Published in final edited form as:
Wiley Interdiscip Rev Cogn Sci. 2010 June ; 1(3): 446–459. doi:10.1002/wcs.58.

Analyzing effective connectivity with fMRI

Klaas Enno Stephan[1,2] and **Karl J. Friston**[2]

[1] Laboratory for Social and Neural Systems Research, Institute for Empirical Research in Economics, University of Zurich, Switzerland. [2] Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London, London, UK.

Abstract

Functional neuroimaging techniques are used widely in cognitive neuroscience to investigate aspects of functional specialization and functional integration in the human brain. Functional integration can be characterized in two ways, functional connectivity and effective connectivity. While functional connectivity describes statistical dependencies between data, effective connectivity rests on a mechanistic model of the causal effects that generated the data. This review addresses the conceptual and methodological basis of established techniques for characterizing effective connectivity using functional magnetic resonance imaging (fMRI) data. In particular, we focus on dynamic causal modeling (DCM) of fMRI data and emphasize the importance of model selection procedures and nonlinear mechanisms for context-dependent changes in connection strengths.

Keywords

dynamic causal modeling; DCM; Bayesian model selection; BMS; synaptic plasticity; gain control; attention; binocular rivalry; neuromodulation

Introduction

Functional integration in neuronal systems can be quantified in two ways, functional connectivity and effective connectivity [1-3]. While functional connectivity only describes statistical dependencies between spatially segregated neuronal events, effective connectivity rests on a mechanistic model of how the data were caused. This article reviews established techniques for characterizing effective connectivity on the basis of fMRI data, focusing on dynamic causal models (DCMs; [4, 5]).

Effective connectivity

The term *effective connectivity* has been defined by various authors in convergent ways. A general definition is that effective connectivity describes the causal influences that neural units exert over another [1]. More specifically, other authors have proposed that “effective connectivity should be understood as the experiment- and time-dependent, simplest possible circuit diagram that would replicate the observed timing relationships between the recorded neurons” [6]. Both definitions emphasize that determining effective connectivity requires a causal model of the interactions between the elements of the neural system of interest.

Address for correspondence: Klaas Enno Stephan, MD Dr.med. PhD Laboratory for Social and Neural Systems Research, Institute for Empirical Research in Economics, University of Zurich, Blümlisalpstr. 10, 8006 Zurich, Switzerland phone +41-44-6345220 fax +41-44-6344907 k.stephan@fil.ion.ucl.ac.uk.

Europe PMC Funders Author Manuscripts

Such causal models can be defined within the general mathematical framework provided by dynamic systems theory [7-9]. A *system* is characterised by time-variant properties x_i ($1 \leq i \leq n$) or *state variables*, which interact with each other, *i.e.* the evolution of each state variable depends on at least one other state variable. For example, the postsynaptic membrane potential depends on which and how many ion channels are open; vice versa, the probability of voltage-dependent ion channels opening depends on the membrane potential. Such functional dependencies can be expressed quite naturally by a set of ordinary differential equations in which a set of parameters θ determine the form and strength of the causal influences between the state variables. In neural systems, these parameters usually include time constants or synaptic strengths of the connections between the system elements. Additionally, in the case of non-autonomous systems (*i.e.* systems that exchange matter, energy or information with their environment) we need to consider the inputs into the system, *e.g.* sensory information entering the brain. Representing the set of all m known inputs by the m-vector function $u(t)$, one can define a general state equation for non-autonomous deterministic systems:

$$\frac{dx}{dt} = F(x, u, \theta) \quad (1)$$

A model whose form follows this general state equation provides a causal description of how system dynamics results from system structure, because it describes (i) when and where external inputs enter the system and (ii) how the state changes induced by these inputs evolve in time depending on the system's structure. Given a particular temporal sequence of inputs $u(t)$ and an initial state $x(0)$, one obtains a complete description of how the dynamics of the system (*i.e.* the trajectory of its state vector x in time) results from its structure by integration of Equation 1:

$$x(\tau) = x(0) + \int_0^{\tau} F(x, u, \theta)\, dt \quad (2)$$

Equation 2 therefore provides a general form for models of effective connectivity in neural systems. (It assumes that all processes in the system are deterministic and occur instantaneously, but can easily be extended, *e.g.* by using stochastic and delay differential equations, respectively [10, 11]). The framework outlined here is concerned with dynamic systems in continuous time and thus uses differential equations. The same basic ideas, however, can also be applied to dynamic systems in discrete time (using difference equations), *e.g.* multivariate/vector autoregressive models (MAR/VAR; [12-14], as well as to "static" systems where the system is at equilibrium at each point of observation. The latter perspective applies to regression-based system models for functional neuroimaging data, *e.g.* psycho-physiological interactions (PPI; [15]), or structural equation modeling (SEM; [16-19]). Readers interested in these classical approaches are referred to the original articles referenced above and to reviews that have compared these approaches (*e.g.* [7, 20]). Here, we focus on that framework for inferring effective connectivity from fMRI data that most closely follows Equation 2 above, *i.e.* dynamic causal modeling ([4, 5]).

Dynamic Causal Modelling (DCM)

An important limitation of classical models of effective connectivity like PPI, SEM or VAR is that they operate at the level of the measured signals. This is a serious problem because the causal architecture of the system that we would like to identify is located at the neuronal level which cannot be investigated directly using non-invasive techniques. In the case of fMRI data, for example, PPI, SEM and VAR are fitted to measured time series which result

from a haemodynamic convolution of the underlying neuronal activity. The absence of a forward model linking neuronal activity to the measured haemodynamic data can render analyses of inter-regional connectivity problematic. For example, different brain regions can exhibit marked differences in neurovascular coupling. It has been shown that these inter-regional differences can lead to false inference about effective connectivity [21]. A similar problem exists for EEG data where changes in neural activity in different brain regions lead to changes in electric potentials that superimpose linearly. The scalp electrodes therefore record a mixture, with unknown weightings, of potentials generated by a number of different sources.

Therefore, to enable inferences about connectivity between neural units we need models that combine two things: (i) a parsimonious but neurobiologically plausible model of neural population dynamics, and (ii) a biophysically plausible forward model that describes the transformation from neural activity to the measured signal (c.f. [13, 22]). Such models make it possible to fit jointly the parameters of the neural and of the forward model such that the predicted time series are optimally similar to the observed time series. In principle, any of the models described above could be combined with a modality-specific forward model, and indeed, VAR models have previously been combined with linear forward models to explain EEG data [23]. So far, however, DCM is the only approach where the marriage between models of neural dynamics and biophysical forward models is a mandatory component.

Since its original inception for fMRI [4], a variety of DCM implementations have been introduced for additional data modalities, including event-related potentials [11, 24] induced responses [25, 26], auto- and cross-spectral densities [27, 28] and phase coupling [29] as measured by local field potential recordings or EEG/MEG. These models, all formulated under the same theoretical framework, have enjoyed considerable success in the practical analysis of neuroimaging data, resulting in more than 100 published studies (as of August 2009). In this chapter, we focus on DCM for fMRI as originally described [4] and on some recent nonlinear extensions of this model [30].

DCM for fMRI uses a simple model of neural dynamics in a system of n interacting brain regions (see Figure 1 for a schematic summary). In its classical form [4], it models the change of a neural state vector x in time, with each region in the system being represented by a single state variable (representing mean regional activity), using the following bilinear differential equation:

$$\begin{aligned} \frac{dx}{dt} &= F(x, u, \theta^n) \\ &= \left(A + \sum_{i=1}^{m} u_j B^{(i)}\right) x + Cu \end{aligned} \quad (3)$$

Note that this neural state equation follows the general form for deterministic system models introduced by Equation 2, *i.e.* the modelled state changes are a function of the system state itself, the inputs u and some parameters $\theta^{(n)}$ that define the functional architecture and interactions among brain regions at a neuronal level. The neural state variables represent a summary index of neural population dynamics in the respective regions. The neural dynamics are driven by experimentally controlled external inputs that can enter the model in two different ways: they can elicit responses through direct influences on specific regions (*e.g.* evoked responses in early sensory cortices; the C matrix) or they can modulate the coupling among regions (*e.g.* during learning or attention; the B matrices). Note that Eq. 3 does not account for conduction delays in either inputs or inter-regional influences. This is not necessary because, due to the large regional variability in hemodynamic response latencies, fMRI data do not posses enough temporal information to enable estimation of inter-regional axonal conduction delays which are typically in the order of 10-20 ms (note

that the differential latencies of the hemodynamic response are accommodated by region-specific biophysical parameters in the hemodynamic model described below). This was verified by Friston *et al.* [4] who showed in simulations that DCM parameter estimates were not affected by introducing artificial delays of up to ± 1 second. In contrast, conduction delays are an important part of DCM for event-related potentials [11].

Given the bilinear state equation (Equation 3), the neural parameters $\theta^{(n)} = \{A, B, C\}$ can be expressed as partial derivatives of F:

$$A = \left.\frac{\partial F}{\partial x}\right|_{u=0}$$

$$B^{(i)} = \frac{\partial^2 F}{\partial x \partial u_i} \quad (4)$$

$$C = \left.\frac{\partial F}{\partial u}\right|_{x=0}$$

As can be seen from these equations, the matrix A represents the endogenous (fixed) connectivity among the regions in the absence of input, the matrices $B^{(i)}$ encode the change in connectivity induced by the i^{th} input u_i, and C embodies the strength of exogenous (direct) influences of inputs on neuronal activity. In most instances, the parameters of primary interest are the modulatory ones (*i.e.* the matrices $B^{(i)}$) since they encode how experimentally controlled manipulations change the connection strengths in the system.

DCM for fMRI combines this model of neural dynamics with an experimentally validated haemodynamic model that describes the transformation of neuronal activity into a BOLD response. This haemodynamic model, which builds on the so-called "Balloon model" [31], consists of a set of differential equations that describe, using a set of parameters $\theta^{(h)}$, how changes in neural activity elicit changes in a vasodilatory signal, blood flow, blood volume and deoxyhemoglobin content [32]. The predicted BOLD signal is a non-linear function of blood volume and deoxyhemoglobin content [33]. The most recent version of this haemodynamic model is summarised by Figure 2 and described in detail by Stephan *et al.* [33].

The combined neural and haemodynamic parameter set $\theta = \{\theta^{(n)}, \theta^{(n)}\}$ is estimated from the measured BOLD data, using a fully Bayesian approach with empirical priors for the haemodynamic parameters and conservative shrinkage priors for the coupling parameters. Details of the parameter estimation scheme, which rests on a fixed-form variational Bayesian algorithm, using a Laplace (*i.e.* Gaussian) approximation to the true posterior, can be found elsewhere [4, 34, 35].

Inference about neuronal mechanisms with DCM

Once the parameters of a DCM have been estimated from measured BOLD data, the posterior distributions of the parameter estimates can be used to test hypotheses about connection strengths. Due to the Laplace approximation, the posterior distributions are defined by their maximum a posteriori (MAP) estimate and their posterior covariance. Usually, the hypotheses to be tested concern context-dependent changes in coupling (*i.e.* the matrices $B^{(i)}$ in Eq. 3). An example, originally reported in [36], is given by Figure 3. Here, DCM was applied to fMRI data from a single subject, testing the hypothesis that in the ventral stream of the visual system a letter decision task increased the strength of interhemispheric connections, but only when the word stimuli were presented in the left visual field and were thus initially received by the non-dominant right hemisphere, necessitating transfer of stimulus information to the specialised left hemisphere. This

hypothesis was tested by constructing a four-area model of ventral stream areas, comprising the lingual and fusiform gyri in both hemispheres (Figure 3A), and comparing the modulatory influences of task, conditional on the visual field of stimulus presentation, for interhemispheric connections in both directions. This comparison, based on the MAP estimates and the posterior covariances of the modulatory parameters, indicated that for this particular subject and for the connections between left and right lingual gyrus the hypothesised asymmetry in interhemispheric transfer existed with a probability of 98.7% (Figure 3B). Other examples of single-subject analyses can be found in [4, 20, 30, 37].

For statistical inference at the group level, various options exist. One commonly used approach, corresponding to a random effects analysis, is to enter the conditional estimates of interest into a classical second-level analysis, *e.g.* a *t*-test on the MAP estimates of a particular parameter across subjects (for examples, see [38-41]). An alternative approach is to use Bayesian statistics at the group level as well. This can be done by computing, for a given parameter, one joint posterior density across all subjects, treating the posterior of one subject as the prior for the next [42]. This approach can be more sensitive; its disadvantage, however, is that it corresponds to a fixed effects analysis and thus does not allow for inference beyond the particular group studied.

Bayesian model selection (BMS)

Model comparison and selection is central to the scientific process, in that it allows one to evaluate different hypotheses about the way data are caused [43, 44]. Nearly all scientific reporting rests upon some form of model comparison, which represents a probabilistic statement about the beliefs in one hypothesis relative to some other(s), given some observations or data. In other words: Given some observed data, which of several alternative models is optimal? The decision cannot be made solely by comparing the relative fit of competing models. One also needs to account for differences in complexity; *i.e.*, the number of free parameters and the degree of their inter-dependency. This is important because as model complexity increases, fit increases monotonically, but at some point the model will start fitting noise that is specific to the particular data (*i.e.*, "over-fitting") and thus becomes less generalizable across multiple realizations of the same underlying generative process. Therefore, the question "What is the optimal model?" can be reformulated as "What is the model that represents the best balance between fit and complexity?" This is the model that maximizes the model evidence:

$$p(y|m) = \int p(y|\theta, m)\, p(\theta|m)\, d\theta \quad (5)$$

Here, the numbers of free parameters (as well as the functional form of the generative model that determines their interdependencies) are subsumed by the integration. Unfortunately, this integral cannot usually be solved analytically; therefore an approximation to the (log of the) model evidence is used instead. This approximation is usually a free energy bound on the log evidence [35]; alternatively, simpler criteria like the Akaike Information Criterion [45] or the Bayesian Information Criterion [46] can be used that are blind to parameter interdependencies (see [47] for a detailed discussion). Given any of these approximations to the log evidence of two models m_i and m_j, the difference in log evidence can be transformed into a *Bayes factor* (*BF*):

$$BF_{ij} = \frac{p(y|m_i)}{p(y|m_i)} \approx \exp\left(F_i - F_j\right) \quad (6)$$

BMS can be applied both to single subjects and whole groups. When the optimal model structure is expected to vary across subjects (*e.g.*, subject-specific cognitive strategies or

different pathophysiological mechanisms in a group of patients), random effects BMS is required. This method rests on a hierarchical model which is optimised to furnish a probability density on the models themselves, using variational Bayes [47]. Specifically, it estimates the parameters of a Dirichlet distribution describing the probabilities for all models considered. These probabilities then define a multinomial distribution over model space, allowing one to compute how likely it is that a specific model generated the data of a randomly chosen subject as well as the exceedance probability of one model being more likely than any other model.

BMS plays a central role for DCM. It is used routinely to select the most likely model amongst a set of alternatives before making inferences about particular parameters, *e.g.*, [33, 38, 42, 48-54]. An alternative use of model selection is to decide about the nature of particular mechanisms without the need for any further inference about particular parameters. For example, BMS has been used to compare DCMs with non-linear versus linear BOLD equations in the haemodynamic forward model [33, 47] or to disambiguate between different possibilities how anatomical connection strength constrains effective connection strength [55]. A particularly interesting approach is to go beyond the comparison of specific models and compare two (or more) partitions of model space [47]. These partitions would typically reflect those components of model structure that one seeks inference about, *e.g.* whether a specific connection should be included in the model or not, whether a particular connection is modulated by one experimental condition or another, or whether certain effects are linear or nonlinear. The advantage of this method is that arbitrarily large set of models can be considered together, allowing one to integrate out uncertainty over any aspect of model structure other than the component of interest.

Nonlinear DCM for fMRI

Since its first description [4], DCM for fMRI has been extended in several ways. For example, an extension of the observation equation takes into account the slice-specific sampling times in multi-slice MRI acquisitions [56]. This enables DCM to be applied to fMRI data from any data acquisition scheme. Another variant represents each region in the model by two state variables and distinguishes between population activity of excitatory and inhibitory neurons [57]. Other work has augmented DCM with a spatial model of the regional time series to which the model is fitted [58].

Here, we focus on what we consider to be a particularly important extension of DCM for fMRI, namely the inclusion of nonlinear modulatory effects [30]. This extension was motivated by two limitations of the original bilinear neuronal state equation in DCM. First, the neuronal origin of the modulatory influence is not specified. Second, the bilinear framework may not be the most appropriate choice for modelling fast changes in effective connectivity, which are mediated by nonlinear effects at the level of single neurons. These mechanisms are instances of “short-term synaptic plasticity” (STP), an umbrella term for a range of processes which alter synaptic strengths with time constants in the range of milliseconds to minutes; *e.g.* NMDA-controlled rapid trafficking or phosphorylation of AMPA receptors, synaptic depression/facilitation or “early LTP”. All these processes are driven by the history of prior synaptic activity and are thus nonlinear [59].

A particularly interesting mechanism, which relies on STP, is “neuronal gain control”. Neuronal gain, *i.e.* the response of a given neuron N_1 to presynaptic input from a second neuron N_2, depends on the history of inputs that N_1 receives from other neurons, e.g. a third neuron N_3. Such a nonlinear modulation or “gating” of the $N_2 \rightarrow N_1$ connection by N_3 has been shown to have the same mathematical form across a large number of experiments (for review, see [60]): the change in the gain of N_1 results from a multiplicative interaction

among the synaptic inputs from N_2 and N_3, *i.e.* a second-order nonlinear effect. Biophysically, neuronal gain control can arise through various mechanisms that mediate interactions among synaptic inputs occurring close in time (see [30] for a discussion of these mechanisms).

Critically, the bilinear framework precludes a representation, at the neuronal level, of the mechanisms described above. As stated in the original DCM paper [4], in order to model processes like neuronal gain control and synaptic plasticity properly, one needs "to go beyond bilinear approximations to allow for interactions among the states. This is important when trying to model modulatory or nonlinear connections such as those mediated by backward afferents that terminate predominantly in the supragranular layers and possibly on NMDA receptors."

Therefore, to enable a realistic representation of how neuronal populations modulate the gains of other populations, one needs to model nonlinear interactions amongst the n states of a given DCM. For this purpose, one can use a two-dimensional Taylor series which is of second order in the states [30]:

$$\begin{aligned} f(x,u) &= \frac{dx}{dt} \\ &\approx f(0,0) + \frac{\partial f}{\partial x}x + \frac{\partial f}{\partial u}u + \frac{\partial^2 f}{\partial x \partial u}xu + \frac{\partial^2 f}{\partial x^2}\frac{x^2}{2} \end{aligned} \quad (7)$$

Setting $D^{(j)} = \frac{1}{2}\frac{\partial^2 f}{\partial x_j^2}\Big|_{u=0} \quad (1 \le j \le n)$ makes Eq. 7 equivalent to:

$$f(x,u) = \frac{dx}{dt} = \left(A + \sum_{i=1}^{m} u_i B^{(i)} + \sum_{j=1}^{n} x_j D^{(j)}\right)x + Cu \quad (8)$$

Here, the $D^{(j)}$ matrices encode which of the n regions gate which connections in the system. Specifically, any non-zero entry $D_{kl}^{(j)}$ indicates that responses of region k to inputs from region l depend on activity in region j. Figure 4 shows a simple example, with synthetic data generated by a nonlinear DCM. This illustrates the sort of dynamics, both at the neuronal and hemodynamic level which this sort of model exhibits.

The nonlinear extension enhances the kind of dynamics that DCM can capture and enables the user to implement additional types of models. Beyond modelling how connection strengths are modulated by external inputs, one can now model how connection strengths are gated by the activity of one or several neuronal populations. This ability is critical for various applications, *e.g.* for marrying reinforcement learning models with DCM [7], but also for mechanistic accounts of the effects of attention. For example, nonlinear DCM was applied to a single-subject data set from a blocked fMRI study of attention to visual motion [17]. Four different models were compared [30], each of which embodied a different explanation for the empirical finding that V5 responses increased during attention to motion, compared to unattended motion. The most likely model was one in which the gain of the V1→V5 connection depended on the activity in the posterior parietal cortex (PPC), a region on which attention exerted a direct effect (this could result, for example, from cholinergic inputs from the brainstem [61]). Analysis of the posterior density of the modulatory parameter in this model indicated that nonlinear gating of the V1→V5 connection by attention could be inferred with 99.1% confidence (see Figure 5). Figure 6 shows the observed and fitted time-series of all areas and highlights the attentional gating effect on V5 activity, such that V5 activity was higher when subjects attended the moving stimuli.

Europe PMC Funders Author Manuscripts

As a second example for the practical utility of nonlinear DCMs, we show the results from a single subject analysis of fMRI data set acquired during an event-related binocular rivalry paradigm [30]. While there is no clear consensus about the mechanisms that underlie binocular rivalry, it has been suggested that it (i) depends on nonlinear mechanisms and (ii) may arise from modulation of connections amongst neuronal representations of the competing stimuli by feedback connections from higher areas [62].

The fMRI data were acquired during a factorial paradigm in which face and house stimuli were presented either during binocular rivalry or during a matched non-rivalry (*i.e.* replay) condition. For the subject studied here, the conventional SPM analysis showed a rivalry × percept interaction in both the right fusiform face area (FFA) and the right parahippocampal place area (PPA): in FFA, the face vs. house contrast was higher during non-rivalry than during rivalry; conversely, in PPA the house vs. face contrast was higher during non-rivalry than during rivalry (both $p<0.05$, small-volume corrected). Additionally, testing for a main effect of rivalry, we replicated previous findings that the right middle frontal gyrus (MFG) showed higher activity during rivalry than during non-rivalry conditions [63].

These SPM results motivated a nonlinear DCM in which the connections between FFA and PPA were modulated by the activity in the MFG (Figure 7). First, the fixed (endogenous) connection strengths between FFA and PPA were negative in both directions, *i.e.* FFA and PPA exerted a mutual negative influence on each other; this could be regarded as a “tonic” or “baseline” reciprocal inhibition. More importantly, however, was that during the presentation of visual stimuli this competitive interaction between FFA and PPA was modulated by activity in the middle frontal gyrus (MFG), which showed higher activity during rivalry vs. non-rivalry conditions. As shown in Figure 7, our confidence about the presence of this nonlinear modulation was very high (99.9%) for both connections.

According to this model, activity levels in the MFG determine activity in FFA and PPA by controlling the influence that face-elicited activations and house-elicited deactivations of FFA have on PPA (and vice versa). For example, the positive nonlinear modulation of the FFA→PPA connection by MFG activity (see Figure 7) means that during face-perception under rivalry conditions (which elicit positive activity in the FFA and MFG, respectively) there is a positive influence of FFA on PPA, overriding the “baseline” inhibition. This means that during binocular rivalry FFA and PPA become more tightly coupled which destroys their stimulus selectivity: their activity becomes very similar, regardless of whether a face or a house is being perceived. In contrast, deactivation of MFG during non-rivalry conditions decreases the influence that FFA has on PPA during house perception; therefore responses in FFA and PPA become less coupled and their relative selectivity for face and house percepts is restored. This dynamic coupling and uncoupling, leading to less selectivity of FFA and PPA during rivalry and higher selectivity during non-rivalry, is clearly visible in Figure 8 which plots the observed and fitted responses of all three areas. Here, the short black arrows indicate blocks with binocular rivalry (when FFA and PPA show very similar time courses) and the long grey arrows denote non-rivalry blocks (when FFA and PPA activities evolve more independently). These changes in effective connectivity over time, which are controlled by the activity level in MFG, provide a nice explanation for the rivalry × percept interaction in FFA and PPA that was identified by the SPM analysis.

Conclusions

In this short review, we have outlined how effective connectivity can be inferred from fMRI data using DCM. We expect that two application domains for DCM will prove to be particularly exciting and fruitful in the near future. The first domain is the integration of the neurophysiological and computational aspects of learning and decision making. For

example, according to theoretical models of learning, the size of prediction errors should control synaptic plasticity, i.e. changes in the strength of synaptic connections, encoding stimulus-stimulus and stimulus-response links [64-66]. In other words, the necessity of reconfiguring neuronal circuits during learning should be inversely proportional to how well those neuronal circuits are capable of predicting sensory stimuli or outcomes of actions. This notion can be tested formally by embedding prediction errors provided by computational models of learning (such as Rescorla-Wagner or temporal difference learning models) into DCMs. A first demonstration of this approach was given by a recent study which combined DCM with a Rescorla-Wagner model and showed that during incidental audio-visual associative learning the plasticity of connections from auditory to visual cortex depended on trial-by-trial prediction errors [67]. A subsequent study extended this finding: combining nonlinear DCM and a hierarchical Bayesian learner, it showed that the degree of trial-by-trial prediction error activity in the putamen controlled the efficacy of visuomotor connections, thus gating the transfer of sensory information depending on how unexpected this information was (den Ouden *et al.*, submitted).

The second application domain concerns the development of DCMs with clinical utility, for example as diagnostic tools. Although DCM has already been applied to some clinical questions (*e.g.*, [39, 68, 69], the critical challenge for the future will be to develop DCMs whose parameter estimates have sufficient sensitivity and specificity to delineate subgroups of patients that are characterized by different pathophysiological mechanisms. This generic framework of model-based inference about pathophysiological processes that cannot be measured directly is likely to be particularly helpful for vaguely defined spectrum diseases. For example, our own work focuses on schizophrenia, trying to establish DCMs, in conjunction with pharmacological challenges and learning paradigms, that can detect specific abnormalities in the regulation of NMDA-dependent synaptic plasticity by neuromodulatory transmitters like dopamine or acetylcholine [70]. Hopefully, neurocomputational models of specific learning and decision-making processes (such as the work by den Ouden *et al.* described above) can be established whose parameters map onto well-defined physiological mechanisms of synaptic plasticity and neuromodulation. These models are not restricted to fMRI, but will also exploit electrophysiological measurements. Careful validation of these models is crucial and will require pharmacological and invasive recording studies in animals. For example, a recent rodent study demonstrated that DCM can correctly infer experimentally induced changes in spike frequency adaptation and postsynaptic efficacy of glutamatergic synapses [27].

Importantly, however, model-based inference on pathophysiology and disease status cannot only proceed on the basis of neurophysiologically interpretable parameter estimates, but could also employ BMS to compare entire models embodying different putative disease mechanisms. This inference on model structure could be particularly useful when disease subgroups differ along more than one pathophysiological dimension.

Acknowledgments

This work was funded by the Wellcome Trust (KJF) and the University Research Priority Program “Foundations of Human Social Behaviour” at the University of Zurich (KES).

References

1. Friston KJ. Functional and effective connectivity in neuroimaging: a synthesis. Hum. Brain Mapp. 1994; 2:56–78.
2. Horwitz B, Tagamets MA, McIntosh AR. Neural modeling, functional brain imaging, and cognition. Trends Cogn Sci. Mar; 1999 3(no. 3):91–98. [PubMed: 10322460]

3. Stephan KE, Riera JJ, Deco G, Horwitz B. The Brain Connectivity Workshops: moving the frontiers of computational systems neuroscience. Neuroimage. Aug 1; 2008 42(no. 1):1–9. [PubMed: 18511300]
4. Friston KJ, Harrison L, Penny W. Dynamic causal modelling. Neuroimage. Aug; 2003 19(no. 4): 1273–302. [PubMed: 12948688]
5. Stephan KE, Harrison LM, Kiebel SJ, David O, Penny WD, Friston KJ. Dynamic causal models of neural system dynamics:current state and future extensions. J Biosci. Jan; 2007 32(no. 1):129–44. [PubMed: 17426386]
6. Aertsen, A.; Preißl, H. Dynamics of activity and connectivity in physiological neuronal networks. In: HG, S., editor. Nonlinear Dynamics and Neuronal Networks. VCH Publishers; New York: 1999. p. 281-302.
7. Stephan KE. On the role of general system theory for functional neuroimaging. J Anat. Dec; 2004 205(no. 6):443–70. [PubMed: 15610393]
8. Jirsa VK. Connectivity and dynamics of neural information processing. Neuroinformatics. Sum; 2004 2(no. 2):183–204. [PubMed: 15319516]
9. Breakspear M. "Dynamic" connectivity in neural systems: theoretical and empirical considerations. Neuroinformatics. 2004; 2(no. 2):205–26. [PubMed: 15319517]
10. Friston KJ, Trujillo-Barreto N, Daunizeau J. DEM: a variational treatment of dynamic systems. Neuroimage. Jul 1; 2008 41(no. 3):849–85. [PubMed: 18434205]
11. David O, Kiebel SJ, Harrison LM, Mattout J, Kilner JM, Friston KJ. Dynamic causal modeling of evoked responses in EEG and MEG. Neuroimage. May 1; 2006 30(no. 4):1255–72. [PubMed: 16473023]
12. Harrison L, Penny WD, Friston K. Multivariate autoregressive modeling of fMRI time series. Neuroimage. Aug; 2003 19(no. 4):1477–91. [PubMed: 12948704]
13. Goebel R, Roebroeck A, Kim DS, Formisano E. Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modeling and Granger causality mapping. Magn Reson Imaging. Dec; 2003 21(no. 10):1251–61. [PubMed: 14725933]
14. Roebroeck A, Formisano E, Goebel R. Mapping directed influence over the brain using Granger causality and fMRI. Neuroimage. Mar; 2005 25(no. 1):230–42. [PubMed: 15734358]
15. Friston KJ, Buechel C, Fink GR, Morris J, Rolls E, Dolan RJ. Psychophysiological and modulatory interactions in neuroimaging. Neuroimage. Oct; 1997 6(no. 3):218–29. [PubMed: 9344826]
16. McIntosh AR, Gonzalez-Lima F. Structural modeling of functional neural pathways mapped with 2-deoxyglucose: effects of acoustic startle habituation on the auditory system. Brain Res. May 3; 1991 547(no. 2):295–302. [PubMed: 1884204]
17. Büchel C, Friston KJ. Modulation of connectivity in visual pathways by attention: cortical interactions evaluated with structural equation modelling and fMRI. Cereb Cortex. 1997; 7(no. 8): 768–78. [PubMed: 9408041]
18. Bullmore E, Horwitz B, Honey G, Brammer M, Williams S, Sharma T. How good is good enough in path analysis of fMRI data? Neuroimage. Apr; 2000 11(no. 4):289–301. [PubMed: 10725185]
19. McIntosh AR, Gonzales-Lima F. Structural equation modelling and its application to network analysis in functional brain imaging. Hum Brain Mapp. 1994; 2:2–22.
20. Penny WD, Stephan KE, Mechelli A, Friston KJ. Modelling functional integration: a comparison of structural equation and dynamic causal models. Neuroimage. 2004; 23(Suppl 1):S264–74. [PubMed: 15501096]
21. David O, Guillemain I, Saillet S, Reyt S, Deransart C, Segebarth C, Depaulis A. Identifying neural drivers with functional MRI: an electrophysiological validation. PLoS Biol. Dec 23; 2008 6(no. 12):2683–97. [PubMed: 19108604]
22. Stephan KE, Harrison LM, Penny WD, Friston KJ. Biophysical models of fMRI responses. Curr Opin Neurobiol. Oct; 2004 14(no. 5):629–35. [PubMed: 15464897]
23. Yamashita O, Galka A, Ozaki T, Biscay R, Valdes-Sosa P. Recursive penalized least squares solution for dynamical inverse problems of EEG generation. Hum Brain Mapp. Apr; 2004 21(no. 4):221–35. [PubMed: 15038004]
24. Kiebel SJ, David O, Friston KJ. Dynamic causal modelling of evoked responses in EEG/MEG with lead field parameterization. Neuroimage. May 1; 2006 30(no. 4):1273–84. [PubMed: 16490364]

25. Chen CC, Kiebel SJ, Friston KJ. Dynamic causal modelling of induced responses. Neuroimage. Jul 15; 2008 41(no. 4):1293–312. [PubMed: 18485744]
26. Chen CC, Henson RN, Stephan KE, Kilner JM, Friston KJ. Forward and backward connections in the brain: a DCM study of functional asymmetries. Neuroimage. Apr 1; 2009 45(no. 2):453–62. [PubMed: 19162203]
27. Moran RJ, Stephan KE, Kiebel SJ, Rombach N, O'Connor WT, Murphy KJ, Reilly RB, Friston KJ. Bayesian estimation of synaptic physiology from the spectral responses of neural masses. Neuroimage. Aug 1; 2008 42(no. 1):272–84. [PubMed: 18515149]
28. Moran RJ, Stephan KE, Seidenbecher T, Pape HC, Dolan RJ, Friston KJ. Dynamic causal models of steady-state responses. Neuroimage. Feb 1; 2009 44(no. 3):796–811. [PubMed: 19000769]
29. Penny WD, Litvak V, Fuentemilla L, Duzel E, Friston K. Dynamic causal models for phase coupling. J Neurosci Methods. Jul 2.2009
30. Stephan KE, Kasper L, Harrison LM, Daunizeau J, den Ouden HE, Breakspear M, Friston KJ. Nonlinear dynamic causal models for fMRI. Neuroimage. Aug 15; 2008 42(no. 2):649–62. [PubMed: 18565765]
31. Buxton RB, Wong EC, Frank LR. Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. Magn Reson Med. Jun; 1998 39(no. 6):855–64. [PubMed: 9621908]
32. Friston KJ, Mechelli A, Turner R, Price CJ. Nonlinear responses in fMRI: the Balloon model, Volterra kernels, and other hemodynamics. Neuroimage. Oct; 2000 12(no. 4):466–77. [PubMed: 10988040]
33. Stephan KE, Weiskopf N, Drysdale PM, Robinson PA, Friston KJ. Comparing hemodynamic models with DCM. Neuroimage. Nov 15; 2007 38(no. 3):387–401. [PubMed: 17884583]
34. Friston KJ. Bayesian estimation of dynamical systems: an application to fMRI. Neuroimage. Jun; 2002 16(no. 2):513–30. [PubMed: 12030834]
35. Friston K, Mattout J, Trujillo-Barreto N, Ashburner J, Penny W. Variational free energy and the Laplace approximation. Neuroimage. Jan 1; 2007 34(no. 1):220–34. [PubMed: 17055746]
36. Stephan KE, Penny WD, Marshall JC, Fink GR, Friston KJ. Investigating the functional role of callosal connections with dynamic causal models. Ann N Y Acad Sci. Dec.2005 1064:16–36. [PubMed: 16394145]
37. Mechelli A, Price CJ, Noppeney U, Friston KJ. A dynamic causal modeling study on category effects: bottom-up or top-down mediation? J Cogn Neurosci. Oct 1; 2003 15(no. 7):925–34. [PubMed: 14628754]
38. Stephan KE, Marshall JC, Penny WD, Friston KJ, Fink GR. Interhemispheric integration of visual processing during task-driven lateralization. J Neurosci. Mar 28; 2007 27(no. 13):3512–22. [PubMed: 17392467]
39. Sonty SP, Mesulam MM, Weintraub S, Johnson NA, Parrish TB, Gitelman DR. Altered effective connectivity within the language network in primary progressive aphasia. J Neurosci. Feb 7; 2007 27(no. 6):1334–45. [PubMed: 17287508]
40. Bitan T, Booth JR, Choy J, Burman DD, Gitelman DR, Mesulam MM. Shifts of effective connectivity within a language network during rhyming and spelling. J Neurosci. Jun 1; 2005 25(no. 22):5397–403. [PubMed: 15930389]
41. Smith AP, Stephan KE, Rugg MD, Dolan RJ. Task and content modulate amygdala-hippocampal connectivity in emotional retrieval. Neuron. Feb 16; 2006 49(no. 4):631–8. [PubMed: 16476670]
42. Garrido MI, Kilner JM, Kiebel SJ, Stephan KE, Friston KJ. Dynamic causal modelling of evoked potentials: a reproducibility study. Neuroimage. Jul 1; 2007 36(no. 3):571–80. [PubMed: 17478106]
43. Pitt MA, Myung IJ. When a good fit can be bad. Trends Cogn Sci. Oct 1; 2002 6(no. 10):421–425. [PubMed: 12413575]
44. Penny WD, Stephan KE, Mechelli A, Friston KJ. Comparing dynamic causal models. Neuroimage. Jul; 2004 22(no. 3):1157–72. [PubMed: 15219588]
45. Akaike H. A new look at the statistical model identification. IEEE Trans. Automatic Control. 1974; 19:716–723.
46. Schwarz G. Estimating the dimension of a model. Ann. Stat. 1978; 6:461–464.

47. Stephan KE, Penny WD, Daunizeau J, Moran RJ, Friston KJ. Bayesian model selection for group studies. Neuroimage. Jul 15; 2009 46(no. 4):1004–17. [PubMed: 19306932]
48. Acs F, Greenlee MW. Connectivity modulation of early visual processing areas during covert and overt tracking tasks. Neuroimage. Jun; 2008 41(no. 2):380–8. [PubMed: 18387824]
49. Grol MJ, Majdandzic J, Stephan KE, Verhagen L, Dijkerman HC, Bekkering H, Verstraten FA, Toni I. Parieto-frontal connectivity during visually guided grasping. J Neurosci. Oct 31; 2007 27(no. 44):11877–87. [PubMed: 17978028]
50. Kumar S, Stephan KE, Warren JD, Friston KJ, Griffiths TD. Hierarchical processing of auditory objects in humans. PLoS Comput Biol. Jun.2007 3(no. 6):e100. [PubMed: 17542641]
51. Leff AP, Schofield TM, Stephan KE, Crinion JT, Friston KJ, Price CJ. The cortical dynamics of intelligible speech. J Neurosci. Dec 3; 2008 28(no. 49):13209–15. [PubMed: 19052212]
52. Noppeney U, Josephs O, Hocking J, Price CJ, Friston KJ. The effect of prior visual information on recognition of speech and sounds. Cereb Cortex. Mar; 2008 18(no. 3):598–609. [PubMed: 17617658]
53. Summerfield C, Koechlin E. A neural representation of prior information during perceptual inference. Neuron. Jul 31; 2008 59(no. 2):336–47. [PubMed: 18667160]
54. Kasess CH, Windischberger C, Cunnington R, Lanzenberger R, Pezawas L, Moser E. The suppressive influence of SMA on M1 in motor imagery revealed by fMRI and dynamic causal modeling. Neuroimage. Apr 1; 2008 40(no. 2):828–37. [PubMed: 18234512]
55. Stephan KE, Tittgemeyer M, Knosche TR, Moran RJ, Friston KJ. Tractography-based priors for dynamic causal models. Neuroimage. Oct 1; 2009 47(no. 4):1628–38. [PubMed: 19523523]
56. Kiebel SJ, Kloppel S, Weiskopf N, Friston KJ. Dynamic causal modeling: a generative model of slice timing in fMRI. Neuroimage. Feb 15; 2007 34(no. 4):1487–96. [PubMed: 17161624]
57. Marreiros AC, Kiebel SJ, Friston KJ. Dynamic causal modelling for fMRI: a two-state model. Neuroimage. Jan 1; 2008 39(no. 1):269–78. [PubMed: 17936017]
58. Woolrich, M.; Jbabdi, S.; Behrens, TE. FMRI Dynamic Causal Modelling with Inferred Regions of Interest; Abstract presented at the annual meeting of the Organsation for Human Brain Mapping; San Francisco: 2009.
59. Zucker RS, Regehr WG. Short-term synaptic plasticity. Annu Rev Physiol. 2002; 64:355–405. [PubMed: 11826273]
60. Salinas E, Sejnowski TJ. Gain modulation in the central nervous system: where behavior, neurophysiology, and computation meet. Neuroscientist. Oct; 2001 7(no. 5):430–40. [PubMed: 11597102]
61. Sarter M, Hasselmo ME, Bruno JP, Givens B. Unraveling the attentional functions of cortical cholinergic inputs: interactions between signal-driven and cognitive modulation of signal detection. Brain Res Brain Res Rev. Feb; 2005 48(no. 1):98–111. [PubMed: 15708630]
62. Blake R, Logothetis NK. Visual competition. Nat Rev Neurosci. Jan; 2002 3(no. 1):13–21. [PubMed: 11823801]
63. Lumer ED, Friston KJ, Rees G. Neural correlates of perceptual rivalry in the human brain. Science. Jun 19; 1998 280(no. 5371):1930–4. [PubMed: 9632390]
64. Friston K. A theory of cortical responses. Philos Trans R Soc Lond B Biol Sci. Apr 29; 2005 360(no. 1456):815–36. [PubMed: 15937014]
65. Montague PR, Dayan P, Sejnowski TJ. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. J Neurosci. Mar 1; 1996 16(no. 5):1936–47. [PubMed: 8774460]
66. Schultz W, Dickinson A. Neuronal coding of prediction errors. Annu Rev Neurosci. 2000; 23:473–500. [PubMed: 10845072]
67. den Ouden HE, Friston KJ, Daw ND, McIntosh AR, Stephan KE. A dual role for prediction error in associative learning. Cereb Cortex. May; 2009 19(no. 5):1175–85. [PubMed: 18820290]
68. Eickhoff SB, Dafotakis M, Grefkes C, Shah NJ, Zilles K, Piza-Katzer H. Central adaptation following heterotopic hand replantation probed by fMRI and effective connectivity analysis. Exp Neurol. Jul; 2008 212(no. 1):132–44. [PubMed: 18501895]

69. Grefkes C, Nowak DA, Eickhoff SB, Dafotakis M, Kust J, Karbe H, Fink GR. Cortical connectivity after subcortical stroke assessed with functional magnetic resonance imaging. Ann Neurol. Feb; 2008 63(no. 2):236–46. [PubMed: 17896791]

70. Stephan KE, Baldeweg T, Friston KJ. Synaptic plasticity and dysconnection in schizophrenia. Biol Psychiatry. May 15; 2006 59(no. 10):929–39. [PubMed: 16427028]

Europe PMC Funders Author Manuscripts

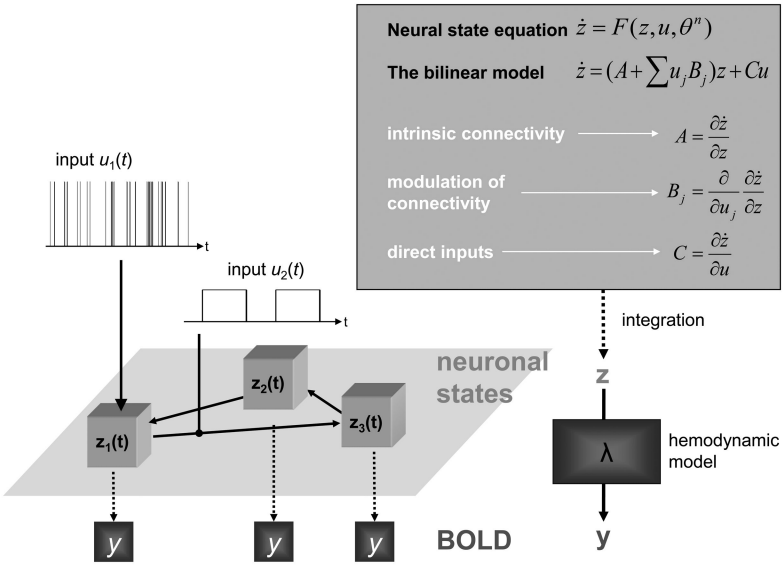

Figure 1.

Schematic summary of the conceptual basis of DCM. The dynamics in a system of interacting neuronal populations (left lower panel), which are not directly observable by fMRI, is modeled using a bilinear state equation (right upper panel). Integrating the state equation gives predicted neural dynamics (z) that enter a model of the hemodynamic response (λ) to give predicted BOLD responses (y) (right lower panel). The parameters at both neural and hemodynamic levels are adjusted such that the differences between predicted and measured BOLD series are minimized. Critically, the neural dynamics are determined by experimental manipulations. These enter the model in the form of external inputs (left upper panel). Driving inputs (u_1; e.g. sensory stimuli) elicit local responses directly that are propagated through the system according to the intrinsic connections. The strengths of these connections can be changed by modulatory inputs (u_2; *e.g.* changes in cognitive set, attention, or learning). In this figure, the structure of the system and the scaling of the inputs are arbitrary. This figure was reproduced, with permission, from Figure 1 in [36].

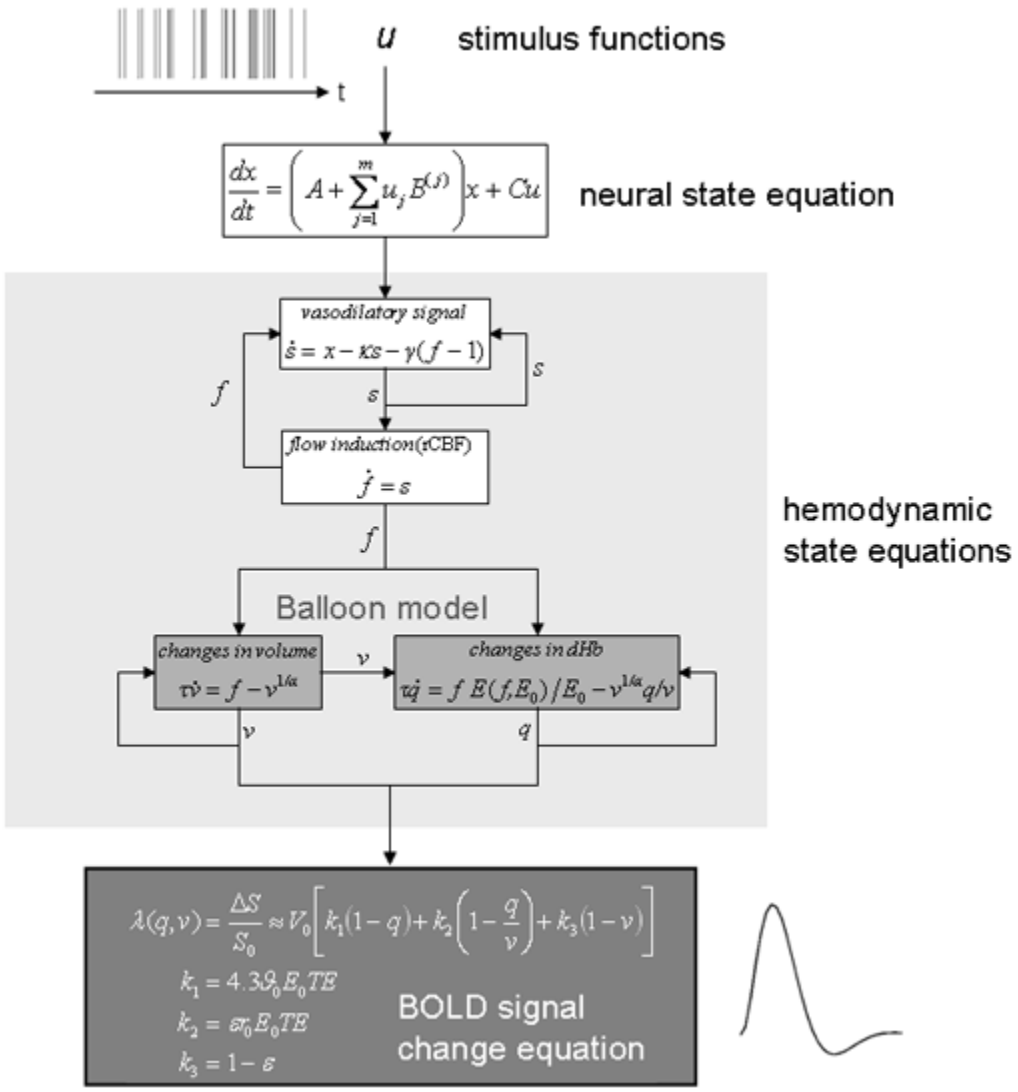

Figure 2.

Schematic summary of the neural state equation and the hemodynamic forward model in DCM; reproduced, with permission, from Figure 1 in [33]. Experimentally controlled input functions u evoke neural responses x, modeled by a bilinear differential state equation, which trigger a hemodynamic cascade, modeled by 4 state equations with 5 parameters. These hemodynamic parameters comprise the rate constant of the vasodilatory signal decay (κ), the rate constant for auto-regulatory feedback by blood flow (γ), transit time (τ), Grubb's vessel stiffness exponent (α), and capillary resting net oxygen extraction (ρ). The so-called Balloon model consists of the two equations describing the dynamics of blood volume (v) and deoxyhemoglobin content (q) (light grey boxes). Integrating the state equations for a given set of inputs and parameters produces predicted time-series for v and q which enter a BOLD signal equation λ (dark grey box) to give a predicted BOLD response.

Europe PMC Funders Author Manuscripts

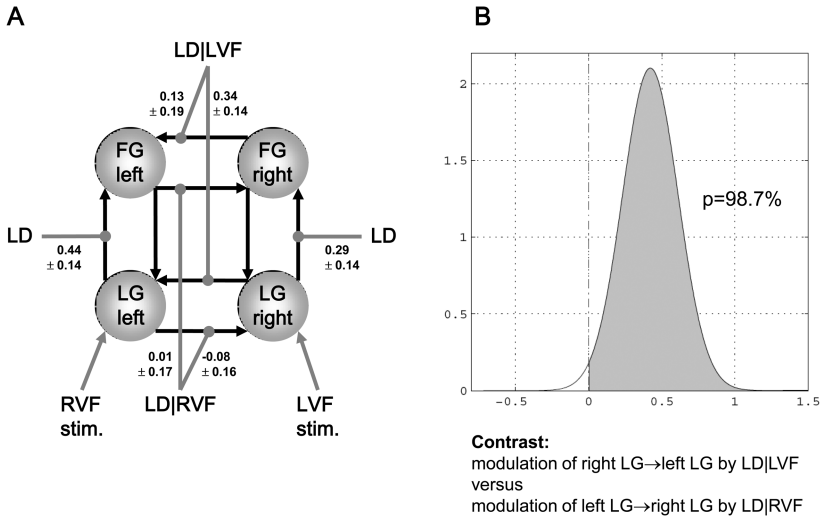

Figure 3.

This figure was adapted, with permission, from Figures 5 and 6 in [36]. It shows an example of a single subject DCM that was used to study asymmetries in interhemispheric connections during a letter decision task. LG = lingual gyrus, FG = fusiform gyrus, LD = letter decisions, LD|VF = letter decisions conditional on the visual field of stimulus presentation.

A. The values denote the maximum a posteriori (MAP) estimates of the parameters (± square root of the posterior variances; units: 1/s=Hz). For clarity, only the parameters of interest, *i.e.* the modulatory parameters of inter- and intra-hemispheric connections, are shown.

B. Asymmetry of callosal connections with regard to contextual modulation. The plots show the probability (98.7%) that the modulation of the right LG → left LG connection is stronger than the modulation of the left LG → right LG connection.

Europe PMC Funders Author Manuscripts

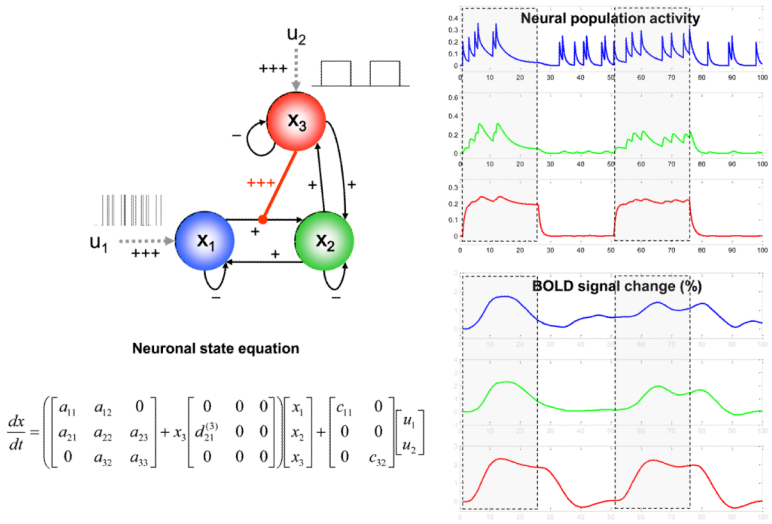

Figure 4.

An example of the neuronal and haemodynamic dynamics that can be accounted for by nonlinear DCMs. The figure is reproduced, with permission, from Figure 2 in [30]. The right panel shows synthetic neuronal and BOLD time-series that were generated using the nonlinear DCM shown on the left. In this model, neuronal population activity x_1 (blue) is driven by irregularly spaced random events (delta-functions). Activity in x_2 (green) is driven through a connection from x_1; critically, the strength of this connection depends on activity in a third population, x_3 (red), which receives a connection from x_2 but also receives a direct input from a box-car input. The effect of nonlinear modulation can be seen easily: responses of x_2 to x_1 become negligible when x_3 activity is low. Conversely, x_2 responds vigorously to x_1 inputs when the $x_1 \rightarrow x_2$ connection is gated by x_3 activity. Strengths of connections are indicated by symbols (−: negative; +: weakly positive; +++: strongly positive).

Europe PMC Funders Author Manuscripts

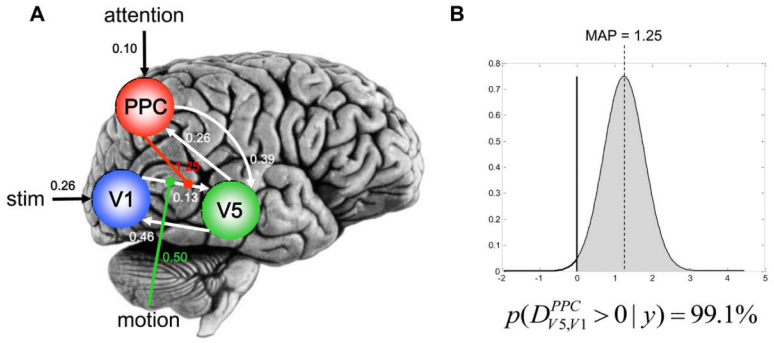

Figure 5.

Application of nonlinear DCM to single subject fMRI data from an attention to motion paradigm [17]. The figure is reproduced, with permission, from Figure 7 in [30].

A. Maximum a posteriori estimates of all parameters. PPC = posterior parietal cortex.

B. Posterior density of the estimate for the nonlinear modulation parameter for the V1→V5 connection. Given the mean and variance of this posterior density, we have 99.1% confidence that the true parameter value is larger than zero or, in other words, that there is an increase in gain of V5 responses to V1 inputs that is mediated by PPC activity.

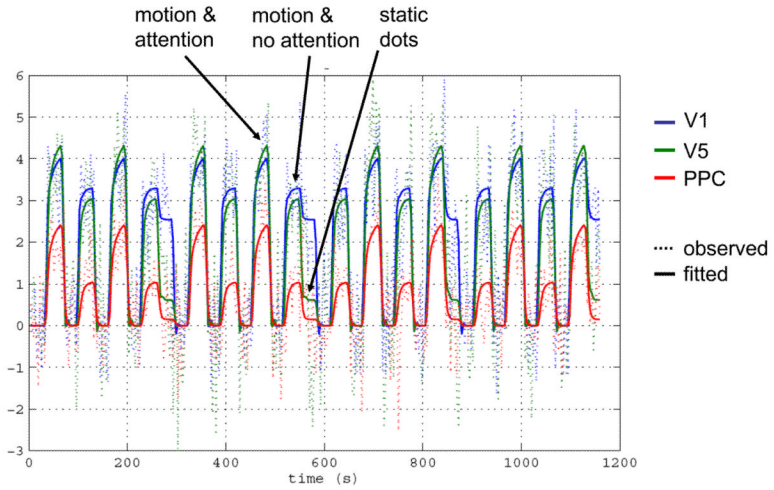

Figure 6.

Fit of the nonlinear model to the attention to motion data in Figure 5. Dotted lines represent the observed data, solid lines the responses predicted by the nonlinear DCM. The increase in the gain of V5 responses to V1 inputs during attention is clearly visible. The figure is reproduced, with permission, from Figure 8 in [30].

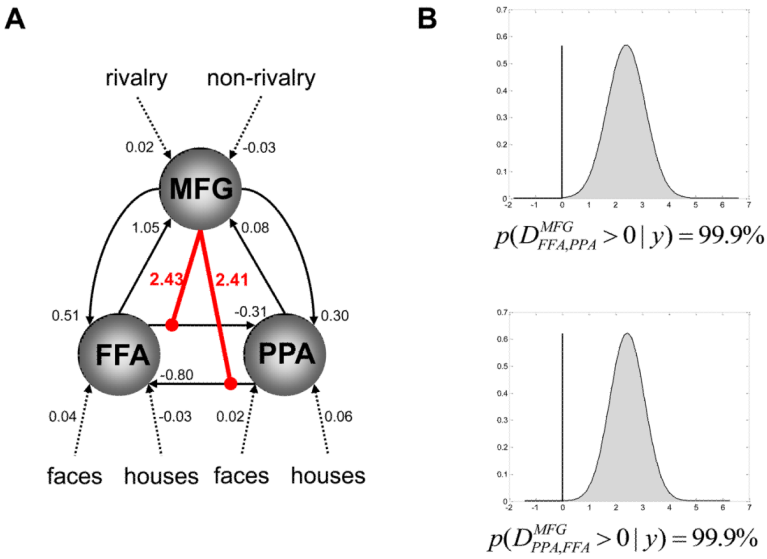

Figure 7.

Application of nonlinear DCM to single subject fMRI data from a binocular rivalry paradigm. The figure is reproduced, with permission, from Figure 9 in [30].

A. The structure of the nonlinear DCM fitted to the binocular rivalry data, along with the maximum a posteriori estimates of all parameters. The intrinsic connections between FFA and PPA are negative in both directions; *i.e.* FFA and PPA mutually inhibited each other. This may be seen as an expression, at the neurophysiological level, of the perceptual competition between the face and house stimuli. This competitive interaction between FFA and PPA is modulated nonlinearly by activity in the middle frontal gyrus (MFG), which showed higher activity during rivalry vs. non-rivalry conditions.

B. Our confidence about the presence of this nonlinear modulation is very high (99.9%), for both connections.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

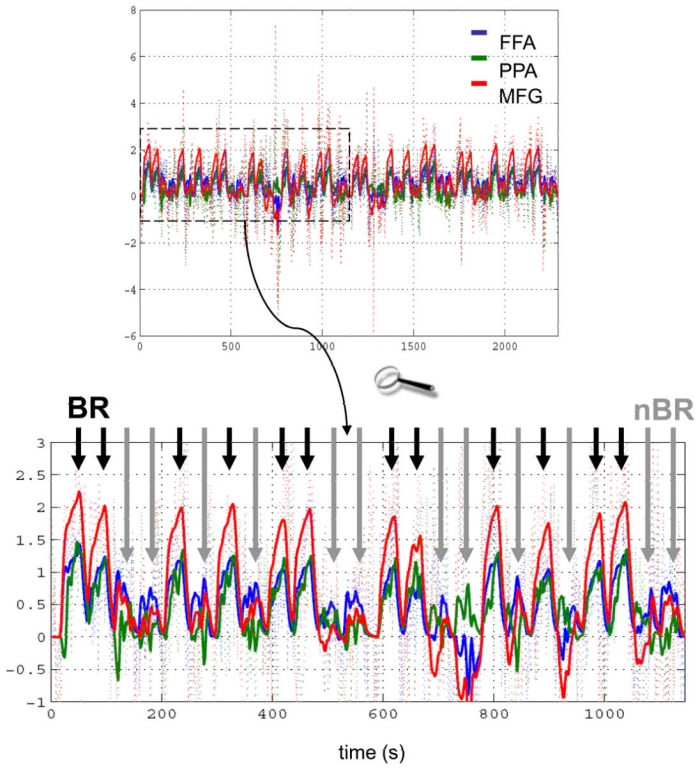

Figure 8.
Fit of the nonlinear model in Figure 7 to the binocular rivalry data. Dotted lines represent the observed data, solid lines the responses predicted by the nonlinear DCM. The upper panel shows the entire time series. The lower panel zooms in on the first half of the data (dotted box). One can see that the functional coupling between FFA (blue) and PPA (green) depends on the activity level in MFG (red): when MFG activity is high during binocular rivalry blocks (BR; short black arrows), FFA and PPA are strongly coupled and their responses are difficult to disambiguate. In contrast, when MFG activity is low, during non-rivalry blocks (nBR; long grey arrows), FFA and PPA are less coupled, and their activities evolve more independently. The figure is reproduced, with permission, from Figure 10 in [30].