Technical Note

Dynamic causal modelling of evoked responses: The role of intrinsic connections

Stefan J. Kiebel,* Marta I. Garrido, and Karl J. Friston

Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London, 12 Queen Square, London, WC1N 3AR, UK

Dynamic causal modelling is an approach to characterising evoked responses as measured by magneto/electroencephalography (M/EEG). A dynamic causal model (DCM) is a spatiotemporal, generative network model for event-related fields/responses (ERP/ERF) data. Using Bayesian model inversion, one can compute the posterior distributions of the DCM's physiological parameters and its marginal likelihood for model comparison. Model comparison can be used to test mechanistic hypotheses about how electrophysiological data were generated. In this work, we look at the relative importance of changes in intrinsic (within source) and extrinsic (between sources) connections in generating mismatch responses. In short, we introduce the modulation of intrinsic connectivity to the DCM framework. This is useful for testing hypotheses about adaptation of neuronal responses to local influences, in relation to influences that are mediated by long-range extrinsic connections (forward, backward, and lateral) from other sources. We illustrate this extension using synthetic data and empirical data from an oddball ERP experiment.

Introduction

In dynamic causal modelling, one views the brain as a dynamic network that produces observable output. This perspective is useful for constructing fully spatiotemporal, generative or forward models for evoked responses as measured with M/EEG (David et al., 2005; Friston et al., 2003). The aim of DCM is to make inferences about effective connections and their changes in different experimental contexts. This is an important advance over conventional analyses of evoked responses, and allows one to test mechanistic hypotheses about distributed responses derived from cognitive or physiological theories. In brief, DCM entails specification of a plausible model of electrodynamic responses. This model is inverted by optimising a variational free-energy bound on the model's evidence to provide

the conditional density of the model parameters and the evidence itself for model comparison. Bayesian model comparison allows one to select the best model without the risk of 'over-fitting' when using classical 'goodness-of-fit' approaches (Penny et al., 2004).

In DCM for M/EEG, a network usually consists of a few sources (between two and eight), which communicate via directed extrinsic connections. Sensory input enters at primary sources, after passing through thalamic structures. The dynamics are described quantitatively by ordinary differential equations using a state–space description. The spatial expression of each source, at the sensors, is described by the lead field, whose free parameters (*e.g.*, dipole locations and moments) are parameters of the DCM. We use a neural mass model to describe the neuronal dynamics of each source (Jansen and Rit, 1995) and established principles for extrinsic connections among sources (David et al., 2005). Although a clear simplification, the ensuing network model is based on anatomical and physiological features of the brain. These neurobiological constraints furnish spatiotemporal priors for model inversion and enforce a biological parameterisation, which distinguishes DCM from conventional inversion techniques.

In the original application of DCM to ERPs, the difference between two evoked responses was attributed exclusively to changes in extrinsic (between-sources) connectivity (David et al., 2006). Extrinsic connections are mediated by axons that leave grey matter and connect to other sources via the white matter (DeFelipe et al., 2002; Peters, 2002). Conversely, intrinsic connections use axons that do not leave grey matter, *i.e.*, they connect neuronal populations horizontally or vertically within or between cortical layers. To test hypotheses about local adaptation of neuronal populations, this paper extends the original formulation and allows changes in intrinsic connections to explain ERP differences.

We use synthetic data to validate our extension and show that DCM can disambiguate between data generated by changes in intrinsic or extrinsic connections. Furthermore, we will illustrate the use of the ensuing DCM in a multi-subject mismatch negativity (MMN) study (Garrido et al., 2006). The MMN has been studied extensively (Naatanen et al., 2005). In brief, it is the differential response to an unexpected (rare) auditory stimulus relative to an expected (standard) stimulus. The conventional understanding of its genesis rests on an auditory network of change-sensitive

---

* Corresponding author. Fax: +44 20 7813 1420.
 E-mail address: skiebel@fil.ion.ucl.ac.uk (S.J. Kiebel).
 Available online on ScienceDirect (www.sciencedirect.com).

populations. In this note, we will consider two recently described MMN hypotheses, which explain the MMN either by adaptation or within a predictive coding framework (Friston, 2005; Garrido et al., 2006; Jaaskelainen et al., 2004). We will show that both these hypotheses can be formulated and tested using DCM. This paper focuses on the methodology and motivation for DCM of intrinsic changes; consequently, we will focus on the grand mean average of the multi-subject study. A subsequent paper will present a full ERP analysis of each subject and discuss the neurobiological implications of the results from a cognitive neuroscience perspective.

In what follows, we first review the DCM approach to ERPs, with a special focus on the experimental modulation of intrinsic parameters. We then briefly review the MMN hypotheses to be tested. In the results section, we will establish face-validity of the model using synthetic data. Finally, we illustrate the approach using real ERP data.

Dynamic causal modelling

Dynamic causal modelling of evoked responses

For completeness, we review the current DCM for evoked responses. Chapter 6 of Friston et al. (2006) provides a comprehensive summary of the larger research context embedding the DCM approach. 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 sources to produce region-specific responses. This system is modelled 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.

DCM starts with a reasonably realistic neuronal model of interacting cortical regions. This model is then supplemented with a spatial forward model of how neuronal activity is transformed into measured M/EEG responses in the sensors. This output is assumed to be the depolarization of large populations of pyramidal cells (Baillet et al., 2001). The spatial model is a forward model of electromagnetic measurements that accounts for volume conduction effects (Mosher et al., 1999). For example, for both EEG and MEG, one can use an equivalent current dipole (ECD) model to construct the observation equations (Kiebel et al., 2006). This makes DCM a full spatiotemporal model of evoked responses (over sensors and peri-stimulus time) and enables us to invert the model, *i.e.*, to compute the posterior distributions of parameters, from all the observed data.

Intrinsic architecture

DCMs for M/EEG adopt a neural mass model (David and Friston, 2003) to explain source activity in terms of the ensemble dynamics of interacting inhibitory and excitatory subpopulations of neurons, based on the model of Jansen and Rit (1995). This model emulates the activity of a source using three neural

subpopulations, each assigned to one of three cortical layers; an excitatory subpopulation in the granular layer, an inhibitory subpopulation in the supra-granular layer and a population of deep pyramidal cells in the infra-granular layer. The excitatory pyramidal cells receive excitatory and inhibitory input from local interneurons (via intrinsic connections, confined to the cortical sheet), and send excitatory outputs to remote cortical sources via extrinsic connections.

Extrinsic architecture

In David et al. (2005) we developed a hierarchical cortical model to study the influence of forward, backward and lateral connections on evoked responses. This model 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). Using these rules, it is straightforward to construct any hierarchical cortico-cortical network model of cortical sources. Under simplifying assumptions, directed connections can be classified as: (i) Bottom-up or forward connections that originate in the infra-granular layers and terminate in the granular layer. (ii) Top–down or backward connections that connect from infragranular to agranular layers. (iii) Lateral connections that originate in infra-granular layers and target all layers. These long-range or extrinsic cortico-cortical connections are excitatory and are mediated through the axonal processes of pyramidal cells. For simplicity, we do not consider thalamic connections, but model thalamic output as a function operating on the input (see below). In the following, we focus on the neuronal state equations (as opposed to the EEG or MEG observation equations), and show where intrinsic and extrinsic connectivity (and their modulation) enter the model.

The neuronal forward model

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

$$\dot{x} = f(x,u,\theta)$$

$$h = g(x,\theta) \tag{1}$$

where x are the neuronal states of cortical sources, 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 'The model priors'). The state equations are ordinary second-order differential equations and are derived from the behaviour of the three neuronal subpopulations, which operate as linear damped oscillators. The integration of the differential equations pertaining to each subpopulation can be expressed as a convolution (David and Friston, 2003). This convolution transforms the average density of its pre-synaptic inputs into an average postsynaptic membrane potential. The convolution kernel is given by

$$p(t)_e = \begin{cases} \dfrac{H_e}{\tau_e} t\exp(-t/\tau_e) & t \geq 0 \\ 0 & t < 0 \end{cases} \tag{2}$$

where subscript "e" stands for "excitatory" and the subscript "i" is used for inhibitory synapses. $H_{e,i}$ controls the maximum post-

synaptic potential and $\tau_{e,i}$ represents a lumped rate constant. An operator S transforms the potential of each subpopulation into mean firing rate, which is the input to other subpopulations. This operator is assumed to be an instantaneous sigmoid nonlinearity

$$S(x) = \frac{1}{1 + \exp(-\rho_1(x - \rho_2))} - \frac{1}{1 + \exp(\rho_1 \rho_2)} \qquad (3)$$

where the free parameters ρ_1 and ρ_2 determine its form (slope and translation). Interactions, among the subpopulations, depend on internal coupling constants $\gamma_{1,2,3,4}$, which control the strength of intrinsic connections and reflect the total number of synapses expressed by each subpopulation (Fig. 1). The integration of this model, to form predicted responses, rests on formulating these two operators (Eqs. (2) and (3)) in terms of a set of differential equations as shown in Fig. 1.

Note that Fig. 1 represents several sources in matrix/vector notation, e.g., a state vector is given by $x_j = [x_j^{(1)}, x_j^{(2)}, \ldots]^T$, where the subscript stands for state j and the superscript for source i. For the i-th source, $x_0^{(i)}, \ldots, x_8^{(i)}$ represent the mean trans-membrane potentials and currents of its three subpopulations. The equations specify the rate of change of voltage as a function of current and how currents change as a function of voltage and current. For schematic reasons we have lumped superficial and deep pyramidal units together, in the infra-granular layer. The matrices A^F, A^B, A^L encode forward, backward and lateral extrinsic connections respectively. 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 depolarization of pyramidal cells x_0 represents a mixture of potentials induced by excitatory and inhibitory [depolarizing and hyperpolarizing] currents, respectively. This pyramidal potential is not only the output to other sources, but also the presumed source of observed MEG/EEG signals. Note that these equations operate with propagation delays between and within sources, giving delayed differential equations: the full equations have been described in David et al. (2005) but delays are omitted here for clarity.

Modulation of extrinsic connectivity

We can model the differential responses to the different stimuli in two ways. The first (David et al., 2006) is that effects of experimental factors are mediated through changes in extrinsic connection strengths. For example, this extrinsic mechanism can be used to explain response changes by modulating forward (bottom–up) or backward (top–down) coupling. The second mechanism, introduced here, is mediated by changing the intrinsic architecture; of the sort mediating local adaptation. In the following we describe how both intrinsic and extrinsic modulations are implemented.

Changes in extrinsic connectivity are expressed as differences in forward, backward or lateral connections that confer a selective sensitivity on each source, in terms of its response to others. The
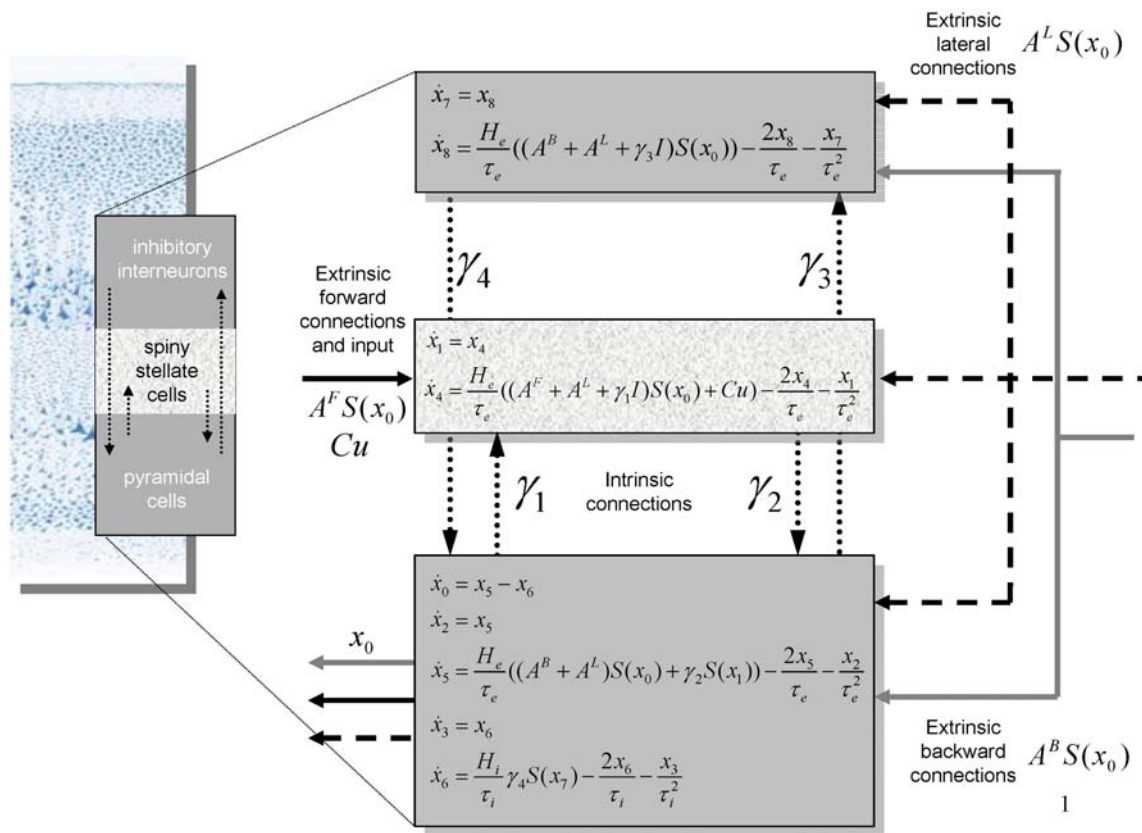

Fig. 1. Neuronal state equations. A source consists of three neuronal subpopulations, which are connected by four intrinsic connections with weights $\gamma_{1,2,3,4}$. Mean firing rates (Eq. (3)) from other sources arrive via forward A^F, backward A^B and lateral connections A^L. Similarly, exogenous input Cu enters receiving sources. The output of each subpopulation is its trans-membrane potential (Eq. (2)).

experimental or stimulus-specific effects are modelled by coupling gains

$$A_{ijk}^{\mathrm{F}} = A_{ij}^{\mathrm{F}} B_{ijk}$$

$$A_{ijk}^{\mathrm{B}} = A_{ij}^{\mathrm{B}} B_{ijk}$$

$$A_{ijk}^{\mathrm{L}} = A_{ij}^{\mathrm{L}} B_{ijk} \qquad (4)$$

Here, A_{ij} encodes the strength of a connection to the i-th source from the j-th and B_{ijk} encodes its gain for the k-th ERP. The superscripts (A, B, or L) indicate the type of connection, i.e., forward, backward or lateral (see also Fig. 1). By convention, we set the gain of the first ERP to unity, so that the gains of subsequent ERPs are relative to the first. The reason we model extrinsic modulations in terms of gain (a multiplicative factor), as opposed to additive effects, is that by construction, connections should be always positive. This is assured; provided both the connection and its gain are positive. In this context, a [positive] gain of less than one represents a decrease in connection strength.

Modulation of intrinsic connectivity

We model the modulation of intrinsic connectivity by a gain on the amplitude H_{e} of the synaptic kernel (Eq. (2)). A gain greater than one effectively increases the maximum response that can be elicited from a source (see Simulations and empirical results). For the i-th source:

$$H_{ek}^{(i)} = H_{\mathrm{e}}^{(i)} B_{iik} \qquad (5)$$

Note that if we considered the gains as elements of a gain matrix, the intrinsic gain would occupy the leading diagonal. Intrinsic modulation can explain important features of typically evoked responses, which are difficult to model with a modulation of extrinsic connections (see Simulations and empirical results).

The aim of introducing intrinsic modulation is to provide for a mechanism that changes the response of a source that is inherently local to that source. Potential candidates for modelling intrinsic effects comprise the four connectivity parameters $\gamma_1, \ldots, \gamma_4$ and the two amplitude parameters H_{e} and H_{i} of the synaptic kernel. Our choice to allow changes in H_{e}, as opposed to other intrinsic parameters is motivated by the fact that this is the only parameter that changes the intrinsic excitability of cells: For example, consider the dynamics of the transmembrane currents of the spiny stellate cells (see Fig. 1); $\tau_e \dot{x}_4 = H_{\mathrm{e}}((A^{\mathrm{F}} + A^{\mathrm{L}} + \gamma_1 I)S(x_0) + Cu) - 2x_4 - x_1/\tau_e$. A change in γ_1 would model a selective change in sensitivity to intrinsic afferents from the pyramidal cells. Conversely, a change in H_{e} models generalised changes in responses to all (extrinsic, intrinsic and exogenous) inputs that is an inherent property of the postsynaptic cells (cf., spike-rate adaptation seen empirically).

Other possible modulations include synaptic time-constants; however, in this work we are interested in evaluating the relative role of intrinsic and extrinsic connectivity, as opposed to synaptic dynamics. In the simulations section we will illustrate the effects of changing all the intrinsic parameters to show that, phenomenologically, modulating H_{e} has the greatest face-validity in terms of changing evoked population responses.

Event-related input

The exogenous input u (Fig. 1) models afferent activity relayed by subcortical structures and is modelled with two components:

The first is a gamma density function (truncated to peri-stimulus time). This models an event-related burst of input that is delayed with respect to stimulus onset and dispersed by subcortical synapses and axonal conduction. Being a density function, this component integrates to unity over peri-stimulus time. The second component is a discrete cosine set modelling systematic fluctuations in input, as a function of peri-stimulus time. In our implementation time is treated as an additional state variable, allowing the input to be computed explicitly during integration. Critically, the event-related input is exactly the same for all evoked responses.

Integration

The equations in Fig. 1, for all areas, can be integrated using the matrix exponential of the systems Jacobian as described in the appendix of David et al. (2006). Note that the integration scheme allows for conduction delays on the connections, which are free parameters of the model. The output of source i is the depolarization of pyramidal cells. This enters the spatial part of the model to generate predicted responses.

The spatial forward model

The dendritic signal of the pyramidal subpopulation of the i-th source $x_0^{(i)}$ is detected remotely in the M/EEG sensors. The relationship between sensor data h and pyramidal activity is linear and instantaneous

$$h = g(x,\theta) = L(\theta^{\mathrm{L}})x_0 \qquad (6)$$

where L is a lead–field matrix (i.e., spatial forward model), which accounts for passive conduction of the electromagnetic field (Mosher et al., 1999). The lead–field is a function of some parameters, i.e., $L(\theta^{\mathrm{L}})$. Here, we assume that each source is well described by a single equivalent current dipole (ECD). Our head model for the EEG-dipoles is based on four concentric spheres, each with homogeneous and isotropic conductivity. The four spheres approximate the brain, skull, cerebrospinal fluid (CSF) and scalp. The lead–field of each ECD is a function of three location and three orientation or moment parameters, θ^{L} (Kiebel et al., 2006).

Data reduction

For computational reasons, it is 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

$$y \leftarrow Ey$$

$$L \leftarrow EL$$

$$\varepsilon \leftarrow E\varepsilon \qquad (7)$$

where ε is the observation error and y is the data (see next subsection). The eigenvectors are computed using principal component analysis or, equivalently, a singular value decomposition (SVD). Because this projection is orthonormal, the independence of the projected errors is preserved and the form of the error covariance components assumed by the observation model remains

unchanged. In this note, we reduce the sensor data to four modes, which usually contain the interesting evoked response components.

The model likelihood

In summary, our DCM comprises a state equation that is based on neurobiological heuristics and an observer equation based on an electromagnetic forward model. By integrating the state equation and passing the ensuing states through the observer equation we generate a predicted measurement. This corresponds to a generalized convolution of the inputs to generate an output $h(\theta)$ (Eq. (6)). This generalized convolution is used for an observation model for the vectorised data[1] y and the associated likelihood

$$y = \text{vec}(h(\theta) + X\theta^X) + \varepsilon$$

$$p(y|\theta, \lambda) = N\big(\text{vec}(h(\theta) + X\theta^X), \text{diag}(\lambda) \otimes V\big) \tag{8}$$

Measurement noise ε is assumed to be zero mean Gaussian and independent over channels, i.e., $\text{Cov}(\text{vec}(\varepsilon)) = \text{diag}(\lambda) \otimes V$, where λ is an unknown vector of mode-specific variances. V represents the error temporal autocorrelation matrix, which we assume is the identity matrix. This is tenable because we typically down-sample the data (to about 8 ms). Low frequency noise or drift components are modelled by X, which is a block diagonal matrix with a low-order discrete cosine set for each evoked response and channel. The order of this set can be determined by Bayesian model selection (see below).

This model is fitted to data by tuning the free parameters θ to minimize the discrepancy between predicted and observed M/EEG time series under model-complexity constraints (more formally, the parameters minimize a variational free energy bound on the marginal likelihood of the model). These parameters specify the constants in the state and observation equations above. In addition to minimizing the prediction error, 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 $p(y|\theta)$, to form a posterior density $p(\theta|y) \propto p(y|\theta)p(\theta)$ according to Bayes' rule. It is this posterior or conditional density we want to estimate. Gaussian assumptions about the errors in Eq. (8) enable us to compute the likelihood from the prediction error. The only outstanding quantities we require are the priors.

The model priors

The parameters of the state equation can be divided into six subsets: (i) extrinsic connection parameters, which specify the coupling strengths among sources, (ii) intrinsic connection parameters, which reflect our knowledge about canonical microcircuitry within a source, (iii) conduction delays, (iv) synaptic and sigmoid parameters controlling the dynamics within an source, (v) input parameters, which control the subcortical delay and dispersion of event-related responses, and, importantly, (vi) intrinsic and extrinsic gain parameters. Table 1 list the priors for these parameters; see also David et al. (2006) for details. Note that 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.

---

[1] Concatenated column vectors of data from each channel.

Inference and model comparison

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

$$p(\theta|y,m) = \frac{p(y|\theta,m)p(\theta,m)}{p(y|m)} \tag{9}$$

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. (6) about the current conditional expectation. The E-step conforms to a Fisher-scoring scheme (Fahrmeir and Tutz, 1994) 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

E−Step $\quad q \leftarrow \min_q F(q, \lambda, m)$

M−Step $\quad \lambda \leftarrow \min_\lambda F(q, \lambda, m)$

$$F(q, \lambda, m) = \langle \ln q(\theta) - \ln p(y|\theta, \lambda, m) - \ln p(\theta|m) \rangle_q$$

$$= D(q \| p(\theta|y, \lambda, m)) - \ln p(y|\lambda, m) \tag{10}$$

Note that the free energy is simply a function of the log-likelihood and the log-prior for a particular DCM and $q(\theta)$. The expression $\langle \cdot \rangle_q$ denotes the expectation under the density q. $q(\theta)$ is the approximation to the posterior density $p(\theta|y, \lambda, m)$ 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 Kullback–Leibler divergence (denoted by $D(\cdot \| \cdot)$), between the real and approximate conditional density minus the log-likelihood. This means that the conditional moments or variational parameters maximize the marginal log-likelihood, while minimizing 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 functional magnetic resonance imaging (Friston, 2002; Friston et al., 2003). Source code for this routine can be found in the Statistical Parametric Mapping software package (see Software Note), in the function 'spm_nlsi_GN.m'.

Bayesian inference proceeds using the conditional or posterior density estimated by iterating Eq. (10). 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 on 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.

Table 1
Prior densities of parameters (for connections to the i-th source from the j-th, in the k-th evoked response)

Extrinsic coupling parameters		$A_{ijk}^F = A_{ij}^F B_{ijk}$	$A_{ij}^F = 32\exp(\theta_{ij}^F)$	$\theta_{ij}^F \sim N(0,\tfrac{1}{2})$
		$A_{ijk}^B = A_{ij}^B B_{ijk}$	$A_{ij}^B = 16\exp(\theta_{ij}^B)$	$\theta_{ij}^B \sim N(0,\tfrac{1}{2})$
		$A_{ijk}^L = A_{ij}^L B_{ijk}$	$A_{ij}^L = 4\exp(\theta_{ij}^L)$	$\theta_{ij}^L \sim N(0,\tfrac{1}{2})$
			$B_{ijk} = \exp(\theta_{ijk}^B)$	$\theta_{ij}^B \sim N(0,\tfrac{1}{2})$
			$C_i = \exp(\theta_i^C)$	$\theta_i^C \sim N(0,\tfrac{1}{2})$
Intrinsic coupling parameters	$\gamma_1 = 128$	$\gamma_2 = \tfrac{4}{5}\gamma_1$	$\gamma_3 = \tfrac{1}{4}\gamma_1$	$\gamma_4 = \tfrac{1}{4}\gamma_1$
Conduction delays (ms)		$\Delta_{ii} = 2$	$\Delta_{ii} = 16\exp(\theta_{ij}^\Delta)$	$\theta_{ij}^\Delta \sim N(0,\tfrac{1}{16})$
Synaptic parameters (ms)		$H_{e,k}^{(i)} = B_{iik} H_e^{(i)}$	$T_e^{(i)} = 8\exp(\theta_i^T)$	$\theta_i^T \sim N(0,\tfrac{1}{8})$
			$H_e^{(i)} = 4\exp(\theta_i^H)$	$\theta_i^H \sim N(0,\tfrac{1}{8})$
			$T_i = 16$	$H_i = 32$
Sigmoid parameters			$\rho_1^{(i)} = \tfrac{2}{3}\exp(\theta_i^{\rho_1})$	$\theta_i^{\rho_1} \sim N(0,\tfrac{1}{8})$
			$\rho_2^{(i)} = \tfrac{1}{3}\exp(\theta_i^{\rho_2})$	$\theta_i^{\rho_2} \sim N(0,\tfrac{1}{8})$
Input parameters (s)			$u(t) = b(t,\eta_1,\eta_2) + \sum \theta_i^C \cos(2\pi(i-1)t)$	$\theta_i^c \sim N(0,1)$
			$\eta_1 = \exp(\theta_1^\eta)$	$\theta_1^\eta \sim N(0,\tfrac{1}{16})$
			$\eta_2 = 16\exp(\theta_2^\eta)$	$\theta_2^\eta \sim N(0,\tfrac{1}{16})$
Spatial (ECD) parameters (mm)			$\theta_i^{pos} \sim N(L_i^{pos}, 32I_3)$	
			$\theta_i^{mom} \sim N(0,8I_3)$	

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

$$p(y|m) = \int p(y|\theta,m)p(\theta|m)\mathrm{d}\theta \qquad (11)$$

Note that the model evidence is simply the normalization constant in Eq. (9). 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, under a normal approximation (Friston et al., 2003), by

$$\ln p(y|m) \approx \ln p(y|\lambda,m) \qquad (12)$$

This is simply the maximum value of the objective function attained by EM (see the **M**-step in Eq. (10)). 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

$$\ln B_{ij} = \ln p(y|m = i) - \ln p(y|m = j) \qquad (13)$$

Conventionally, strong evidence in favour of one model requires the difference in log-evidence to be three or more (Penny et al., 2004). This threshold criterion (i.e., the Bayes factor) plays a similar role as a p-value of $0.05 = 1/20$ in classical statistics (used to reject the null hypothesis in favour of the alternative model). A difference in log-evidence of greater than three (i.e., a Bayes factor more than $\exp(3) \sim 20$) indicates that the data provide strong evidence in favour of one model over the other. This is a standard way to assess the differences in log-evidence quantitatively.

In the next section, we look at how DCMs can be formulated to test hypotheses framed in cognitive or physiological terms. In this case, we look at the DCMs entailed by different mechanistic ideas concerning the MMN.

Mismatch negativity: Hypotheses and models

In this section, we briefly review three hypotheses about the genesis of the mismatch negativity (MMN), the 'traditional', 'adaptation', and 'predictive coding' hypotheses. Here, we use the 'adaptation' and 'predictive coding' hypotheses to motivate the potential importance of both intrinsic and extrinsic mechanisms for the MMN that can be expressed formally as a DCM.

The traditional hypothesis (change-sensitive neurons)

The term 'mismatch negativity' describes an evoked response component elicited by the presentation of a rare auditory stimulus in a sequence of repetitive standard stimuli (Naatanen, 2003). The rare stimulus typically causes a more negative response over peri-stimulus time. The difference between oddball and standard tone reaches a minimum at 85 ms, and exhibits a second minimum later between 100 and 200 ms. The sustained interest in the MMN derives mainly from the finding that the MMN can be elicited in the absence of attention to the auditory stimuli and without task requirements. As the name suggests, the MMN is thought to be the consequence of change-sensitive processes. The MMN seems to be related to the 'automatic' detection of auditory changes (i.e., an oddball preceded by a series of standard tones), and their pre-attentive processing. Because of its robustness, the MMN paradigm is attractive for clinical purposes, and engages sensory and memory systems, and may correlate with involuntary attention switches (Naatanen, 2003).

The adaptation hypothesis (feature-sensitive neurons)

While the traditional hypothesis is largely an explanation for the MMN at the cognitive level, it does not offer a mechanistic

explanation of how the MMN is generated at a neuronal level. May et al. (1999) addressed this question and suggested, based on experimental results combined with computational modelling, that the MMN can be explained by neuronal adaptation, either caused locally or by backward/lateral connectivity. This theme was later experimentally pursued by Jaaskelainen et al. (2004), who found that their data (M/EEG and fMRI) could be modelled by two differentially adapting populations in auditory cortex. Importantly, these sources were found to be at the same location as the sources that generate the evoked auditory response per se. Note that Jaaskelainen et al. (2004) did not discuss the question how the adaptation of feature-specific neuronal populations is caused. In this note, we assume that their hypothesis postulates an intrinsic adaptation of feature-specific neuronal populations.

The predictive coding hypothesis

Recently, predictive coding has been advanced as a parsimonious explanation for several phenomena (Friston, 2005; Rao and Ballard, 1999). It explains the MMN in terms of predictive coding, which assumes that the brain infers the causes of its sensory inputs by predicting them and adjusting the predictions based upon the mismatch or prediction error. Consequently, one would expect that the brain supports an online representation and, in particular, predictions about the auditory environment (see also Winkler et al., 1996). With respect to the MMN, the mechanism for computing the prediction and its dynamic comparison with auditory input is assumed to be implemented by hierarchically organised sources in auditory cortex. Primary auditory sources receive input, which is compared with top–down predictions. The resulting error is then passed up to the next level of the hierarchy so that the prediction can be adjusted. This prediction is then passed back down the hierarchy in an attempt to explain away the prediction error. These recurrent dynamics are presumed to cause the MMN which represents a failure to suppress prediction error. In this model one could interpret the N1-component as an error signal due to the occurrence of an auditory tone that is rapidly explained away by top–down predictions. When the tone is rare or unpredictable this 'explaining-away' will take longer, leading to the emergence of the MMN. Crucially, to the extent that predictions are derived from higher levels of cortical processing, the MMN is a consequence of difference in both intrinsic and extrinsic influences; with intrinsic connections constructing the prediction error and extrinsic connection passing messages between hierarchical levels. As with the adaptation hypothesis, the MMN is modelled as an integral part of the auditory network response. The difference between the adaptation and predictive coding hypotheses is that predictive coding does not make an assumption about feature-sensitive populations. Rather, the whole system makes ongoing predictions about the input. For example, as noted by Naatanen et al. (2005), the adaptation hypothesis would fail to explain missing (but expected) tones (which cause an MMN), or the MMN response to a violation in some predictable feature-changes (e.g., a tone ladder).

In terms of causal architectures, the adaptation of feature-selective populations can be explained by purely intrinsic neuronal mechanisms, whereas hierarchical inference using predictive coding invokes recurrent interactions that are mediated by extrinsic forward and backward connections. Under the predictive coding hypothesis, differences between responses to predictable and unpredictable stimuli would be manifest as changes in both intrinsic and extrinsic coupling.

We now describe how one can express the adaptation and predictive coding hypotheses in DCM. The idea is to model the data using plausible but competing models that embody the different hypotheses, invert these models and use their evidences to find the best model or hypothesis. As mentioned in the Introduction, we will illustrate the approach using the grand mean (i.e., the average of the evoked responses from all subjects). The analysis of the grand mean is quasi-standard in most ERP/ERF analyses and is informative, because it reveals features that are common to single-subject responses. Practically, because of its high signal-to-noise ratio, people often use the grand mean for source localization or feature selection (e.g., identifying peaks in peri-stimulus time and channels).

DCM for the adaptation hypothesis

The architecture of the DCMs we tested was motivated by recent electrophysiological and neuroimaging studies looking at the sources underlying the MMN (Doeller et al., 2003; Opitz et al., 2002). We use a five-source network (Fig. 2). As described in Garrido et al. (2006) the input feeds via subcortical structures into two bilateral sources in posterior auditory cortex (lA1 and rA1). These have reciprocal forward and backward connections to two bilateral sources in anterior auditory cortex, i.e., superior temporal gyri (lSTG and rSTG). These two sources are laterally and reciprocally connected via the corpus callosum. The fifth source is located in the right inferior frontal gyrus (rIFG) and is reciprocally connected with rSTG. The prior expectations of the location of all five sources are listed in Table 2. The evoked response to standard tones is modelled by this network. The response to rare tones is modelled by the same network, except for a gain in the intrinsic excitability of all sources.

DCM for the predictive coding hypothesis

We use the same network as for the adaptation hypothesis. The difference is that we allowed extrinsic changes. Following Garrido et al. (2006), we used three different models with changes only in the extrinsic connections: The first model allows for a modulation of forward connections (F), the second for a modulation of backward connections (B), and the third model for both (FB). We will refer to these as extrinsic models. However, most formulations of predictive coding would predict both intrinsic and extrinsic modulation within the same model. In this context, it is possible that the mismatch response comprises two components: an N1 adaptation, mediated by intrinsic changes and a later MMN proper, mediated by changes in extrinsic connections. In terms of the theoretical perspectives on these mechanisms, one could construe intrinsic adaptation as that part of predictive coding which involves changes in lateral interactions within a cortical level (Friston, 2005).

From all possible combinations of intrinsic and extrinsic modulation, we focussed on two types: First, a mixture of F, B, and FB models with intrinsic modulation of the first two sources (lA1 and rA1). This corresponds to a predictive coding mechanism for all hierarchical levels but allows for intrinsic adaptation of the primary auditory responses. The intrinsic modulations serve as a mechanism to explain differences in the N1-component (a large response around 100 ms in peri-stimulus time), while extrinsic changes may explain later difference corresponding to the MMN per se. The resulting models are called FI, BI, and FBI. Second,
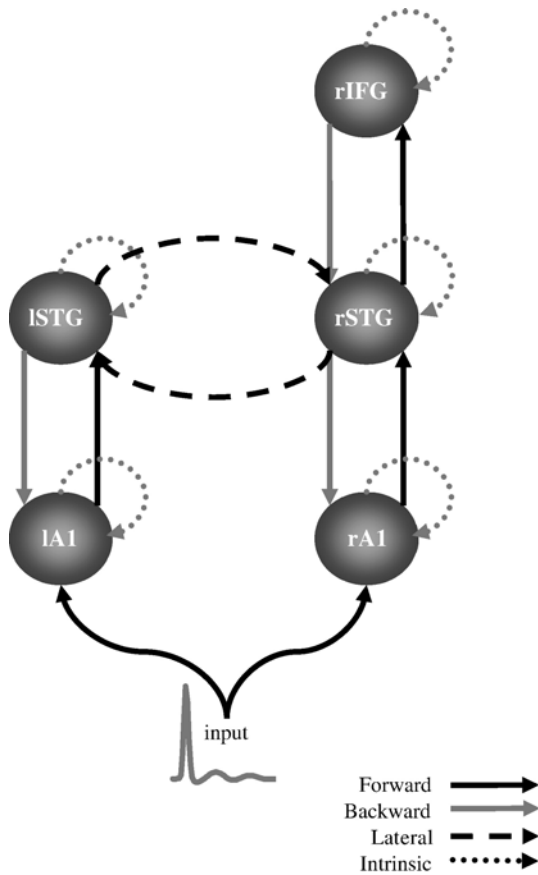
Fig. 2. Neuronal network used for explaining the mismatch response with DCM. Different models are derived from gain modulation of subsets of the indicated intrinsic and extrinsic connections. For example, model **F** allows for the modulation of extrinsic forward connections only, whereas model **FBA** uses all possible modulations (as shown here). See text for further description.

we allow for intrinsic modulations of all sources in the three extrinsic models. We call the resulting models **FA**, **BA**, **FBA**. In summary, there are ten models: One intrinsic adaptation model **A**, three extrinsic models (**F**, **B**, **FB**), and six combined models (**FI**, **BI**, **FBI**), and (**FA**, **BA**, **FBA**). This is complemented by a null model, in which we do not allow for any modulation (**0**). See also Table 3 for a list of all models.

Simulations and empirical results

Simulations

In this section, we motivate and validate the extension of DCM using simulated data in which the true parameters and their

Table 2
Prior coordinates L_i^{pos} of equivalent current dipoles for MMN model, in MNI space (mm)

Left primary auditory cortex (**lA1**)	−42	−22	7
Right primary auditory cortex (**rA1**)	46	−14	8
Left superior temporal gyrus (**lSTG**)	−61	−32	8
Right superior temporal gyrus (**rSTG**)	59	−25	8
Right inferior frontal gyrus (**rIFG**)	46	20	8

changes are known. With these simulations we establish face-validity, *i.e.*, we check whether the model inversion is veridical, given data that have been generated from the same class of models. We performed three sets of simulations. In the first, we compared four potential mechanisms of intrinsic adaptation. These were modelled as gains on parameters (i) γ_1, (ii) γ_2, (iii) H_e, and (iv) H_i. We show that the modulation of either γ_2 or H_e provides a suitable modulation of intrinsic dynamics in phenomenological terms (i.e., establish face-validity). In the second simulations, we demonstrate that the modulation of intrinsic versus extrinsic connectivity parameters leads to distinct dynamics in simple networks. We illustrate this by inverting DCMs of synthetic data generated by a two-source network. We will use model comparison to examine whether one can identify the correct model, given alternative models (i.e., establish predictive validity). In the third and final simulations, we will generate realistic data using the MMN models defined above. The aim of these simulations is the same as above, i.e., to assess whether the correct model can be identified using model comparison, using responses observed through the lead field.

Synthetic data: single-source models

We looked at the effects of changing the parameters γ_2 and H_e and on γ_1, and H_i, using a single-source system (Fig. 3A). For these simulations, we set all the parameters to their prior expectations (Table 1). Note that we did not use a lead field for generating data, because we wanted to focus on the neuronal responses. We varied each of the four intrinsic parameters around a gain of one (i.e., no modulation). We generated data by integrating the resulting system (0 to 300 ms in peri-stimulus time) using a gamma-function as input. In Fig. 3C, one can see that increasing both γ_2 and H_e produce marked quantitative increases in evoked responses. This is evidently not the case for γ_1, because of the saturation imposed by the sigmoid non-linearity, the magnitude of the modulated response reaches a maximum only slightly higher than the non-modulated response (see Fig. 3B). Rather, a high γ_1 tends to prolong the response. Modulation of H_i causes a change in the response after the peak, *i.e.*, the undershoot is more pronounced for higher values and is attenuated with low values.

The response changes due to the modulation of γ_2 and H_e are very similar. The only difference is that H_e, given the same peak response magnitude, causes a slightly earlier return to baseline and a slightly more negative undershoot (Fig. 3B). This behaviour is

Table 3
Modulations used in eleven models for the MMN

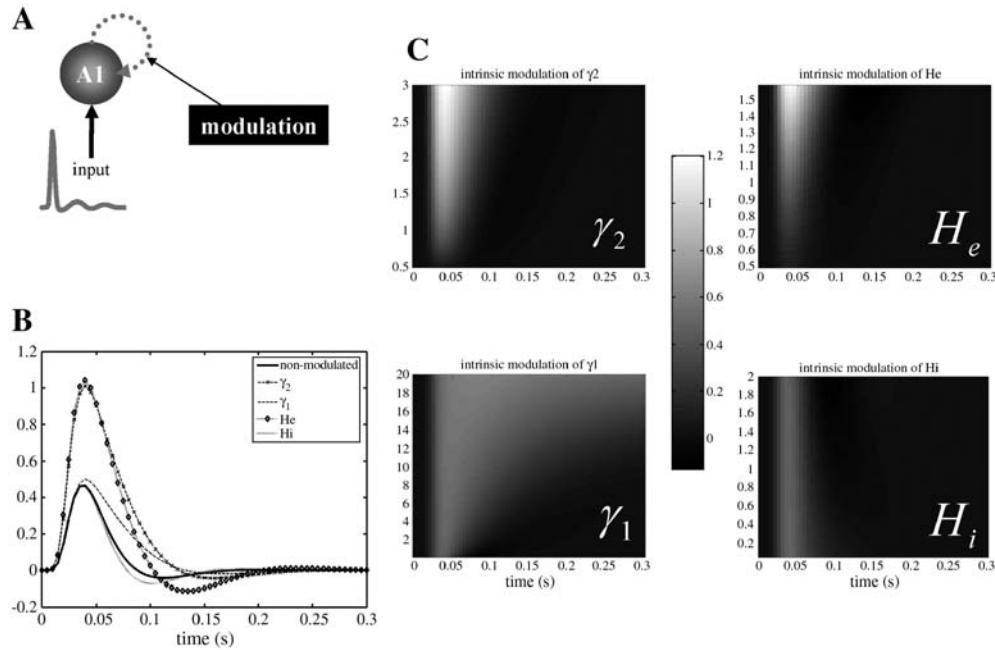|	Extrinsic forward	Extrinsic backward	Extrinsic lateral	Intrinsic bilateral A1	Intrinsic all sources
F	X				
B		X			
FB			X		
FI	X			X	
BI		X		X	
FBI			X	X	
FA	X				X
BA		X			X
FBA			X		X
A					X
0					

Fig. 3. First set of simulations. (A) The generating network. The intrinsic connectivity of the single source changed between two evoked responses. We generated data for a modulation of four different parameters: γ_1, γ_2, H_e, and H_i. (B) The non-modulated response and responses for gain factors 3.1 (γ_1), 2.28 (γ_2), 1.48 (H_e), and 1.43 (H_i). The modulation of γ_2 and H_e results in an amplitude modulation, whereas a modulation of γ_1 and H_i changes the response after the peak. (C) The modulated responses for all four parameters over some representative range. We selected gains such that the largest amplitude modulation was three times the amplitude of the non-modulated response (~0.45, see B).

due to the concurrent modulation of inhibitory interneuron responses to excitatory input, which hyperpolarises pyramidal cells. Phenomenologically, γ_2 and H_e are both candidates for modelling adaptation or changes in excitability. We selected H_e because its effects are closer to a simple response modulation and it avoids specific assumptions about the synaptic mechanisms entailed by changing γ_2.

Synthetic data: two-source models

In these simulations, we show that the change in sensor space, due to a modulation of H_e can be distinguished from changes caused by an extrinsic connectivity modulation. This is important because detectable differences are necessary to identify changes in intrinsic and extrinsic parameters that are conditionally independent. In the following, we present a case study to show the changes one can typically expect under different modulation mechanisms.

We generated data from a two-source network (Fig. 4). The first source **A1** received the input. Source **A1** is connected with **A2** via forward and backward connections. As in the first set of simulations, the data are observed directly at the source level. We used four models: (i) intrinsic modulation of the first source (**i1**), (ii) intrinsic modulation of the second source (**i2**), (iii) extrinsic forward modulation (**F**), and (iv) extrinsic backward modulation (**B**).

The intrinsic parameters were set to their prior mean (Table 1). The forward connectivity (**A1** to **A2**) was two and the backward connectivity (**A2** to **A1**) was five. For model **i1** and **i2** we used an intrinsic gain (on H_e) of 1.4. For models **F** and **B** we used a gain of three on both the forward and backward connections. We generated 300 ms of data, starting at stimulus presentation. The responses of all four models (without noise) are shown in Fig. 5A. The

responses of the four models have distinct characteristics: For example, the response of **i1** is unique in the sense that no other model generates an amplitude modulation for the **A1** source. Similarly, it is difficult to emulate the response of model **B** with any other model: With **B** the dynamics only change after 100 ms, whereas all other models produce some change in the early response component, either in the first or second source.

This intuition is supported by inverting each of the four models using the four data sets (after the addition of white noise with 5% standard deviation of the peak amplitude of the second source). In Fig. 5B we show all log-evidences for the sixteen inversions. Usually, a difference of about three or more constitutes strong evidence for one model over another (Penny et al., 2004). All four models are identified correctly. The smallest differences in model evidence were seen for the data-model combination **i2**−**i2** and **i2**−**F** (867−852=15). The differences between the responses of **i2** and **F** are subtle (Fig. 5A); in the second source, the response of the **i2** model is slightly more delayed than the **F** model.

Two observations are important and hold for networks with more than two sources: (i) The modulation of intrinsic parameters within the first source cannot be emulated by changes in extrinsic parameters and represents a potentially important degree of freedom for DCMs of evoked responses. (ii) Intrinsic modulation of higher sources resembles the effects of increasing forward connectivity to that source. However, the extrinsic changes will be specific to the afferent connection eliciting a response.

Synthetic data: five-source models

In the final simulations we used a realistic five-source network to show that (i) models with a combination of intrinsic and extrinsic modulations can be inverted successfully and (ii) that the
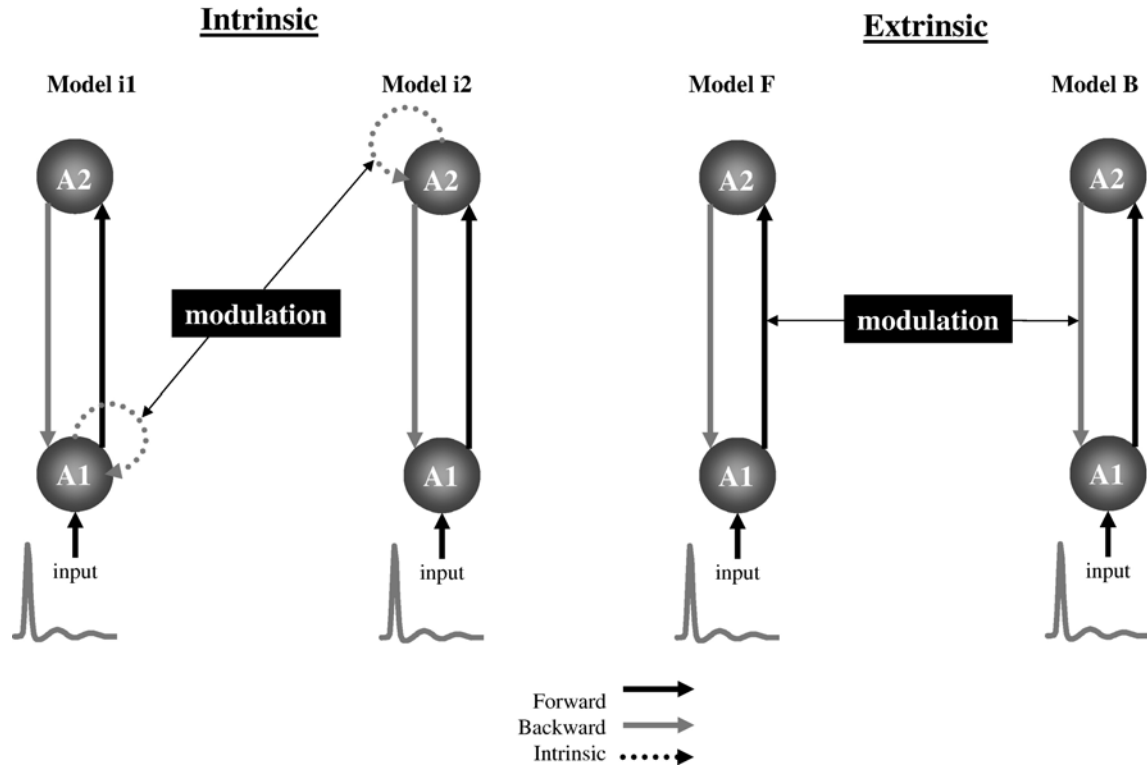
Intrinsic

Extrinsic

Fig. 4. Second set of simulations, two-source models. We used four different models (**i1**, **i2**, **F**, **B**) to generate data. Models **i1** and **i2** implement a modulation of the intrinsic parameter H_e, for the first and second sources. Models **F** and **B** implement a modulation of the forward and backward connection. Each model resulted in four times-series, two for each source, with and without modulation.

correct model can be identified on the basis of its log-evidence. For each of ten models (**A**, **F**, **B**, **FB**, **FI**, **BI**, **FBI**, **FA**, **BA**, **FBA** – see above), data were generated using both the neuronal state equations (Fig. 1) and the M/EEG observation equations (Eq. (6)). We added white noise with a standard deviation of 0.5. The model parameters were taken from the analysis of the real ERP data presented in the next section (i.e., the conditional means). For each model, this gave two ERPs (standard and rare responses), over channels and peri-stimulus time 0 to 200 ms. As in the previous section, we inverted each model using all ten synthetic data sets. In Table 4, we list all the log-evidences for the ensuing hundred inversions. Each column contains the log-evidences for each data set.

This matrix is shown as an image in Fig. 6 (after subtracting the maximum from each column): Most of the maximum values (coded as white) can be found on the diagonal, i.e., the true model had the greatest evidence, among all models. In three cases, the maximum log-evidences were for other models: for data generated by **FA**, **BA**, and **FBA** the maximum evidences were for models **FI**, **BI**, and **FBI**. However, in none of these cases did we found strong evidence in favour of the wrong model. The evidence for the true model was always close to the maximum log-evidence (i.e., a difference of less than three) and therefore cannot be discounted as a viable model. There were other cases where the log-evidences of models for a given data set were similar. This was expected

Table 4
Log-evidences for third set of simulations (see text)

Data										
Model	F	B	FB	FI	BI	FBI	FA	BA	FBA	A
F	−375.0	−481.8	−432.4	−482.7	−797.6	−577.3	−515.9	−624.9	−585.6	−657.9
B	−943.8	−402.8	−920.0	−939.5	−689.1	−937.0	−927.9	−699.2	−949.1	−906.1
FB	−376.5	−610.7	−378.2	−474.8	−809.8	−561.3	−510.3	−779.5	−581.7	−640.6
FI	−382.5	−652.6	−485.6	−387.9	−569.2	−401.3	−395.2	−583.9	−411.6	−677.1
BI	−729.0	−412.8	−595.2	−710.9	−399.7	−692.5	−684.0	−401.9	−708.1	−681.6
FBI	−384.3	−626.2	−387.3	−390.3	−564.1	−385.6	−396.6	−650.3	−402.2	−509.4
FA	−386.2	−510.0	−414.2	−391.0	−482.9	−405.8	−398.5	−510.9	−416.2	−385.2
BA	−583.6	−416.4	−492.8	−513.1	−406.1	−463.1	−444.3	−402.0	−471.-9	−391.0
FBA	−388.5	−557.9	−391.6	−394.0	−449.3	−389.0	−396.5	−511.5	−405.5	−493.9
A	−454.5	−612.6	−470.6	−408.4	−454.2	−449.3	−403.2	−591.1	−481.7	−385.3

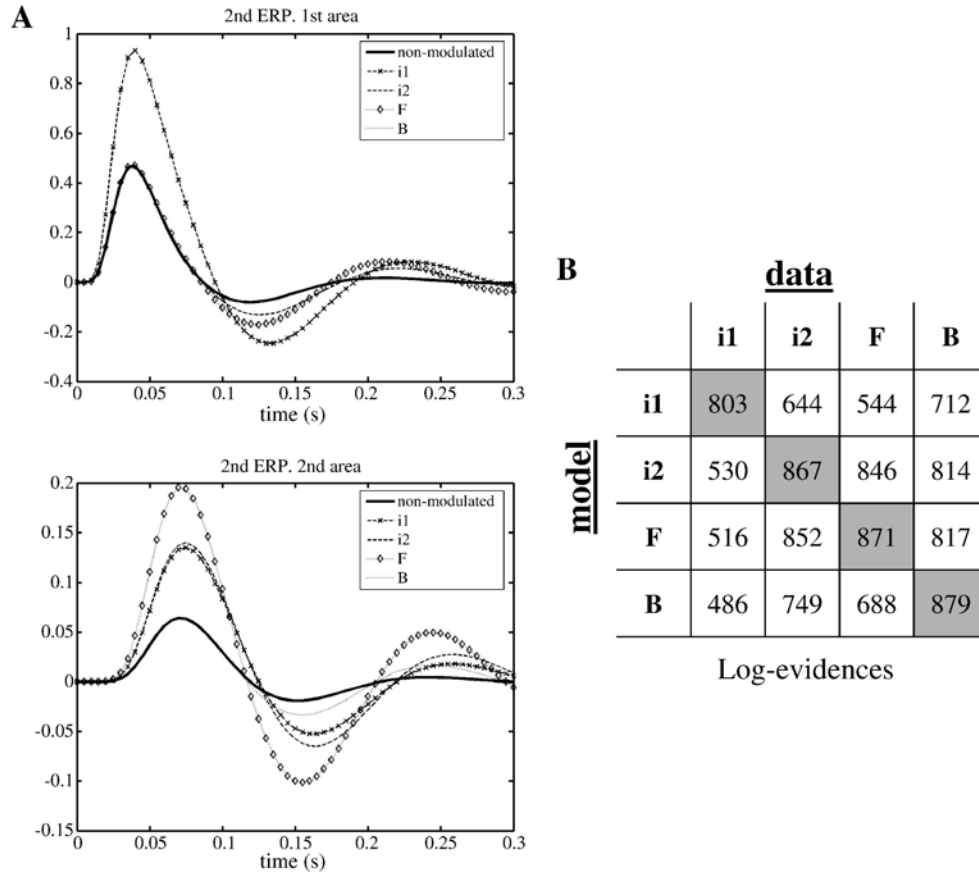Grey elements: model with maximum log evidence for each data set (see also Fig. 6).

Fig. 5. Second set of simulations: results. (A) Plots of generated data for each of the four models (see Fig. 2). (B) White noise (see text) was added to the data and for each data set, we inverted each of the four models and show all log model evidences. Grey cells indicate the best and, in all cases, correct model.
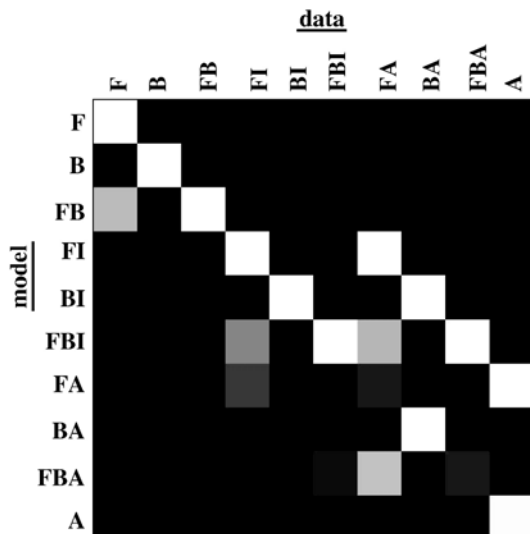

Fig. 6. Third set of simulations: log-model evidences, normalised for visualization purposes (see text for details). Each column shows the log-evidences for a given data set. White blocks indicate the highest model evidence, for each column. Grey values indicate log-evidences up to a difference of minus five, relative to the maximum for a given data set. Black indicates models whose log-evidences were five or more under the maximum.

because many of our models are nested, *i.e.*, one model is an extension of another (*e.g.*, **FBA** is an extension of **FBI**, because **FBA** is equal to model **FBI** with the addition of three intrinsic modulations). In cases when a model extension actually explains an important part of the data, one finds large log-evidences differences between nested models. For example, with data generated by model **FB**, and explained by model **B** and **FB**. For these data, model **B** (−920; see Table 4) has a much lower model evidence than model **FB** (−378). In this case, changes in forward connectivity are necessary to explain these data.

Mismatch negativity study

We studied a group of fourteen healthy volunteers aged 24–35 (5 females). Each subject gave signed informed consent before the study, conducted under local ethical committee guidelines. Subjects sat on a comfortable chair in front of a desk in a dimly illuminated room. Electroencephalographic activity was measured during an auditory 'oddball' paradigm; subjects heard "standard" (1000 Hz) and "deviant" tones (2000 Hz), occurring 80% (480 trials) and 20% (120 trials) of the time, respectively, in a pseudo-random sequence. The stimuli were presented binaurally via headphones for 15 min every 2 s. The duration of each tone was 70 ms with 5 ms rise and fall times. The subjects were instructed not to move, to keep their eyes closed and to count the deviant tones.

Acquisition and pre-processing

EEG data were recorded with a Biosemi system and 128 scalp electrodes at a sampling rate of 512 Hz. Vertical and horizontal eye movements were monitored using EOG (electro-oculogram) electrodes. Data was epoched offline, with a peri-stimulus window of −100 to 400 ms, down-sampled to 200 Hz, band-pass filtered between 0.5 and 40 Hz and re-referenced to the average of the right and left ear lobes. Trials in which the absolute amplitude of the signal exceeded 100 μV were excluded. Two of the subjects were eliminated from further analysis due to excessive numbers of trials containing artefacts. In the remaining subjects an average 18% of trials were excluded. We formed the grand mean of all remaining subjects, i.e., the average over all ERPs. For computational expediency the dimensionality of the data was reduced to four spatial modes (see above). These were the principal modes of a singular value decomposition of the channel data between 0 and 200 ms, from both trial types. The four principal eigenvariates preserved 93.4% of the experimental variance.

Dynamic causal modelling

We assumed five cortical sources, modelled as equivalent current dipoles (ECDs), over left and right primary auditory cortices (**lA1**, **rA1**), left and right superior temporal gyrus (**lSTG**, **rSTG**) and right inferior frontal gyrus (**rIFG**). The prior means of the dipole locations are provided in Table 2. The prior variance was 32 mm in each direction. The parameters encoding orientation (moments in three orthogonal directions) had a prior mean of zero and a variance of eight.

As described above, we inverted eleven DCMs. An adaptation model **A**, three extrinsic models (**F**, **B**, **FB**), the mixed models (**FI**, **BI**, **FBI**, **FA**, **BA**, **FBA**), and a null model (**0**). In Fig. 7, we plot the log-evidence for each model, minus the log-evidence of the null model.[2] The intrinsic adaptation (**A**) has the largest log-evidence (**159.8**), followed by **FBI** (**158.0**), and **FBA** (**156.1**). Because of the negligible difference (**1.8**) between their log-evidences, the **A** and **FBI** models are the best among the models, for these data. As previously established by Garrido et al. (2006), the **B** model (**95.9**) is not a compelling model. However, note that once the **B** model is complemented with intrinsic modulations, the ensuing models **BI** (**132.9**) and **BA** (**145.7**) are much better. For the forward models **F** (**144.0**), **FI** (**145.5**), and **FA** (**146.8**), model evidences remain relatively unchanged. This suggests that including intrinsic modulations in a forward-modulation model offers only a slight advantage.

In Fig. 8, the posterior means of the coupling gains of the two best models (**A** and **FBI**) are shown: we only show changes that have a posterior probability of 95% or more. For the **FBI** model, this probability is more than 99% for the two intrinsic modulations, two of the forward connections, and the left backward connection. For the **A** model, this is true for the intrinsic adaptation of four sources (both **A1** and **STG**).

In summary, one can conclude that intrinsic adaptation is necessary to explain the grand mean data and if changes in extrinsic connections cause a mismatch response, these changes are likely to involve both forward and backward connections. We will

---

[2] This subtraction is just for visualisation purposes. It does not change the inference made, because we subtract the same value from all log-evidences.
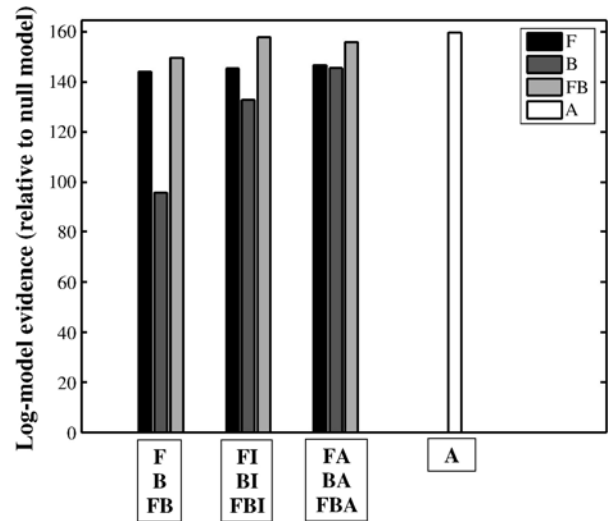

Fig. 7. Mismatch data: This plot shows log-evidences of ten models, adjusted with respect to the null model. See text for a description of the models. The two best models are the adaptation model **A**, and model **FBI**, which uses a mixture of extrinsic and intrinsic modulation.

present a fuller analysis of this issue (and the implications for the various theories of the MMN) using single-subject analysis and a more comprehensive search of model space in a subsequent paper (Garrido et al., in preparation).

Discussion

We have complemented the original DCM (David et al., 2006) by allowing changes in within-source or intrinsic connectivity. These intrinsic changes can be used to test hypotheses which postulate local neuronal adaptation as a mechanism for the genesis of evoked responses. We have demonstrated the usefulness of this extension using the grand mean of a multi-subject mismatch ERP study. We anticipate that similar intrinsic adaptation models will be useful in testing hypotheses about other ERP/ERF phenomena.

Changes in intrinsic connectivity were modelled by a gain on the parameter H_e of the kernel modelling postsynaptic responses to excitatory pre-synaptic inputs (Eq. (2)). Although H_e is not a connectivity parameter *per se*, it does act as one, because it increases the influence the three sub-populations have on each other (*i.e.*, it encodes effective post-synaptic receptor density). We have shown that changes in this parameter emulate an amplitude modulation of a source's response, which provides an intuitive interpretation. We have shown, using model inversion and simulated data, that one can disambiguate between competing models, which explain differences between evoked responses either by intrinsic or by extrinsic mechanisms. In the third simulations, we generated synthetic data based on the posterior estimates from real evoked responses: as an aside, we note that DCM can be used to produce realistic evoked responses (over all peri-stimulus times and channels) that may be useful for developing and validating other spatiotemporal models for evoked responses.

We found that models based solely on changes in extrinsic connectivity are not the best. Rather, one should augment these models by intrinsic modulations. Our results (Fig. 7) suggest that
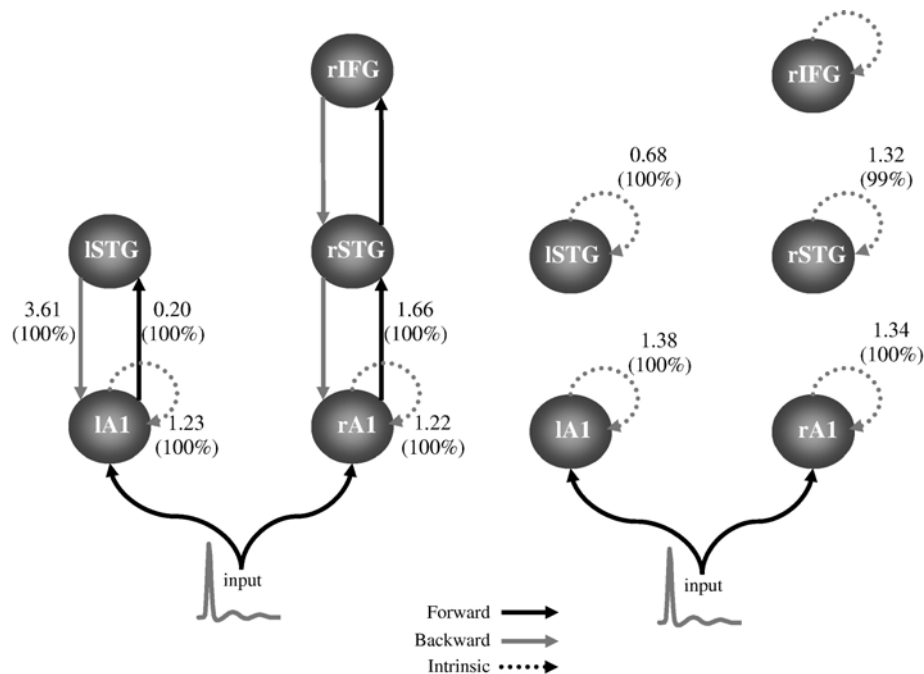
Fig. 8. Mismatch data: These two graphs show the posterior means of gains with a conditional probability of greater than 95% of being present. Left: FBI model. Right: adaptation model (**A**).

only the first two auditory (**A1**) areas need additional intrinsic modulation, but this will be examined more thoroughly, using multiple subject analyses. We also found that a pure 'local adaptation' model attained the highest model evidence. Although our results were computed on grand mean data, this points to the 'explanatory' power of intrinsic mechanisms: The number of parameters in these models is relatively low because it scales with the number of sources. Conversely, the number of extrinsic parameters increases quadratically with the number of sources. The local adaptation model explains the data well, so models with more (e.g., extrinsic) parameters must counter their increased complexity with a better fit. In the case of the combined **FBI** model, the increased complexity and accuracy were balanced so that there was no real difference between the simple adaptation (**A**) and the more complicated (**FBI**) model.

The interesting difference between intrinsic and extrinsic mechanisms seems to lie in the modulation of backward connectivity. A modulation of backward connectivity can explain a change in the time-course of a lower area, late in peri-stimulus time. For example, in our second simulations, there is a late change in the dynamics of the first area, but nearly no change in the dynamics of the second area. This effect is unique to a modulation of backward connectivity and cannot be explained by intrinsic mechanisms. We are currently pursuing this by looking at the log-evidences of models for early and late ERP components.

Conclusion

We have shown that the modulation of intrinsic connectivity is a useful extension of DCM for evoked responses. Mechanistic theories that invoke local adaptation of neuronal populations can be explicitly tested and compared with competing hypotheses.

Software note

All procedures described in this note have been implemented as Matlab (MathWorks) code. The source code is freely available as the 'DCM for evoked responses' toolbox of the Statistical Parametric Mapping package (SPM5) under http://www.fil.ion. ucl.ac.uk/spm/.

Acknowledgments

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2007.02.046.

References

Baillet, S., Mosher, J.C., Leahy, R.M., 2001. Electromagnetic brain mapping. IEEE Signal Process. Mag. 18, 14–30.

David, O., Friston, K.J., 2003. A neural mass model for MEG/EEG: coupling and neuronal dynamics. NeuroImage 20, 1743–1755.

David, O., Harrison, L., Friston, K.J., 2005. Modelling event-related responses in the brain. NeuroImage 25, 756–770.

David, O., Kiebel, S.J., Harrison, L.M., Mattout, J., Kilner, J.M., Friston, K.J., 2006. Dynamic causal modeling of evoked responses in EEG and MEG. NeuroImage 30, 1255–1272.

DeFelipe, J., Alonso-Nanclares, L., Arellano, J.I., 2002. Microstructure of the neocortex: comparative aspects. J. Neurocytol. 31, 299–316.

Doeller, C.F., Opitz, B., Mecklinger, A., Krick, C., Reith, W., Schroger, E., 2003. Prefrontal cortex involvement in preattentive auditory deviance

detection: neuroimaging and electrophysiological evidence. Neuro-Image 20, 1270–1282.

Fahrmeir, L., Tutz, G., 1994. Multivariate Statistical Modelling Based on Generalized Linear Models. Springer-Verlag, New York.

Felleman, D.J., Van Essen, D.C., 1991. Distributed hierarchical processing in the primate cerebral cortex. Cereb. Cortex 1, 1–47.

Friston, K.J., 2002. Bayesian estimation of dynamical systems: an application to fMRI. NeuroImage 16, 513–530.

Friston, K., 2005. A theory of cortical responses. Philos. Trans. R. Soc., B Biol. Sci. 360, 815–836.

Friston, K.J., Harrison, L., Penny, W., 2003. Dynamic causal modelling. NeuroImage 19, 1273–1302.

Friston, K.J., Ashburner, J., Kiebel, S.J., Nichols, T., Penny, W.D., 2006. Functional Brain Imaging: Statistical Parametric Mapping. Elsevier.

Garrido, M., Kilner, J.M., Kiebel, S.J., Stephan, K.E., Friston, K.J., 2006. Dynamic causal modeling of the mismatch negativity. Ref Type: Unpublished Work.

Jaaskelainen, I.P., Ahveninen, J., Bonmassar, G., Dale, A.M., Ilmoniemi, R.J., Levanen, S., Lin, F.H., May, P., Melcher, J., Stufflebeam, S., Tiitinen, H., Belliveau, J.W., 2004. Human posterior auditory cortex gates novel sounds to consciousness. Proc. Natl. Acad. Sci. U. S. A. 101, 6809–6814.

Jansen, B.H., Rit, V.G., 1995. Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. Biol. Cybern. 73, 357–366.

Kiebel, S.J., David, O., Friston, K.J., 2006. Dynamic causal modelling of evoked responses in EEG/MEG with lead field parameterization. NeuroImage 30, 1273–1284.

May, P., Tiitinen, H., Ilmoniemi, R.J., Nyman, G., Taylor, J.G., Naatanen, R., 1999. Frequency change detection in human auditory cortex. J. Comput. Neurosci. 6, 99–120.

Mosher, J.C., Leahy, R.M., Lewis, P.S., 1999. EEG and MEG: forward solutions for inverse methods. IEEE Trans. Biomed. Eng. 46, 245–259.

Naatanen, R., 2003. Mismatch negativity: clinical research and possible applications. Int. J. Psychophysiol. 48, 179–188.

Naatanen, R., Jacobsen, T., Winkler, I., 2005. Memory-based or afferent processes in mismatch negativity (MMN): a review of the evidence. Psychophysiology 42, 25–32.

Opitz, B., Rinne, T., Mecklinger, A., von Cramon, D.Y., Schroger, E., 2002. Differential contribution of frontal and temporal cortices to auditory change detection: fMRI and ERP results. NeuroImage 15, 167–174.

Penny, W.D., Stephan, K.E., Mechelli, A., Friston, K.J., 2004. Comparing dynamic causal models. NeuroImage 22, 1157–1172.

Peters, A., 2002. Examining neocortical circuits: some background and facts. J. Neurocytol. 31, 183–193.

Rao, R.P., Ballard, D.H., 1999. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat. Neurosci. 2, 79–87.

Winkler, I., Karmos, G., Naatanen, R., 1996. Adaptive modeling of the unattended acoustic environment reflected in the mismatch negativity event-related potential. Brain Res. 742, 239–252.