, Fries P, and Singer W (2001) Dynamic predictions: oscillations and synchrony in top-down processing. *Nat Rev Neurosci* **2**: 704-716.

Epstein R, Kanwisher N (1998) A cortical representation of the local visual environment. *Nature* **392**: 598-601.

Felleman DJ, Van Essen DC (1991) Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex* **1**: 1-47.

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

Friston KJ, Harrison L, Penny W (2003) Dynamic causal modeling. *NeuroImage* **19**: 1273-1302.

Friston KJ. (2003). Learning and inference in the brain. *Neural Networks.* **16**:1325-1352

Grave de Peralta Menendez R, Gonzalez-Andino SL (1998) A critical analysis of linear inverse solutions to the neuroelectromagnetic inverse problem. *IEEE Trans Biomed Eng.* **45**: 440-448.

Roebroeck A, Formisano E, Goebel R (2005) Mapping direct influence over the brain using Granger causality and fMRI. *Neuroimage* **25**: 230-242.

Harrison L, Penny WD, Friston K (2003) Multivariate autoregressive modeling of fMRI time series. *NeuroImage* **19**: 1477-1491.

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* **293**: 2425-2430.

Haxby JV, Hoffman EA, Gobbini MI (2002) Human neural systems for face recognition and social communication. *Biol. Psychiatry* **51**: 59-67.

Ishai A, Ungerleider LG, Martin A, Schouten JL, Haxby JV (1999) Distributed representation of objects in the human ventral visual pathway. *Proc Natl Acad Sci USA* **96**: 9379-9384.

Jansen BH, Rit VG (1995) Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. *Biol. Cybern* **73**: 357-366.

Kaas JH, Hackett TA (2000) Subdivisions of auditory cortex and processing streams in primates. *Proc Natl Acad Sci USA* **97**: 11793-11799.

Kloeden PE, Platen E (1999). *Numerical solution of stochastic differential equations.* Berlin: Springer-Verlag.

Liley DT, Cadusch PJ, Dafilis MP (2002) A spatially continuous mean field theory of electrocortical activity. *Network* **13**: 67-113.

Linden DE, Prvulovic D, Formisano E, Vollinger M, Zanella FE, Goebel R, Dierks T (1999) The functional neuroanatomy of target detection: an fMRI study of visual and auditory oddball tasks. *Cereb Cortex* **9**: 815-823.

Lopes da Silva FH, Hoeks A, Smits H, Zetterberg LH (1974) Model of brain rhythmic activity. The alpha-rhythm of the thalamus. *Kybernetik* **15**: 27-37.

Mangin J-F, Frouin V, Bloch I, Régis J, López-Krahe J (1995) From 3D Magnetic Resonance Images to Structural Representations of the Cortex Topography Using Topology Preserving Deformations. *Journal of Mathematical Imaging and Vision* 297-318.

McIntosh AR, Gonzalez-Lima F (1994) Network interactions among limbic cortices, basal forebrain, and cerebellum differentiate a tone conditioned as a Pavlovian excitor or inhibitor: fluorodeoxyglucose mapping and covariance structural modeling. *J Neurophysiol* **72**: 1717-1733.

Melcher JR, Kiang NY. (1996). Generators of the brainstem auditory evoked potential in cat. III: Identified cell populations. *Hear Res*. **93**: 52-71

Mosher JC, Leahy RM, Lewis PS (1999) EEG and MEG: forward solutions for inverse methods. IEEE Trans *Biomed Eng.* **46**: 245-259.

Penny W, Stephan K, Mechelli A, Friston K (2004) Comparing dynamic causal models. *NeuroImage*.

Press WH, Teukolsky SA, Vetterling WT, Flannery BP (1992) *Numerical recipes in C.* Cambridge MA USA: Cambridge University Press.

Rennie CJ, Robinson PA, Wright JJ (2002) Unified neurophysical model of EEG spectra and evoked potentials. *Biol. Cybern* **86**: 457-471.

Riera JJ, Watanabe J, Kazuki I, Naoki M, Aubert E, Ozaki T, and Kawashima R (2004) A state-space model of the hemodynamic approach: nonlinear filtering of BOLD signals. *NeuroImage* **21**: 547-567.

Robinson PA, Rennie CJ, Wright JJ, Bahramali H, Gordon E, Rowe DL (2001) Prediction of electroencephalographic spectra from neurophysiology. *Phys. Rev E* **63**: 021903.

Robinson PA, Rennie CJ, Rowe DL, O'Connor SC. (2004) Estimation of multiscale neurophysiologic parameters by electroencephalographic means. *Hum Brain Mapp.* **23**: 53-72

Romanski LM, Tian B, Fritz J, Mishkin M, Goldman-Rakic PS, Rauschecker JP (1999) Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat Neurosci* **2**:1131-1136.

Rosenblum MG, Pikovsky AS, Kurths J, Osipov GV, Kiss IZ, Hudson JL (2002) Locking-based frequency measurement and synchronization of chaotic oscillators with complex dynamics. *Phys Rev Lett*. **89**: 264102

Schnitzler A, Gross J. (2005) Normal and pathological oscillatory communication in the brain. *Nat Rev Neurosci*. **6**: 285-96

Smith PH, Populin LC (2001) Fundamental differences between the thalamocortical recipient layers of the cat auditory and visual cortices. *J Comp Neurol* **436**: 508-519.

Stam CJ, Pijn JP, Suffczynski P, and Lopes da Silva FH (1999) Dynamics of the human alpha rhythm: evidence for non-linearity? *Clin Neurophysiol* **110**: 1801-1813.

Tass PA. (2004) Transmission of stimulus-locked responses in two oscillators with bistable coupling. *Biol. Cybern*. **91**: 203-11

Ulanovsky N, Las L, Nelken I. (2003) Processing of low-probability sounds by cortical neurons. Nat *Neurosci*. **6**: 391-8.

Vuilleumier P, Armony JL, Driver J, Dolan RJ (2001) Effects of attention and emotion on face processing in the human brain: an event-related fMRI study. *Neuron*. **30**: 829-841.

Wendling F, Bartolomei F, Bellanger JJ, Chauvel P (2002). Epileptic fast activity can be explained by a model of impaired GABAergic dendritic inhibition. *Eur J Neurosci* **15**:1499-1508.

Figure Legends

Figure 1: Schematic of the DCM used to model a single source. This schematic includes the differential equations describing the dynamics of the source or regions states. Each source is modeled with three subpopulations (pyramidal, spiny-stellate and inhibitory interneurons) as described in (Jansen and Rit, 1995). These have been assigned to granular and agranular cortical layers, which receive forward and backward connection respectively.

Figure 2: Left: Form of the synaptic impulse response function, converting synaptic input (discharge rate) into mean transmembrane potential. Right: The nonlinear static transformation of transmembrane potential into synaptic input. In this figure, the constants are set to unity, with the exception of $r = 0.56$. See main text for details.

Figure 3: Typical hierarchical network composed of three cortical areas. Extrinsic inputs evoke transient perturbations around the resting state by acting on a subset of sources, usually the lowest in the hierarchy. Interactions among different regions are mediated through excitatory connections encoded by coupling matrices.

Figure 4: Log-normal densities on $\exp(\theta)$ entailed by Gaussian priors on θ with a prior expectation of zero and variances of ½ and 1/16. These correspond to fairly uninformative (allowing for changes up to an order of magnitude) and informative (allowing for changes up to a factor of two) priors respectively.

Figure 5: Model definition for the category-selectivity paradigm: The sources comprising the DCM are connected with forward (solid), backward (broken) or lateral (gray) connections as shown. **V1**: primary visual cortex, **RS**: retrosplenial cortex, **PPA**: parahippocampal place area, **IOG**: inferior occipital gyrus, **STS**: superior temporal sulcus, **FFA**: fusiform face area (left is on the left). Insert: Transparent views of the subject's cortical mesh from the top-right, showing the sources that defined the lead field for the DCM: a bilateral extrinsic input acts on the primary visual cortex (red). Two pathways are considered: (i) bilaterally from occipital regions to the parahippocampal place area (blue) through the retrosplenial cortex (green, laterally interconnected), (ii) in the right hemisphere, from primary visual areas to inferior occipital gyrus (yellow) which projects to the superior temporal sulcus (cyan) and the lateral fusiform gyrus (magenta). The superior temporal sulcus and lateral fusiform gyrus are laterally connected

Figure 6: DCM results for the category-selectivity paradigm: Left: Predicted (thick) and observed (thin) responses in measurement space. These are a projection of the scalp or channel data onto the first three spatial modes or eigenvectors of the channel data (Faces: gray. Houses: black). The predicted responses are based on the conditional expectations of the DCM parameters. The agreement is evident. Right: Reconstructed responses for each source and changes in coupling for the DCM modeling category-specific engagement of forward connections, in the ventral visual system. As indicated by the predicted responses in **PPA** and **FFA**, these changes are sufficient to explain an increase response in **PPA** when perceiving houses and, conversely, an increase in **FFA** responses during face perception. The coupling differences mediating this category-selectivity are shown alongside connections, which showed category-specific

differences (highlighted by solid lines). Differences are the relative strength of forward connections during house presentation, relative to faces. The percent conditional confidence that this difference is greater than zero is shown in brackets. Only changes with 90% confidence or more are reported and are highlighted in bold.

Figure 7: DCM specification for the auditory oddball paradigm: Left: Graph depicting the sources and connections of the DCM using the same format as Figure 5: **A1**: primary auditory cortex, **OF**: orbitofrontal cortex, **PC**: posterior cingulate cortex, **STG**: superior temporal gyrus. Insert: localized sources corresponding to the lead fields that entered the DCM: a bilateral extrinsic input acts on primary auditory cortex (red) which project to orbitofrontal regions (green). In the right hemisphere, an indirect pathway was specified, via a relay in the superior temporal gyrus (magenta). At the highest level in the hierarchy, orbitofrontal and left posterior cingulate (blue) cortices were assumed to be laterally and reciprocally connected. Lower right: Results of the Bayesian model selection among DCMs allowing for learning-related changes in forward **F**, backward **B**, forward and backward **FB** and all connections **FBL**. The graph shows the Laplace approximation to the log-evidence and demonstrates clearly that the **FB** model supervenes. The log-evidence is expressed relative to a DCM in which no connections were allowed to change.

Figure 8: DCM results for the auditory oddball (**FB** model). This figure adopts the same format as Figure 6. Here the oddball-related response show many components and are expressed most noticeably in mode 2. The mismatch response is expressed in nearly every source (black: oddballs, gray: standards), and there are widespread learning-related changes in connections (solid lines: changes with more than 90% conditional confidence). In all connections the coupling was stronger during oddball processing, relative to standards.

Figure 9: Conditional density of a contrast averaging over all learning-related changes in backward connections. It is evident that change in backward connections is unlikely to be zero or less given our data and DCM.

Figure 10: Results of simulations showing true and conditional estimates of the connection whose gain was changed (top panel: right **A1** to **STG**) and one whose gain remained the same (lower panel: right **OF** to **A1**). The solid lines are the conditional expectations and the broken lines are the true values. The gray areas encompass the 90% confidence region, based on the conditional variance. In all cases the true values falls within the 90% confidence region (just). These simulations used the conditional expectations and maximum likelihood variance components from the empirical analysis (using the **FB** model) to demonstrate, heuristically, sensitivity and specificity of conditional inferences with this DCM.

Figure 11: Results of simulations showing true and conditional estimates of coupling-gain (right **A1** to **STG**) as a function of error variance. The format of this figure is the same as Figure 10. The variance of simulated observation error was scaled, from half to twice the maximum likelihood estimates of the error variance from the empirical analysis (using the **FB** model). These simulations demonstrate, heuristically, how conditional uncertainty increases with noise. Note that even at high levels of noise the 90% confidence

interval still permits an inference that this connection changed (*i.e.* zero gain falls well outside the gray region).

Tables

Table 1: Prior densities of parameters

(for connections to the *i*-th source from the *j*-th, in the *k*-th ERP)

Extrinsic coupling parameters	$C^F_{ijk} = C^F_{ij} G_{ijk}$	$C^F_{ij} = \exp(\theta^F_{ij})$	$\theta^F_{ij} \sim N(\ln 32, \frac{1}{2})$
	$C^B_{ijk} = C^B_{ij} G_{ijk}$	$C^B_{ij} = \exp(\theta^B_{ij})$	$\theta^B_{ij} \sim N(\ln 16, \frac{1}{2})$
	$C^L_{ijk} = C^L_{ij} G_{ijk}$	$C^L_{ij} = \exp(\theta^L_{ij})$	$\theta^L_{ij} \sim N(\ln 4, \frac{1}{2})$
		$G_{ijk} = \exp(\theta^G_{ijk})$	$\theta^G_{ij} \sim N(0, \frac{1}{2})$
		$C^U_i = \exp(\theta^U_i)$	$\theta^U_i \sim N(0, \frac{1}{2})$
Intrinsic coupling parameters	$\gamma_1 = 1$	$\gamma_2 = \frac{4}{5}$	$\gamma_3 = \frac{1}{4}$ $\gamma_4 = \frac{1}{4}$
Conduction delays (ms)	$\Delta_{ii} = 2$	$\Delta_{ij} = \exp(\theta^\Delta_{ij})$	$\theta^\Delta_{ij} \sim N(\ln 16, \frac{1}{16})$
Synaptic parameters (ms)	$T_i = 16$	$T^{(i)}_e = \exp(\theta^T_i)$	$\theta^T_i \sim N(\ln 8, \frac{1}{16})$
	$H_i = 32$	$H^{(i)}_e = \exp(\theta^H_i)$	$\theta^H_i \sim N(\ln 4, \frac{1}{16})$
Input parameters (sec)	$u(t) = b(t, \eta_1, \eta_2) + \sum \theta^c_i \cos(2\pi(i-1)t)$		$\theta^c_i \sim N(0,1)$
	$\eta_1 = \exp(\theta^\eta_1)$		$\theta^\eta_1 \sim N(\ln 96, \frac{1}{16})$
	$\eta_2 = \exp(\theta^\eta_2)$		$\theta^\eta_2 \sim N(\ln 1024, \frac{1}{16})$