ELSEVIER

NeuroImage

www.elsevier.com/locate/ynimg
NeuroImage 30 (2006) 1243 – 1254

Large-scale neural models and dynamic causal modelling

Lucy Lee,[a] Karl Friston,[a,*] and Barry Horwitz [b]

[a]*Wellcome Department of Imaging Neuroscience, 12 Queen Square, London WC1N 3BG, UK*
[b]*Brain Imaging and Modelling Section, National Institute on Deafness and other Communication Disorders, National Institutes of Health, Bethesda, MD 20892-2320, USA*

Received 21 December 2004; revised 4 November 2005; accepted 10 November 2005
Available online 4 January 2006

Dynamic causal modelling (DCM) is a method for estimating and making inferences about the coupling among small numbers of brain areas, and the influence of experimental manipulations on that coupling [Friston, K.J., Harrison, L., Penny, W., 2003. Dynamic causal modelling. Neuroimage 19, 1273–1302]. Large-scale neural modelling aims to construct neurobiologically grounded computational models with emergent behaviours that inform our understanding of neuronal systems. One such model has been used to simulate region-specific BOLD time-series [Horwitz, B., Friston, K.J., Taylor, J.G., 2000. Neural modeling and functional brain imaging: an overview. Neural Netw. 13, 829–846]. DCM was used to make inferences about effective connectivity using data generated by a model implementing a visual delayed match-to-sample task [Tagamets, M.A., Horwitz, B., 1998. Integrating electrophysiological and anatomical experimental data to create a large-scale model that simulates a delayed match-to-sample human brain imaging study. Cereb. Cortex 8, 310–320]. The aim was to explore the validity of inferences made using DCM about the connectivity structure and task-dependent modulatory effects, in a system with a known connectivity structure. We also examined the effects of misspecifying regions of interest. Models with hierarchical connectivity and reciprocal connections were examined using DCM and Bayesian Model Comparison [Penny, W.D., Stephan, K.E., Mechelli, A., Friston, K.J., 2004. Comparing dynamic causal models. Neuroimage 22, 1157–1172]. This approach revealed strong evidence for those models with correctly specified anatomical connectivity. Furthermore, Bayesian model comparison favoured those models when bilinear effects corresponded to their implementation in the neural model. These findings generalised to an extended model with two additional areas and reentrant circuits. The conditional uncertainty of coupling parameter estimates increased in proportion to the number of incorrectly specified regions. These results highlight the role of neural models in establishing the validity of estimation and inference schemes. Specifically, Bayesian model comparison confirms the validity of DCM in relation to a well-characterised and comprehensive neuronal model.

Keywords: Dynamic causal modelling; fMRI; Large-scale neural model; Validation

Introduction

The organisation of the brain follows two principles: functional specialisation and functional integration. Functional specialisation implies that anatomically distinct cortical (and subcortical) areas are specialised to perform certain aspects of perceptual, cognitive or motor processing. Given a high degree of functional specialisation, no cortical area can perform a meaningful function in isolation. In order to support even a simple perceptual or motor function, it is necessary for these anatomically distinct functionally specialised areas to interact, via extrinsic connections among cortical areas. Functional integration describes the pattern of connections, established between cortical areas, which is unique to a particular function. Traditionally, the high spatial resolution available with functional neuroimaging data has lent itself to analyses of functional specialisation (Zeki et al., 1991). Analyses of neural activity based solely on this principle provide a limited account of the neuronal substrate of the process under investigation. Therefore, alternative approaches have been developed to investigate task-dependent changes in the integration of functionally specific areas. For a review of three methods of analysing functional integration, see Ramnani et al. (2004); for a discussion of some conceptual issues associated with connectivity analysis, see Lee et al. (2003) and Horwitz (2003).

Dynamic causal modelling (DCM) is a method for estimating and making inferences about the coupling among a small number of specified brain areas, and the influence of experimental manipulations on that coupling (Friston et al., 2003) that has been developed specifically to work with functional neuroimaging data. It is used to test specific null hypotheses about the coupling between cortical areas. It does not attempt to provide a realistic model of how the brain actually works. DCM treats the brain as a dynamic, deterministic input-state-output system where the inputs are conventional stimulus functions describing experimental manipulations. The state variables describe regional neuronal activity and biophysical states specify the haemodynamic response in each region, given its neuronal activity. The outputs are the measured regional blood oxygenation-dependent (BOLD) signal. Within the framework of DCM, the neurodynamics of any given

* Corresponding author. Fax: +44 2078131420.
E-mail address: k.friston@fil.ion.ucl.ac.uk (K. Friston).
Available online on ScienceDirect (www.sciencedirect.com).

1053-8119/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2005.11.007

region are treated as deterministic: changes in neuronal activity in a region can only be caused by direct inputs to that area (if specified), intrinsic activity in the area (via self-connections) and activity in other connected regions. Any of the connections can be modulated by experimental variables. The modeller chooses the regions of interest, specifies which connections are present and indicates where the experimental manipulations are expected to act, either as direct inputs to a region or by modulating the strength of connections between regions.

There are a number of issues associated with DCMs (Penny et al., 2004). Of particular importance is the problem of specifying the latent (interregional) connectivity structure because detailed knowledge of the anatomical connectivity of the human brain is unknown (Passingham et al., 2002). Another important issue is the specification of those connections that are modulated by experimentally controlled variables, which reflects the hypotheses under examination. When working with experimental data, these are empirical questions. Using simulated data based on simple neuronal models, Penny et al. demonstrated that Bayesian model comparison can be used to distinguish between competing models.

This paper aims to provide a validation of DCM by using synthetic fMRI time-series data simulated by a neurobiologically informed computational model (see Horwitz et al., 2000 for an overview of employing computational neural modelling in conjunction with functional neuroimaging). One such model has been used to simulate region-specific BOLD time-series (Horwitz and Tagamets, 1999). Large-scale neural models can be used to partially validate techniques such as DCM because the underlying anatomical structure and neuronal physiology are fully known. As such, their use can permit an examination of the neural substrates underlying the interpretation of DCMs. Our aim was to establish the face validity of inferences made using DCM in relation to a known (model) architecture. This is clearly not possible using data from the brain, whose connectivity structure and mechanisms, mediating phenomena like attention, are unknown.

Neuronal model

Neuronal model

A neurobiologically grounded computational model was used to simulate region-specific BOLD fMRI time-series data. This previously published large-scale model (Tagamets and Horwitz, 1998) simulates the ventral visual processing stream during a delayed match to sample (DMS) task for object shape (an analogous model for processing auditory objects was developed by Husain et al., 2004). The model contains four major brain regions: primary visual cortex (V1/V2), occipitotemporal cortex (V4), inferior temporal cortex (IT) and prefrontal cortex (PFC) (Fig. 1). Each region is composed of one or more arrays of 9 × 9 neuronal assemblies of basic units, representing a simplified cortical column. The basic unit underlying all regions in the model consists of a coupled inhibitory and excitatory element with the excitatory element providing a greater contribution to the overall synaptic activity by virtue of a larger synaptic weight. The excitatory element provides interarea connections. Spontaneous low-level random activity occurs in all units at all times. This model was able to perform a DMS task for object shape; the simulated electrical activities of the excitatory units in each region were similar to experimental electrical recordings from nonhuman primates and the simulated summed synaptic activities in each region matched human functional brain imaging data obtained during visual DMS tasks (Tagamets and Horwitz, 1998). For details of the parameters used in the model and a thorough discussion of the assumptions employed, see Tagamets and Horwitz (1998) and Horwitz and Tagamets (1999).

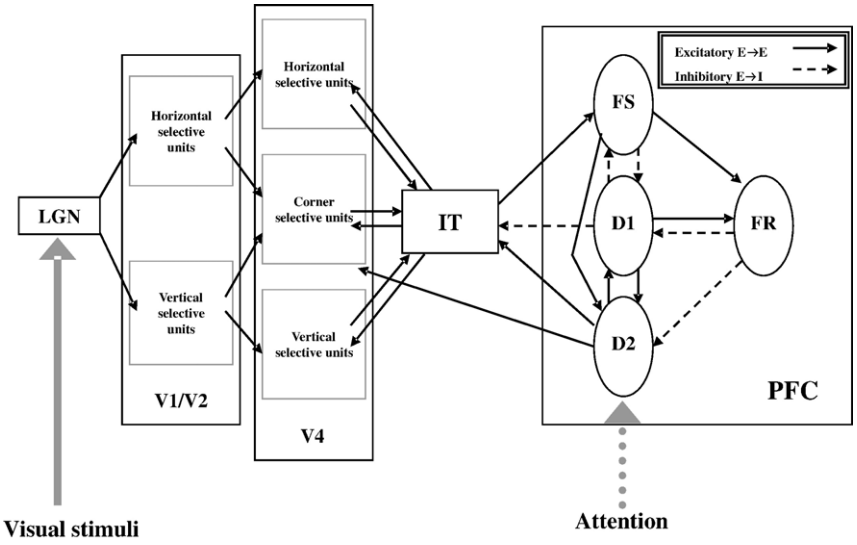

Fig. 1. Network diagram of the visual object processing model. The regions form a network with forward and backward connections, some of which are excitatory (excitatory to excitatory neuronal populations: solid lines) and others are inhibitory (excitatory to inhibitory populations: dashed lines). In the prefrontal (PFC) region, FS contains stimulus sensitive units, D1 and D2 contain units active during the delay period of the task and FR contains units whose activity increases if there is a match between the first and second stimuli of a trial. The solid grey arrow shows the site of action of the visual stimuli, the dotted grey arrow shows the site of action of attention.

Delayed match to sample task

The DMS task involves presentation of a shape, a variable duration delay period, followed by presentation of a second stimulus (same or different). The task requires the model to decide if the second stimulus matches the first. The stimulus shapes used are shown in Fig. 2.

The arrival of a stimulus activates orientation selective (e.g., horizontal and vertical) units in V1/V2 which respond best to small sections of horizontal or vertical lines. The next stage in the visual processing stream (V4) contains units that respond to longer horizontal and vertical lines, and to corners. The third stage (IT) contains units that respond to integrated information from V4, i.e., they form a 'percept' of the entire stimulus. This representation is then transmitted to the first of four distinct types of PFC units. The cue arrives at cue-selective units (FS), which are active only during stimulus presentation. Information is then passed to two types of delay units: type D2 is active during the initial presentation and the delay period; type D1 is only active during the delay. The D2 units provide feedback to earlier areas. The fourth type of PFC module consists of response units (FR), whose activity increase when the second cue matches the first. The PFC units become more active during stimulus presentation and maintain a representation of the stimulus when it is no longer present. An additional experimental factor (attention) is implemented via a low-level, diffuse increase in activity in the D2 units. This modulates how the delay units respond to a given stimulus: the higher the attention, the better the representation that is maintained during the delay. The effect of attention of other areas is mediated via connections from D2. Although the various neurons that comprise the PFC are likely to be intermixed in the same anatomical area in the brain (Funahashi et al., 1993), these modules are treated as if there were separate regions in this paper because the anatomical connections between these modules are different from the other components of the model. The connectivity of the network and site of action of inputs are shown in Fig. 1. A detailed discussion of the development of the original model is provided in Tagamets and Horwitz (1998).

A recent upgrade of the model (Horwitz et al., 2005) is used in this work. This incorporates a group of nonspecific neurons in each region, which are connected to each other in the same way as are the task-specific neurons. The input to the nonspecific neurons consists of noise patterns that are presented asynchronously and randomly relative to the presentation times of the stimuli to the task-specific elements. The net effect of this arrangement is that the task-specific neurons in each module receive random neural activity from the nonspecific neurons that varies for each trial.

Study design

The parameters used to simulate the fMRI data conform to a 2 × 2 factorial design. The two factors are 'Task' and 'Attention'. The 'Task' factor had two levels: visual inputs were either coherent shapes (active condition) or scrambled shapes (control condition). The 'Attention' factor had three levels (low, medium and high). Additional variability between trials was induced by setting the duration of the delay period to 1.5 s, 3 s or 6 s. This variability in the delay between presentation of the first and second visual stimuli makes it possible to disambiguate between activity in the prefrontal regions in response to visual stimuli, and the sustained prefrontal activity in the absence of visual input that enables the model to perform the 'match' element of the task.

Simulation of fMRI time-series

The haemodynamic fMRI data were obtained by integrating the absolute values of the synaptic activities within the different regions over time. The total synaptic activity in a region is the total sum of absolute values of the inputs to the inhibitory and excitatory units (inter- and intraregional inputs) at each point in time. The window of integration was 50 ms, following which the resulting values were convolved with a Poisson function representing a haemodynamic delay of 6 s, and down-sampled to simulate a repetition time (TR) of 2 s. Details of this approach can be found in Husain et al. (2004).

Dynamic causal modelling

DCM was used to make inferences about effective connectivity using data generated by this large-scale neural model while implementing a visual delayed match-to-sample task (Tagamets and Horwitz, 1998). DCM relies on the specification of a simple

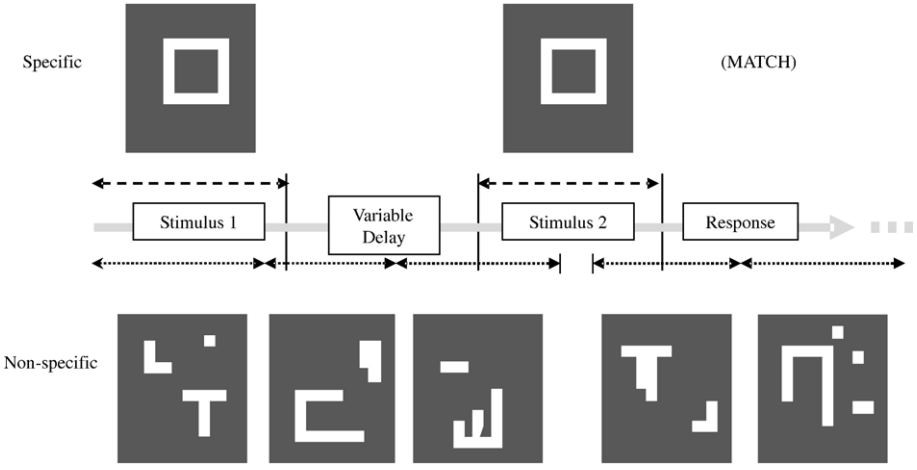

Fig. 2. Schematic of the delayed match to sample (DMS) task. The shape stimuli presented during the task are shown at the top of the figure. The degraded stimuli, presented during the control condition, are shown at the foot of the figure. Each shape is presented for 1 s, the response period (which includes the intertrial interval) is 6 s, and the delay period is 1.5 s, 3 s or 6 s.

but reasonably realistic neuronal model of interacting regions. The neuronal states can be described by Eq. (1)

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

F is a nonlinear function describing the influences among the brain regions included in the model on their mean synaptic activity (z) and the experimental inputs (u). θ are the parameters of the model. A bilinear approximation of this equation provides a reparameterisation in terms of the effective connectivity encoded by the matrices A, B and C:

$$\dot{z} \approx \mathrm{A}z + \Sigma u_j \mathrm{B}^j z + \mathrm{C}u = \left(\mathrm{A} + \Sigma u_j \mathrm{B}^j\right) z + \mathrm{C}u \tag{2}$$

The matrix A represents first-order or latent connectivity i.e., coupling among regions in the absence of experimental inputs or perturbations. This is the first term of a Taylor Expansion of the system equation (Eq. (2)). Put simply, this quantifies how the rate of change in activity in a target region depends on the activity of a source region. Constraints on these connections can embed knowledge of anatomical connections (if available) to create a plausible network. The matrix C specifies the direct influence of extrinsic inputs (u) on regions. The matrix B specifies the changes in first order or latent connections (self-connections or connections between regions) induced by the jth input (u). The parameters of the neuronal state equation are therefore the coupling matrices A, B^j and C. Priors on the coupling parameters ensure that the system remains dissipative, i.e., returns to a steady state in the absence of input.

The haemodynamic model comprises state variables describing the translation of regional neuronal activity into regional haemodynamic responses. These include a vasodilatory signal (s), normalised flow (f), normalised venous volume (v) and normalised deoxyhaemoglobin content (q) (Friston et al., 2000; Mechelli et al., 2001). Empirically determined priors for the biophysical parameters are based on previous data (Friston et al., 2000).

Combining the neuronal and biophysical states gives a full forward model, whose parameters are estimated using a fully Bayesian approach to derive maximum a posteriori (MAP) estimates and conditional covariances. The units of the coupling parameters are per unit time; a strong connection corresponds to an influence that is expressed quickly with a small time constant. The conditional density of a variable is the posterior probability density of that variable having observed another, e.g., the posterior probability density of the model parameters given the data. Mathematically speaking, this density is conditioned on the data. The conditional densities can be used (i) to make inferences about coupling given a particular model, M_i using $p(\theta|y, M_i)$ (the probability of the parameters, given the data and the model) or (ii) used to compute the evidence for that model $p(y|M_i)$. The evidence for several models can then be used to compute Bayes Factors, for Bayesian model comparison. Following Penny et al. (2004), we approximated $p(y|M_i)$ with the Bayesian and Akaike's Information Criterion (AIC and BIC). This renders model comparison less dependent on priors.

DCM uses a bilinear reparameterisation of nonlinear neuronal dynamics to enable inferences about effective connectivity. The bilinear approximation reduces any nonlinear model to the same form. It could be applied to the differential equations used to generate data in the large-scale neural model, or to the (unknown) analytic equations governing real neuronal dynamics and the generation of fMRI or ERP data in vivo. This approximation is particularly useful because it parameterises effective connectivity into components that are context-invariant (latent part) and those that are input or context-sensitive (the bilinear part). The key issue is whether the bilinear approximation is sufficient to capture features of interest. This is an empirical issue. If the DCM models the data well, then conditional inferences will be significant,

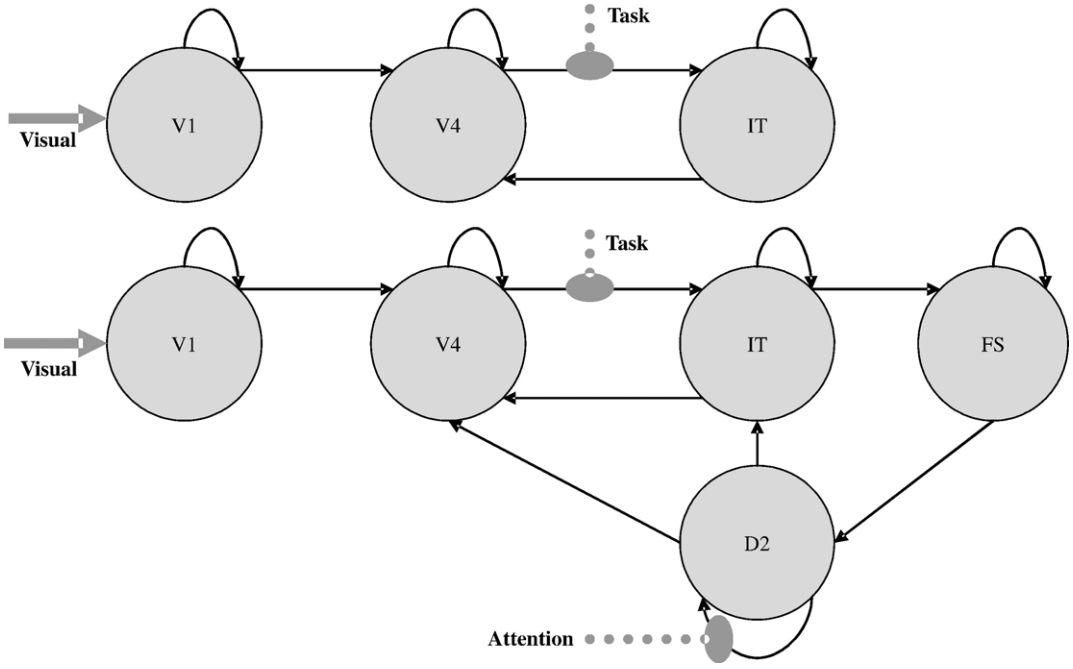

Fig. 3. Schematic of the dynamic causal model (DCM) architectures that correctly represent part of the underlying structure of the computational model used to simulate the fMRI time-series data. Not all areas of the computational model are included in the DCM. Direct inputs are shown in solid grey; modulatory inputs are shown in dotted grey.

otherwise they will not. Similarly, in comparing DCMs, a significant difference in their relative log-evidence is evidence that the difference between the DCMs is supported by the data.

In short, DCM is not used to emulate neuronal dynamics to explore emergent behaviours (this is the goal of the neuro-biologically inspired computational model); rather, it is designed to test hypotheses specified operationally by the DCMs. Using one DCM, it is only possible to address quantitative questions about a specific model. For example, in this paper, we required a 90% confidence that a contrast of parameters was greater than zero. However, Bayesian model comparison also allows us to answer qualitative questions that are framed in terms of competing models. Heuristically, instead of asking 'given this model, what do the data tell us about its parameters?', we can now ask 'given two models (descriptions of the system), do the data support one model significantly more than the alternative?' The criterion for this is the difference in log-evidence, as described by Penny et al. (2004) in terms of Bayes factors. In this paper, we required 'strong' evidence for one model relative another (i.e., a relative probability of 20 or more).

Two models were used to explore the validity of inferences made using DCM: a simple three-area model consisting of V1/2, V4 and IT and a five-area model extended to explain responses in areas FS and D2 (see Fig. 1). The reduced model does not incorporate all the regions used to derive the data. This is likely to reflect the situation using in vivo fMRI data where it is unlikely that all the nodes of a network will be known.

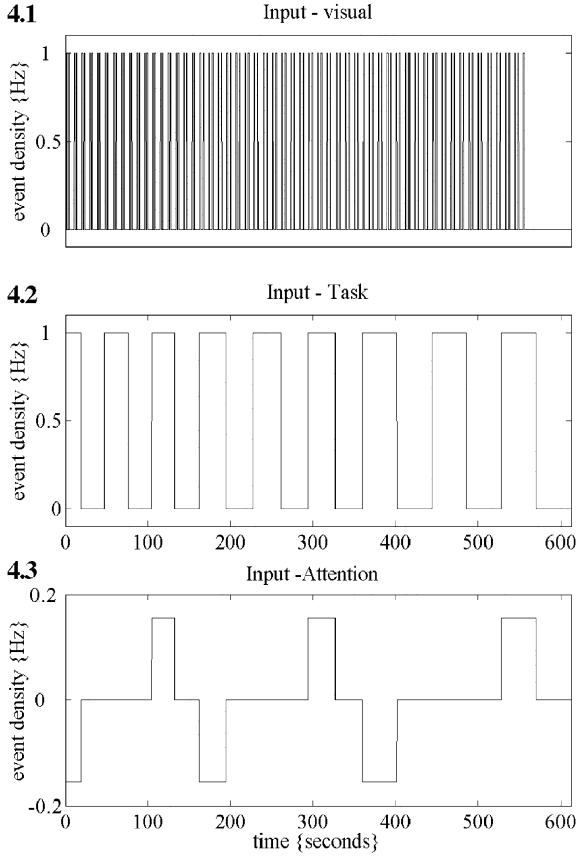

Fig. 4. Inputs to the DCMs. The plots show the driving input (visual) (4.1) and the modulatory inputs (task (4.2) and attention (4.3)). The latter enters only the extended five-area DCM. The stimuli used as driving inputs to the large-scale neural model are in Fig. 2.

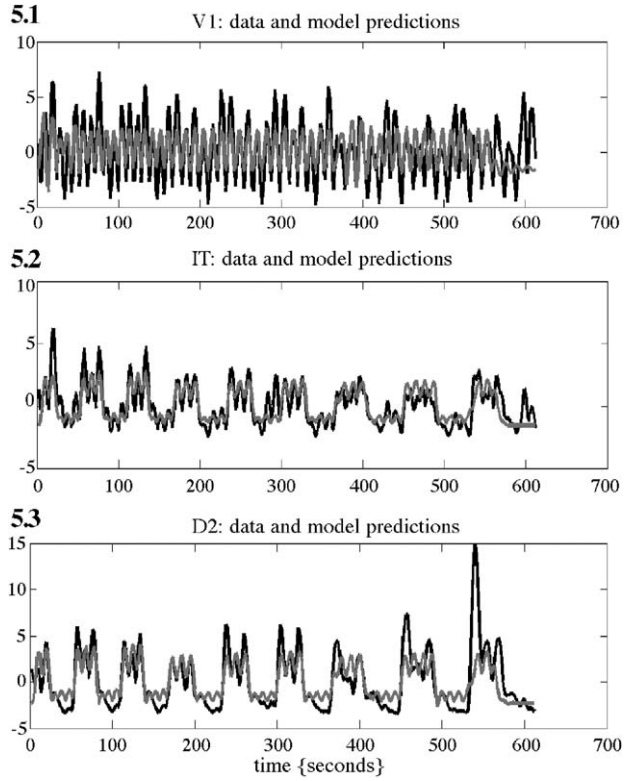

Fig. 5. Outputs of the DCM (grey) and the simulated fMRI time-series data derived from the large-scale computational model (black) for three regions: V1 (5.1), IT (5.2) and D2 (5.3). It can be seen that the output of the DCM is not a perfect fit for the data from the large-scale model, which is noisier due to the activity of the nonselective neuronal populations. However, the DCM output does capture the major changes in the simulated fMRI data.

The 'correct' DCMs can be seen in Fig. 3. For both models, the direct sensory inputs comprise a series of events with duration of 1 s occurring every time a visual stimulus was 'presented'. In both models, the factor 'task' (specific vs. nonspecific shapes) modulated the strength of connections from V4 to IT, because in the computational model, there is very little activity in IT when noise patterns are presented and significant neural activity when shapes are used as the input stimuli. In the five-area model, the second factor 'attention', which had three levels (high, medium, low), entered as an additional modulatory variable on the self-connections of the prefrontal D2 area. The format of the inputs is shown in Fig. 4. The output of the DCMs (predicted fMRI time-series) and the simulated fMRI time-series data derived from the large-scale computational model are shown in Fig. 5.

Simulations

Determining latent connectivity

For the simple model, three possible connectivity structures were compared: the correct connectivity, a complete hierarchical connectivity specification and full latent connectivity. The models used can be seen in Fig. 6. In each case, visual inputs entered directly to V1 and the factor 'task' modulated the connections from

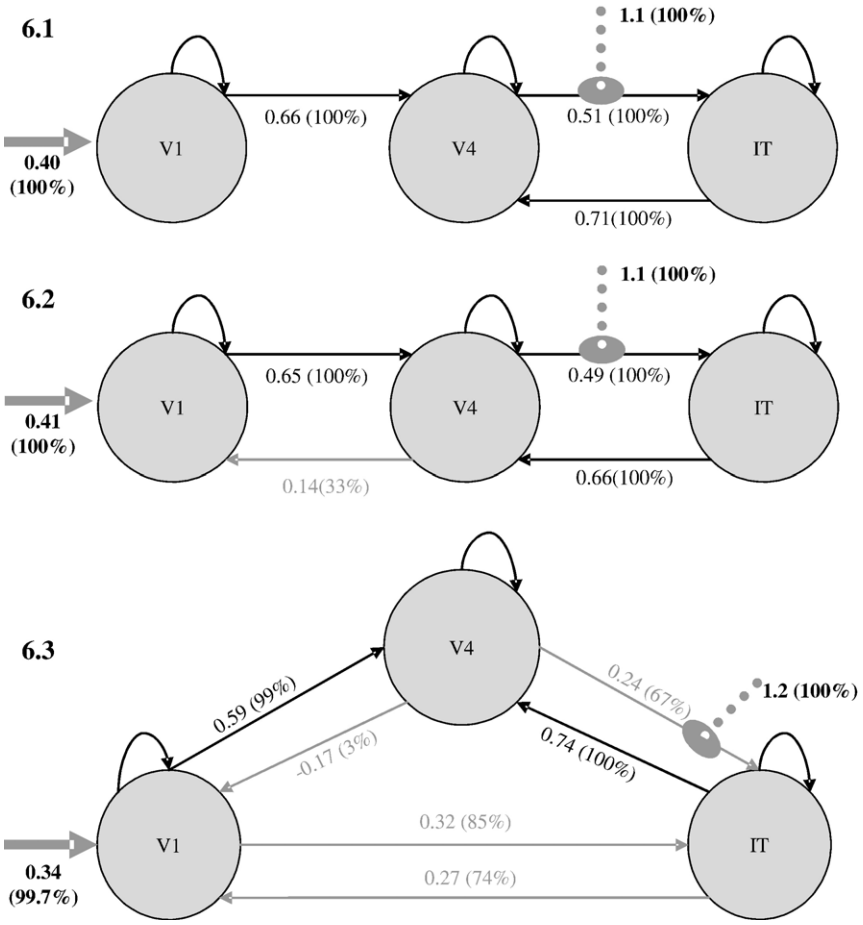

Fig. 6. (6.1) Coupling parameters for the simple model, with interarea connectivity and modulatory input specified correctly with respect to the underlying large-scale neural model. The conditional density (expectations and variances) of the coupling parameters were computed using Expectation Maximisation (EM) as described in Friston et al. (2003). The conditional expectations correspond to the posterior estimates of coupling strength, while the conditional variances were used to compute the confidence that the parameters were greater than threshold (this is simply the appropriate area under the Gaussian conditional density). The posterior parameter estimates for the coupling parameters are shown in black and grey; the values in brackets are the confidence that these values exceed a threshold of ln2/4 Hz. Coupling parameters exceeding threshold with a confidence of greater than 90% are shown in black. The posterior parameter estimates for the coupling parameters for direct visual inputs are shown next to the solid grey arrows. The posterior parameter estimates for the coupling parameters for modulatory effect of task (shapes vs. degraded stimuli) are shown next to the dotted grey markers. The values in brackets are the percentage confidence that these values are greater than zero. (6.2) Coupling parameters for the simple model, specified with full interarea connectivity. Inputs are specified and displayed as in panel 6.1. (6.3) Coupling parameters for the simple model, specified as a hierarchy. Inputs are specified and displayed as in panel 6.1.

V4 to IT. The latent connectivity was assessed in two ways (see Penny et al. (2004) for a complete discussion of how these comparisons are performed): (1) Conditional on each model we asked, do the coupling parameters exceed a threshold of ln2/4 per second (i.e., the influence of one region on another is expressed within a time frame of 4 s) with 90% confidence? (2) Using Bayesian model comparison, does one of the models provide a substantially better explanation of the observed responses?

The same approach was adopted for the extended model. The three connectivity specifications were: the correct connectivity, a hierarchical connectivity and full connectivity structure (see Fig. 7).

Determining modulatory connectivity

The validity of inferences made using DCM, regarding the site of action of modulatory inputs, was examined using the three-area model. Using a model with correctly specified latent connectivity, the modulatory input 'task' was allowed to act on each interarea connection and the self-connections of V4 and IT. The models are shown in Fig. 8. Bayesian model comparison was then used to determine if one of the models provided a substantially better explanation of the observed responses.

Examining the impact of misspecified regions of interest

The third aim of this validation was to examine the impact of using 'invalid' region of interest (ROI) data when estimating DCMs. This would be equivalent to including fMRI data in a DCM analysis from a region of the brain that was misspecified, or where the signal was dominated by nonspecific responses (noise or artefact). This is possible using the simulated fMRI times-series data because the ROI data comprise a mixture of activity from 'specific' and 'nonspecific' neurons. The valid ROI time-series data were derived from the average activity in the specific and nonspecific neurons; the invalid time-series data were derived using just the nonspecific neuronal activity. Two potential effects

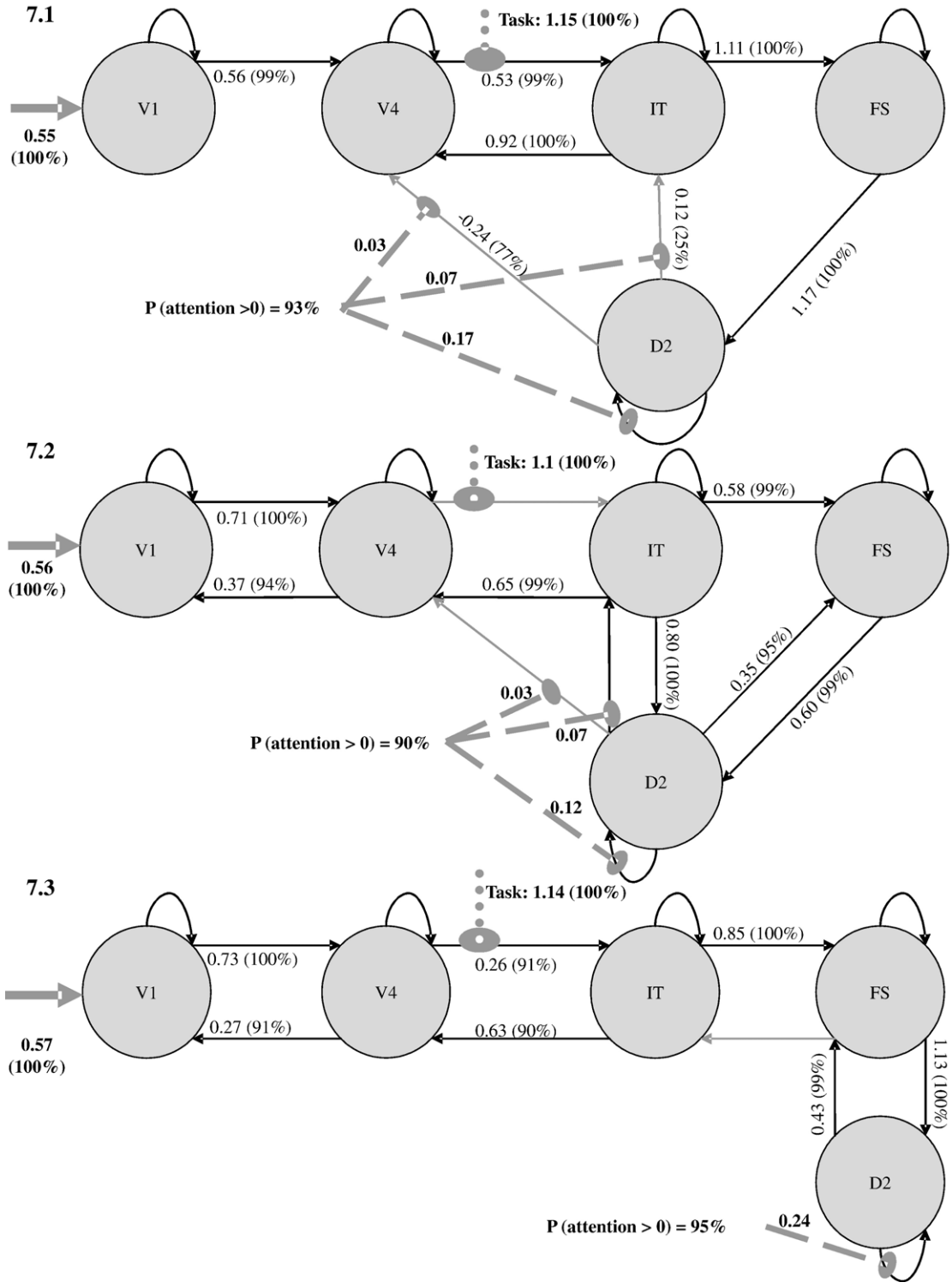

Fig. 7. (7.1) Extended model, with correctly specified interarea connections. Two sets of bilinear effects are specified: shapes vs. scrambled images as in the 3-area model, and the effects of different levels of attention influencing the connections from D2 to D2, IT and V4. Self-connections, direct inputs and modulatory effects of task are displayed as described in Fig. 6. Posterior parameter estimates for the modulatory effect of attention are shown next to dashed grey markers; the values in brackets are the percentage confidence that these values are greater than zero. (7.2) Extended model, results of specifying full interarea connections. Only posterior parameter estimates that exceed criterion are displayed. (7.3) Extended model, results of specifying a hierarchical connection structure.

of misspecified data were examined: (1) the differential impact of misspecifying different regions, and (2) the effects of increasing the proportion of corrupted regions. Bayesian model comparison cannot be used to compare models using different data; therefore, a measure of the conditional uncertainty of the model fit (the trace of the posterior covariance was used). This is the sum of the posterior

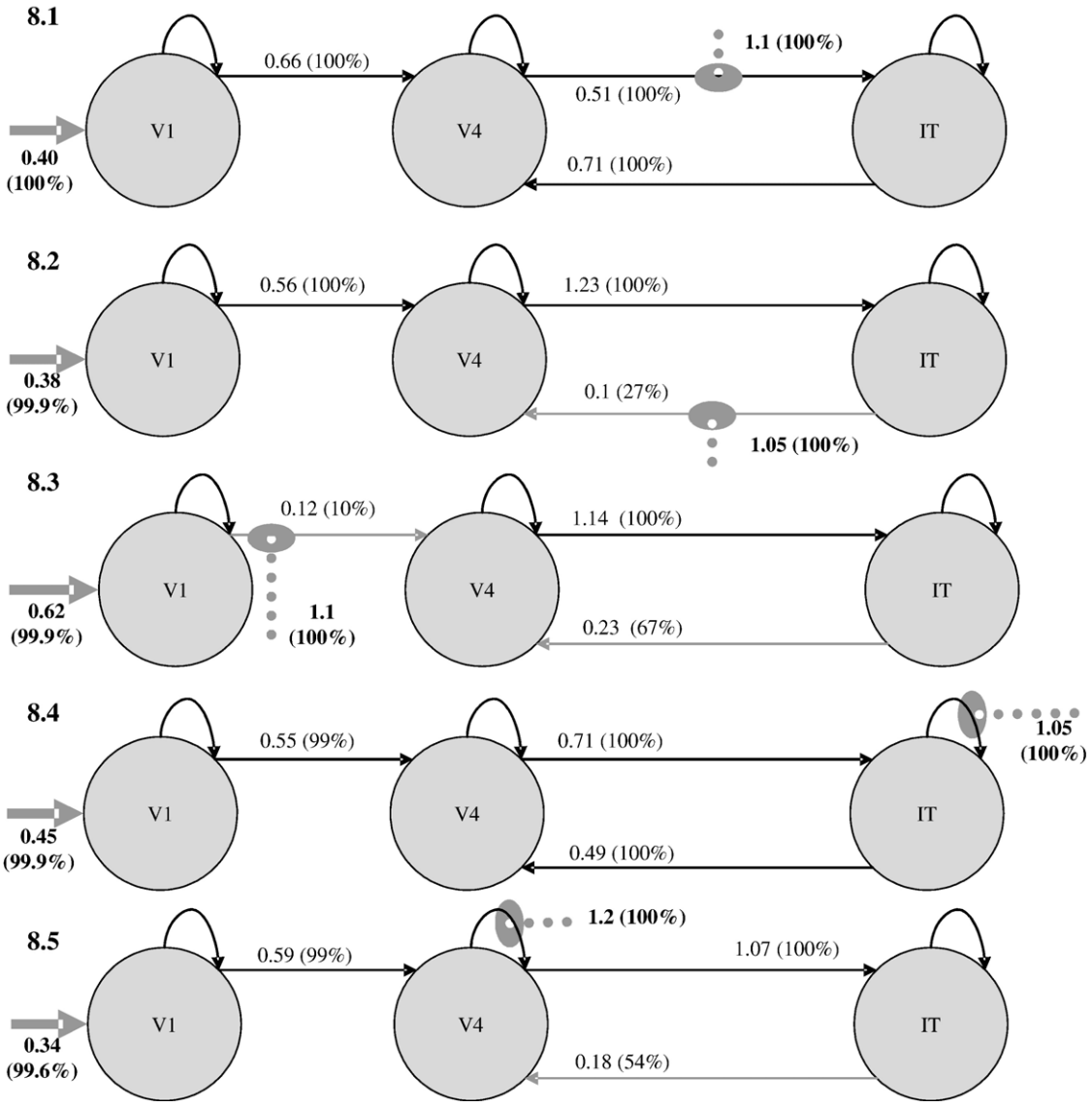

Fig. 8. (8.1–8.5) Simple models, with interarea connections specified correctly. The site of action of modulatory, bilinear effects varies over panels 8.1–8.5. The effect of misspecification of bilinear effects on the strengths of the intrinsic connections can be seen. In all models, DCM suggests that there is a significant bilinear effect.

variance, which increases as the precision of the parameter estimates decreases.

Results

Determining latent connectivity

The combination of DCM and Bayesian model comparison revealed strong evidence for those models with correctly specified connections. Figs. 6.1–6.3 show the three patterns of connectivity specified for the simple three-area model. It can be seen that all the coupling parameters in Fig. 6.1 (correctly specified model) exceed threshold with greater than 90% confidence. In Figs. 6.1 and 6.3, the posterior probability of correctly specified connections exceeded 90% confidence, whereas the incorrect connections do not. The model shown in Fig. 6.2 reduces to the correct model when the coupling parameters that do not exceed threshold are excluded. Bayesian model comparison suggested that there was positive evidence for the 'correct' model (model 1) over the two alternatives (Fig. 9.1).

Figs. 7.1–7.3 shows the patterns of connectivity specified for the more comprehensive five-area model. As with the three-area model, it can be seen that the values of the coupling parameters in the correctly specified model (Fig. 7.1) exceed threshold with greater than 90% confidence or, in the case of the backwards connections from D2 to IT and V4, are significantly modulated by the factor 'attention'. The results shown in Fig. 7.2 are derived from a DCM in which full connectivity was specified. A number of connections were inferred incorrectly: D2 to FS, IT to D2 and V4 to V1. All of the connections present in the computational model either exceed threshold or were significantly modulated by the factors task and attention. Fig. 7.3 shows the estimated coupling parameters when a simple but incorrect hierarchical structure was specified. Bayesian model comparison of the three models

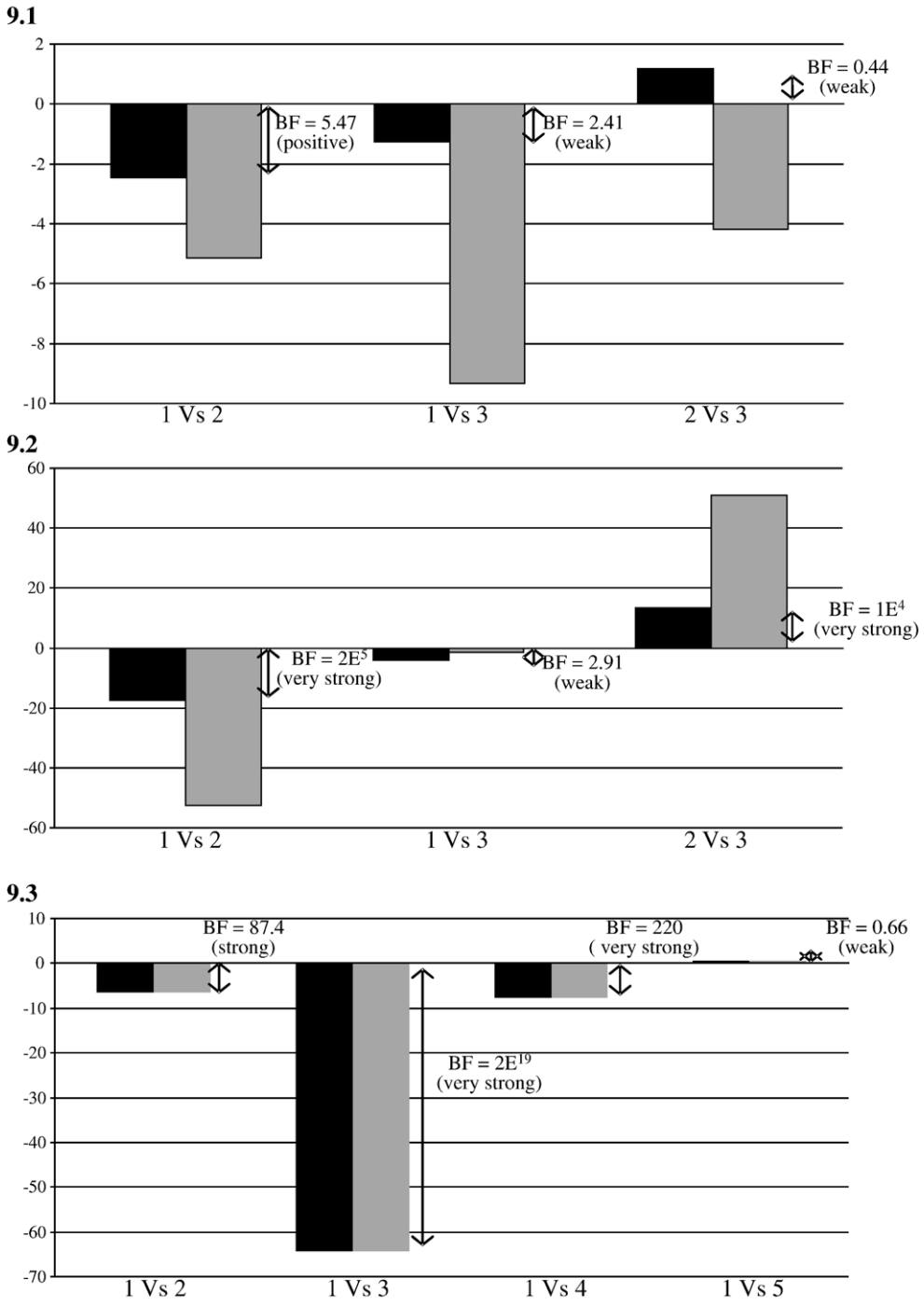

Fig. 9. (9.1) Bayesian model comparison of the three-area models with differently specified intrinsic connectivity. Akaike's Information Criterion (black) and Bayesian Information Criterion (grey) values were used to compute Bayes Factors (BF) for the comparison of the models shown in Figs. 6.1–6.3. The bars represent the log-BF with negative values indicating the first model is more likely. When both AIC and BIC Bayes Factors are greater than e = 2.7183, this is considered consistent evidence in favour of one model. (9.2) BF for the comparison of the five-area models in Figs. 7.1–7.3. (9.3) BF for the comparison of modulatory inputs shown in Figs. 8.1–8.5 and described in Simulations.

provided convincing evidence that the correct model (Fig. 7.1) provides the best explanation of the observed responses (Fig. 9.2).

Determining modulatory connectivity

The effects of incorrectly specifying the site of action of the modulatory input 'task' were explored using the simple three area model. Figs. 8.1–8.5 show the site of action of the factor 'task', and the coupling parameters estimated for these models. In each case, the effect of 'task' is greater than zero (at 90% confidence). The results of Bayesian model comparison are shown in Fig. 9.3. It can be seen that models 1 and 5 (Figs. 8.1 and 8.5) provide the best explanation of the observed responses, but cannot be distinguished from each other. Model 1 was the correct model.

Incorrectly specified regional data

Figs. 10 and 11 show the effects of substituting the simulated fMRI times-series data with invalid time-series data derived from nonspecific neuronal activity. From Figs. 10.1 and 10.2, it can be seen that the overall effect of substituting increasing numbers of corrupted areas increases the conditional uncertainty, as indexed by the posterior covariance. For both the three-area model and the extended five-area model, the precision of the posterior estimates of the coupling parameters tended to decrease as the proportion of ROIs containing nonspecific neuronal activity increased. Effectively, this means that corrupted data or misspecified regional activity increases conditional uncertainty and, by virtue of the shrinkage priors used in DCM (see Penny et al., 2004, p. 1160), decreases the MAP estimates of coupling. This means that one is less likely to infer a connection is present. This is shown in Fig. 11, where increasing the proportion of corrupted regions (relative to the total number of regions) reduces the proportion of latent connections and bilinear effects that are inferred to be present.

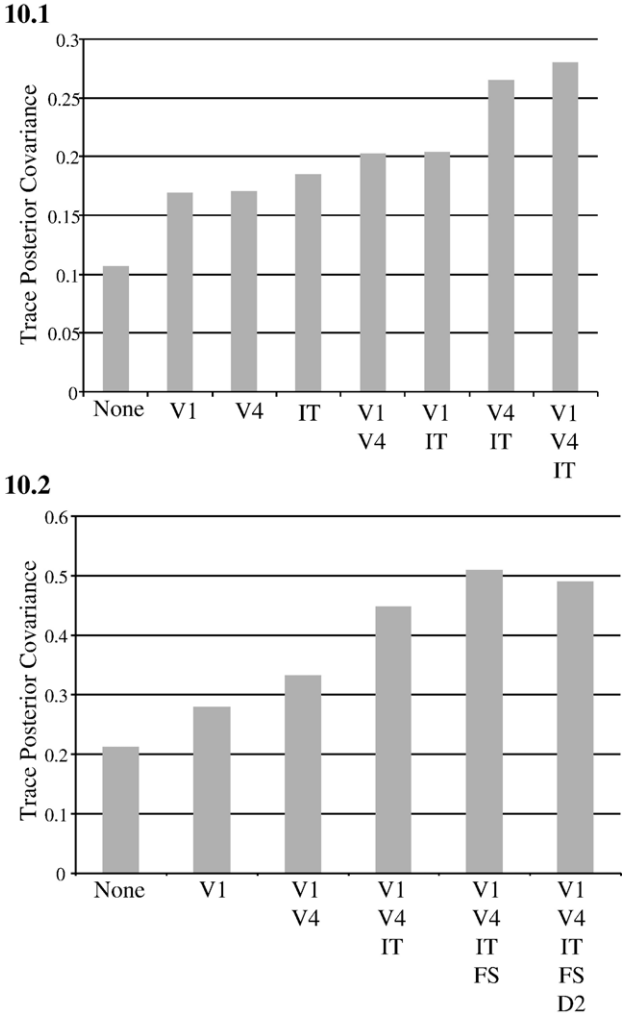

Fig. 10. (10.1) Results of comparing Posterior Covariance as a measure of the effects of substituting corrupted data into the three-area DCM. The heights of the bars represent the trace of the Posterior Covariance. The effects of increasing the number of misspecified areas can be seen. (10.2) As above, but for the extended five-area model, showing the effects of increasing the number of misspecified areas.

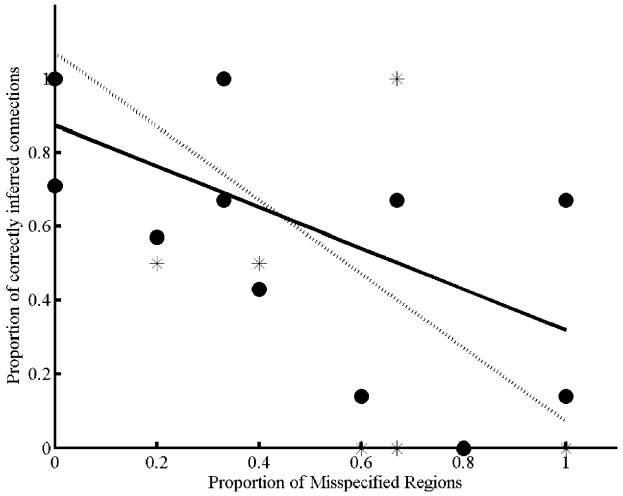

Fig. 11. Linear regression of the effects of increasing the proportion of corrupted regions (relative to the total number of regions) on the proportion of latent connections (black line and dots) and bilinear effects (grey line, asterisks) that are inferred to be present. The correlation coefficients are −0.56 for the latent connections and −0.68 for the bilinear effects.

Discussion

Penny et al. (2004) have previously described the use of Bayesian inference procedures in the context of dynamic causal models. With a single DCM, the posterior distribution can be used to make inferences about changes in coupling produced by experimental manipulations. These inferences are contingent on the model: which regions are connected, and where the modulatory inputs act. Penny et al. established the use of Bayes factors to inform decisions regarding the connectivity of DCMs and as a means to compare models with different structures.

The results presented in this paper extend the findings of Penny et al. and provide further support for the use of AIC and BIC approximations to the model evidence as described previously (Penny et al., 2004). In the present work, synthetic fMRI time-series data generated from a neurobiologically inspired computational model were analysed using DCM. Using inference based on the confidence that the rate of change of effects mediated via coupling parameters exceeded plausible thresholds and Bayesian model comparison, DCM analysis was generally able to correctly identify the connectivity of the computational model (Figs. 6, 7 and 8). In the three-area model, the connectivity structure was detected using the posterior distributions for the coupling parameters alone, with Bayes factors confirming that the correct model provided the best explanation of the data. With the extended five-area model, the posterior estimates of the coupling parameters were less discriminating, but the model comparison procedure provided conclusive evidence in favour of the model with correctly specified connectivity as this provided the most accurate and parsimonious explanation of the data.

Furthermore, Bayesian model comparison favoured those models when the bilinear (modulatory) effects corresponded to their actual implementation in the large-scale neural model. This can be seen in Figs. 8 and 9.3. Models 1 and 5 provide equally good

explanations of the observed responses. These models are the closest to the architecture of the computational model. In the model, activity in area IT varies as a function of the input stimuli (shapes vs. scrambled images) whose effects arrive at IT via activity in V4. In the DCM, the factor 'task' either increases the activity in area V4 via the V4 to V4 self-connections (model 5) or increases the influence of activity in V4 on IT by strengthening the connection (model 1). It is interesting to note that models 1 and 5 are similar in that both speak to a task-dependent interaction between the activity in Areas V4 and IT. In model 1, this is mediated by changes in the connections from V4 and IT. In model 5, the self-connections to V4 are modulated directly. Bayesian model selection was unable to disambiguate these two models because they provide equally accurate and parsimonious explanations of the data. Models 2 and 4 fail to capture the direction of the interaction from V4 to IT because the effect of task is mediated by IT to IT connections or IT to V4 connections, respectively. This has important implications for the interpretation of experimental data: with unreliable a priori knowledge of the anatomical connectivity (frequently the case in human studies), it would be impossible to decide if models 1 or 5 were correct. This would not prevent the experimenter from inferring correctly that the factor 'task' had a nontrivial effect on the interaction between regions V4 and IT.

Fig. 7.1 shows that the magnitude of the changes in effective connectivity, induced by changes in the level of attention, is less than those induced by the factor 'task'. This reflects differences in the magnitude of the effects of attention and task on the neuronal activity within the computational model. Inference about bilinear effects on the coupling parameters was based on a threshold of zero. Bilinear effects represent a change in the coupling which can be negative or positive. Quantitatively small bilinear effects are always interesting, provided we can be confident they are not zero. For example, a 10% change in the strength of a connection with an initial rate of 1 s^{-1} corresponds to a bilinear effect of 0.1%. Although this appears to be a very small effect, it represents a 10% difference in the effective connectivity between the two regions. Hence, it is appropriate to use a threshold of zero for inference about the bilinear (modulatory) effects of experimental manipulations.

Bayesian model comparison provides a means of deciding which of several models provides the best explanation of the observed data; therefore, it is not possible to compare models of different data. With increasing nonspecific activity, the posterior covariance (covariance of the posterior probability distribution) increases, i.e., the precision of the parameter estimates falls. The effects of misspecifying the region of interest data are clearly demonstrated in Fig. 11 where it can be seen that the ability of DCM to properly estimate the coupling parameters and the effects of modulatory inputs decreases as the proportion of incorrectly specified regions increases.

It is important to note that, in both DCMs, even the 5-region model, crucial brain areas that make important contributions to task performance were not included (see Fig. 1). Nevertheless, DCM was usually able to determine the simplified model that corresponded most closely to the actual configuration of regions mediating the task. This is important because it is likely the case for experimental data that the network of brain areas crucial for task performance will contain more regions than DCM or similar methods can handle. Therefore, it is reassuring that good approximations to the actual networks were obtained when using a subset of regions and constrained hypotheses about the interactions between the areas.

It should be noted that although model selection and comparison allow one to adjudicate among different models (i.e., architectures), it does not ensure that the best model is the true model. In other words, model selection with DCM only allows relative statements about some specified models, it does not explore model space exhaustively to find the best of all models and, even if it did, the best model may or may not be the true model; it is simply the most accurate and parsimonious model.

In summary, these findings confirm that DCM can be used at the level of a single model to interrogate data with respect to a particular set of connection strengths or changes in coupling. The use of thresholds, based on posterior densities, provides a means of deciding which connections are expressed functionally or are significantly modulated. These quantitative inferences are conditional on the DCM in question. The architectures of qualitatively different models can be compared using Bayesian model selection. This comparison does not depend on the quantitative parameter estimates for any particular model because the log-evidence is not a function of the parameters. Model selection can be used to disambiguate two models based on their respective evidence, given the data. This is useful in cases where a priori knowledge of the anatomical connectivity and/or site of action of experimental manipulations is imprecise.

Conclusions

These results highlight the essential role of neural models in establishing the validity of estimation and inference schemes, providing further support for the use of DCM as a method of analysis for fMRI data that is able to discern latent connectivity and the modulatory effects of designed experimental manipulations on the strength of connections between brain regions. Specifically, Bayesian model comparison confirms the validity of DCM in relation to a well-characterised and comprehensive model. These findings also speak to the importance of using informed and complicated neural models to test methods of data analysis because in the model, unlike the brain, everything is known. Hence, the results obtained by an analysis technique can be checked for accuracy, completeness and interpretability.

Acknowledgments

This work was supported by the Wellcome Trust (LL and KJF) and the NIDCD Division of Intramural Research (BH). We wish to thank Brent Warner for help with the simulations, and Dr. Fatima Husain for discussions.

References

Friston, K.J., Mechelli, A., Turner, R., Price, C.J., 2000. Nonlinear responses in fMRI: the Balloon model, Volterra kernels, and other hemodynamics. NeuroImage 12, 466–477.

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

Funahashi, S., Chafee, M.V., Goldman-Rakic, P.S., 1993. Prefrontal neuronal activity in rhesus monkeys performing a delayed anti-saccade task. Nature 365, 753–756.

Horwitz, B., 2003. The elusive concept of brain connectivity. NeuroImage 19, 466–470.

Horwitz, B., Tagamets, M.A., 1999. Predicting human functional maps with neural net modeling. Hum. Brain Mapp. 8, 137–142.

Horwitz, B., Friston, K.J., Taylor, J.G., 2000. Neural modeling and functional brain imaging: an overview. Neural Netw. 13, 829–846.

Horwitz, B., Warner, B., Fitzer, J., Tagamets, M.A., Husain, F.T., Long, T.W., 2005. Investigating the neural basis for functional and effective connectivity. Application to fMRI. Philos. Trans. R. Soc. Lond., B Biol. Sci. 360, 1093–1108.

Husain, F.T., Tagamets, M.A., Fromm, S.J., Braun, A.R., Horwitz, B., 2004. Relating neuronal dynamics for auditory object processing to neuroimaging activity: a computational modeling and an fMRI study. NeuroImage 21, 1701–1720.

Lee, L., Harrison, L.M., Mechelli, A., 2003. A report of the functional connectivity workshop, Dusseldorf 2002. NeuroImage 19, 457–465.

Mechelli, A., Price, C.J., Friston, K.J., 2001. Nonlinear coupling between evoked rCBF and BOLD signals: a simulation study of hemodynamic responses. NeuroImage 14, 862–872.

Passingham, R.E., Stephan, K.E., Kotter, R., 2002. The anatomical basis of functional localization in the cortex. Nat. Rev., Neurosci. 3, 606–616.

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

Ramnani, N., Behrens, T.E., Penny, W., Matthews, P.M., 2004. New approaches for exploring anatomical and functional connectivity in the human brain. Biol. Psychiatry 56, 613–619.

Tagamets, M.A., Horwitz, B., 1998. Integrating electrophysiological and anatomical experimental data to create a large-scale model that simulates a delayed match-to-sample human brain imaging study. Cereb. Cortex 8, 310–320.

Zeki, S., Watson, J.D., Lueck, C.J., Friston, K.J., Kennard, C., Frackowiak, R.S., 1991. A direct demonstration of functional specialization in human visual cortex. J. Neurosci. 11, 641–649.