Bayesian fusion and multimodal DCM for EEG and fMRI

Huilin Wei [a,b], Amirhossein Jafarian [b], Peter Zeidman [b], Vladimir Litvak [b], Adeel Razi [b,c,d,e], Dewen Hu [a,*], Karl J. Friston [b,**]

[a] College of Intelligence Science and Technology, National University of Defense Technology, Changsha, Hunan, China
[b] Wellcome Centre for Human Neuroimaging, UCL Institute of Neurology, University College London, London, United Kingdom
[c] Turner Institute for Brain and Mental Health, School of Psychological Sciences, Monash University, Clayton, Australia
[d] Monash Biomedical Imaging, Monash University, Clayton, Australia
[e] Department of Electronic Engineering, NED University of Engineering and Technology, Karachi, Pakistan

ARTICLE INFO

ABSTRACT

This paper asks whether integrating multimodal EEG and fMRI data offers a better characterisation of functional brain architectures than either modality alone. This evaluation rests upon a dynamic causal model that generates both EEG and fMRI data from the same neuronal dynamics. We introduce the use of *Bayesian fusion* to provide informative (empirical) neuronal priors – derived from dynamic causal modelling (DCM) of EEG data – for subsequent DCM of fMRI data. To illustrate this procedure, we generated synthetic EEG and fMRI timeseries for a mismatch negativity (or auditory oddball) paradigm, using biologically plausible model parameters (i.e., posterior expectations from a DCM of empirical, open access, EEG data). Using model inversion, we found that Bayesian fusion provided a substantial improvement in marginal likelihood or model evidence, indicating a more efficient estimation of model parameters, in relation to inverting fMRI data alone. We quantified the benefits of multi-modal fusion with the information gain pertaining to neuronal and haemodynamic parameters – as measured by the Kullback-Leibler divergence between their prior and posterior densities. Remarkably, this analysis suggested that EEG data can improve estimates of haemodynamic parameters; thereby furnishing proof-of-principle that Bayesian fusion of EEG and fMRI is necessary to resolve conditional dependencies between neuronal and hae-modynamic estimators. These results suggest that Bayesian fusion may offer a useful approach that exploits the complementary temporal (EEG) and spatial (fMRI) precision of different data modalities. We envisage the pro-cedure could be applied to any multimodal dataset that can be explained by a DCM with a common neuronal parameterisation.

1. Introduction

Dynamic causal modelling (DCM) was initially proposed to infer effective connectivity from fMRI data using parsimonious one-state-per-region model generators of neuronal activity (Friston et al., 2003, 2014; Li et al., 2011). DCM was subsequently developed to characterise the canonical microcircuits – with laminar structure – underlying event related or induced responses measured with EEG or MEG (David et al., 2006; Chen et al., 2008; Friston et al., 2012). In general, dynamic causal models have a neuronal part and a measurement (or observation) part; for example, the standard haemodynamic model in DCM for fMRI

(Stephan et al., 2007, 2015). These models specify how experimental stimuli induce neuronal dynamics and then how neuronal responses cause observable responses, such as EEG sensor data and BOLD signals. A recent development, the canonical microcircuit DCM for fMRI, offers a canonical microcircuit neural mass model that underwrites neuronal dynamics (Friston et al., 2017). This neuronal model is based on suc-cessive model developments and comparison using DCM for M/EEG (Brown and Friston, 2012; Pinotsis et al., 2012, 2014; Moran et al., 2013; Bhatt et al., 2016). Compared with the previous one-state-per-region neuronal model (commonly used in DCM for fMRI), the canonical microcircuit neural mass model furnishes a description of neuronal

---

activity at the mesoscale (i.e., within-region; while extant models only describe neuronal activity at the macroscale; i.e., between-region). Using canonical microcircuit neural mass models for fMRI also brings the following benefits: first, it provides a generative or forward model of measurable signals that retains a degree of neurobiological plausibility. Experimental inputs (e.g., visual stimuli, attention modulation, etc.) perturb neuronal dynamics within coupled canonical microcircuits, each comprising a laminar specification of cell types and their inter-connectivity. These neuronal populations elaborate forward and back-ward connections to model cortical hierarchies, where these laminar specific connections also play an important role in neuronal process theories based upon predictive coding (Friston, 2005; Bastos et al., 2012, 2015; Markov et al., 2013). Second, it equips a haemodynamic model with a laminar specific neuronal drive. These neuronal afferents further characterise a laminar specification of neurovascular architecture that may be useful for modelling high resolution fMRI (Heinzle et al., 2016; Markuerkiaga et al., 2016; Scheeringa et al., 2016; Huber et al., 2018; Kashyap et al., 2018; Uludag and Blinder, 2018).

However, from a computational perspective, inverting such neural mass models using fMRI data may not be efficient (due to its poor temporal resolution), which means that some model parameters could not be optimised during model inversion (Friston et al., 2017). In contrast, there is more information in electromagnetic data about underlying neuronal and synaptic parameters, these parameters can be estimated with greater conditional certainty (David et al., 2006). This intuitively suggests that a multimodal fusion strategy should be considered. In response to this, a straightforward approach is to treat the neuronal and haemodynamic model parameters as conditionally independent, i.e., they can be efficiently estimated separately using EEG and fMRI data (Rosa et al., 2011). Along these lines, a recently introduced method within the DCM framework begins by deriving neuronal responses from a neural mass model of MEG data, which are then used to drive a haemodynamic model fitted to fMRI data (Jafarian et al., 2019). This approach provides an efficient pipeline for multi-modal integration. However, in this paper, we preclude the use of independence assumptions about neuronal and hae-modynamic parameters. Therefore, we instead perform multimodal *Bayesian fusion*, in which the (posterior) estimates of the parameters from one imaging modality are used as the priors for the other. This enables all parameters to be informed by both the M/EEG and fMRI data. Specifically, under this scheme, one can first employ electrophysiological measurements such as M/EEG to constrain posterior estimates of canonical microcircuitry, then take these posteriors as (empirical) neuronal priors for a subsequent inversion using fMRI data. This second inversion furnishes (electrophysiologically) informed estimates of regionally specific haemodynamic parameters. This form of fusion, which rests upon a DCM with a common neuronal parameterisation, is known as *Bayesian fusion* via *Bayesian belief updating* (Friston et al., 2017). The resulting multimodal Bayesian fusion thereby provides a characterisation of functional brain architectures that properly combines two sources of information that are well-resolved either in time (electromagnetic) or space (fMRI).

In this technical note, we assess the benefits of multimodal Bayesian fusion when making inferences about neuronal and haemodynamic responses using a canonical microcircuit DCM. The following paper comprises three sections. The first section introduces the theoretical foundations of multimodal DCM and Bayesian fusion. In this section, we combine the two strands of Bayesian modelling to provide a generative model of DCM suitable for EEG-fMRI fusion. We describe this generative model in detail, and introduce a pipeline for implementing multimodal Bayesian fusion. The second section provides an illustrative application of Bayesian fusion using a synthetic multimodal dataset. We first call on the generative model described previously and biologically plausible parameters (i.e., posterior expectations based upon a standard DCM analysis of empirical EEG data) to generate EEG and fMRI timeseries for a mismatch negativity paradigm (Garrido et al., 2007). Having simulated multimodal data (with appropriate levels of measurement noise), we

implement Bayesian fusion, especially focusing on the Bayesian belief updating procedure. We also conduct model inversion using just fMRI data, for later comparison with Bayesian fusion. The quality of the ensuing inversions is assessed in terms of the log evidence (as approximated by the variational free energy). We hypothesised that the unimodal fMRI analyses would be less able to identify detailed changes in microcircuitry (i.e., mismatch effects), whereas the Bayesian fusion would provide more precise estimates – an accompanying increase in the efficiency of parameter estimation. The last section assesses the usefulness of EEG and fMRI data in terms of their relative information; i.e., information gain as measured by the Kullback-Leibler divergence (Kullback and Leibler, 1951; Zeidman et al., 2018), pertaining to neuronal and haemodynamic parameters associated with the two data modalities and their fusion. This *Bayesian data comparison* provides a quantitative way to evaluate the benefits of multimodal Bayesian fusion – and the degree to which distinct data modalities resolve uncertainty about unknown model parameters.

2. Methods and materials

2.1. Multimodal DCM for EEG and fMRI

Dynamic causal modelling suitable for multimodal fusion adopts a common neuronal model but modality specific observation models, to explain multiple measurements (e.g. M/EEG, fMRI timeseries) as the observable consequence of neuronal activity. Fig. 1 provides a schematic summary of the generative model of multimodal DCM for EEG and fMRI, in which the neuronal part is described by the canonical microcircuit neural mass model (Friston et al., 2017), while the observation part is a standard electromagnetic forward model for EEG (Kiebel et al., 2006) and the haemodynamic model for fMRI (Stephan et al., 2007), respectively. All corresponding variables are listed in Table 1 and Table 2.

The right panel of Fig. 1 illustrates the structure of the canonical microcircuit and the equations modelling neuronal dynamics. The canonical microcircuit (per region or source) comprises four neuronal populations, corresponding to spiny stellate cells (denoted by population 1), superficial pyramidal cells (population 2), deep pyramidal cells (population 4), and inhibitory interneurons (population 3). Population 1, 2, and 4 model the granular, supragranular, and infragranular excitatory cells respectively, while inhibitory interneurons constitute the only inhibitory population. Note that the inhibitory interneurons in Fig. 1 are shown to be restricted to supragranular layer. In reality, inhibitory interneurons are distributed over the cortical layers; however, they are usually represented in a single layer as the primary source of inhibition originates from cells in the same layer. This reflects the prevalence of inhibitory intralaminar connections, evidence for which comes primarily from the optogenetics literature, e.g., Katzel et al. (2011). The four populations are coupled via inter- and intralaminar intrinsic connectivity (Thomson and Bannister, 2003), where light blue connections in the figure denote intrinsic excitatory connectivity (mediated by intrinsic connectivity strength $a_{ij}, a_{ij} > 0$, ij indicates the direction from j to i). The pink connections represent intrinsic inhibitory connectivity (mediated by $a_{ij}, a_{ij} < 0$), and the pink circles denote the recurrent (self) connections that are universally inhibitory (mediated by self-inhibition strength G_i). In addition to the intrinsic (within-region) connectivity, there exists extrinsic (between-region) connectivity (dark blue connections in Fig. 1) with empirically and statistically established laminar-specific origins and targets (Haeusler and Maass, 2007; Sohal et al., 2009; Ninomiya et al., 2012; Friston et al., 2017). These extrinsic connections can be divided into forward and backward connections in cortical hierarchies; where forward connections arise from superficial pyramidal cells of a lower-level region and target spiny stellate cells in the granular layer of a higher region. In Fig. 1 these forward connections are mediated by $A_{kl}^{(12)}$. In the current canonical microcircuit model, these forward connections are supplemented with weak (possibly polysynaptic) connections to deep
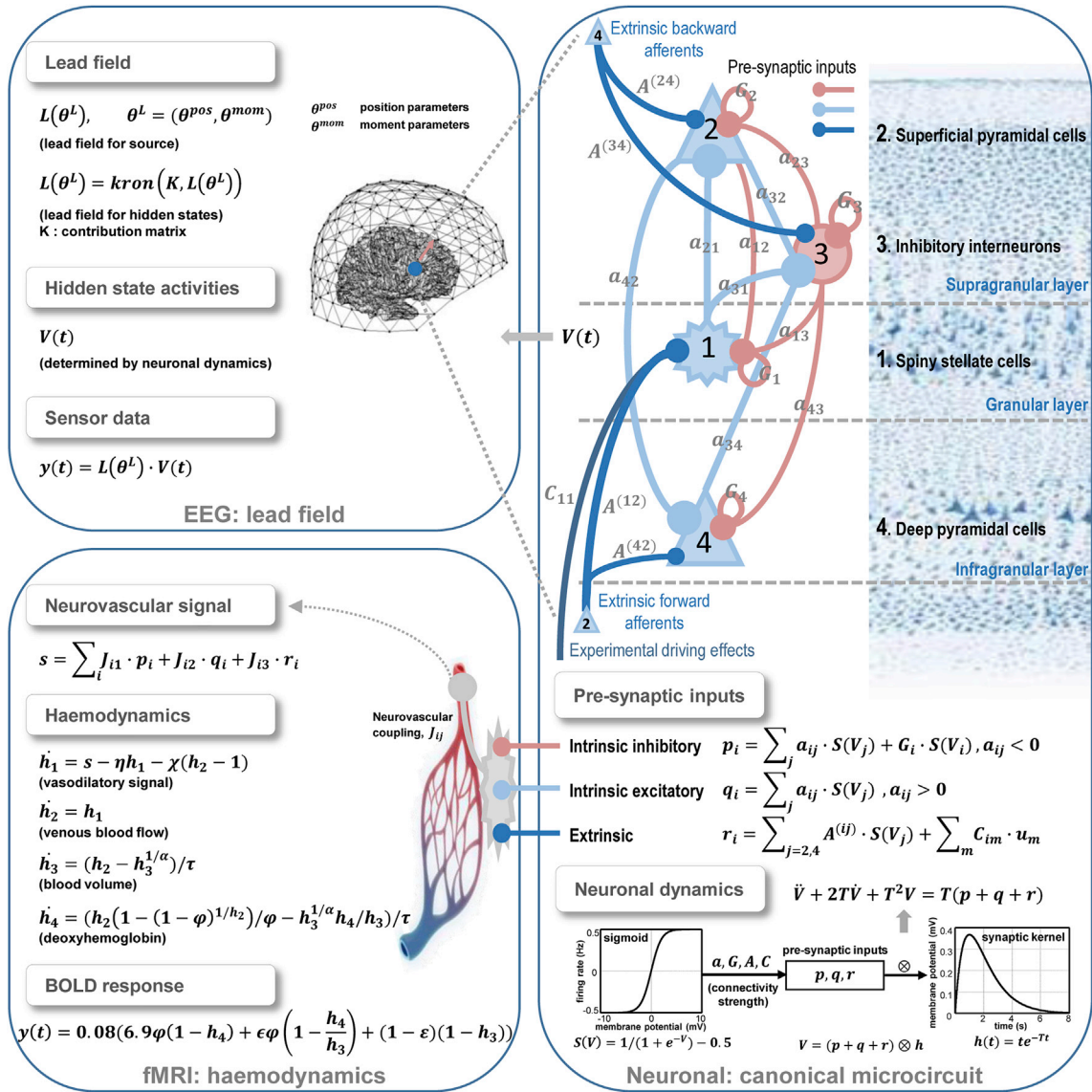
Fig. 1. The generative model of multimodal DCM for EEG and fMRI (per region). This model comprises two parts: the common neuronal part (right panel), and a modality-specific observation part (left panel). Responses to experimental inputs – perturbing neuronal dynamics – are modelled with a canonical microcircuit. This microcircuit comprises four neuronal populations denoted in distinct cortical layers (where inhibitory interneurons distributed over cortical layers are lumped together into a single population), each population receives three sorts of pre-synaptic inputs (scaled by respective connectivity strength parameters), corresponding to intrinsic inhibitory afferents (denoted by pink lines and circles), intrinsic excitatory afferents (light blue lines), and extrinsic (forward and backward) afferents (dark blue lines). Each population is equipped with a hidden state, whose dynamics are described by the second-order ordinary differential equation in the figure. The observation models are a standard lead field for EEG and a standard haemodynamic model for fMRI. For EEG, the lead field (determined by some EEG spatial parameters) is a spatial linear mapping that transforms neuronal activity into the EEG sensor data. For fMRI, the observation model comprises neurovascular coupling and haemo-dynamics. The neurovascular coupling describes how neuronal dynamics induce a vasoactive signal. The neurovascular signal is generated via a linear combination of the pre-synaptic inputs (scaled by the neurovascular coupling parameters). The subsequent haemodynamics describe how neurovascular signal drives a standard haemodynamic model to generate the final BOLD response. Please see Tables 1 and 2 for a description and parameterisation of the variables.

pyramidal cells (mediated by $A^{(42)}_{kl}$, see also Table 1) to ensure dynamical stability as described in (Bastos et al., 2015). Extrinsic backward connections arise from deep pyramidal cells of a higher-level region and target superficial pyramidal cells and inhibitory interneurons in the level below (mediated by $A^{(24)}_{kl}$, $A^{(34)}_{kl}$). Intrinsic and extrinsic connections are based on laminar specific anatomical connectivity in the visual hierarchy (Ninomiya et al., 2012) and the computational architectures implied by predictive coding models of hierarchical message passing in the brain (Bastos et al., 2012, 2015). The validity of the canonical microcircuit model used here was established using nonhuman primate ECoG data and Bayesian model comparison (Bastos et al., 2015; Friston et al., 2017). It should be noted that alternative extrinsic connectivity models can be

evaluated within the current framework using Bayesian model comparison in the usual way. Experimental inputs can directly drive or modulate neuronal responses. The experimental driving effects are determined by the parameters of a C matrix, which target spiny stellate cells in each region. The experimental modulatory effects are specified by changes in intrinsic and/or extrinsic connectivity strength, which are parameterised as B matrices.

Based on the architecture of the canonical microcircuit, the neuronal dynamics are described by the second-order ordinary differential equations of motion for $V(t)$ as shown in Fig. 1, where $V(t)$ are the mean transmembrane potentials of populations (i.e., hidden neuronal states). In more detail, experimental inputs change hidden neuronal states, where a sigmoid activation function transforms the transmembrane

Table 1
Neuronal parameters.

	Description	Parameterisation	Prior
$T_i^{(k)}$	Synaptic rate constant of the i-th neuronal population in region k	$\exp(\theta_T) \cdot T_i$ $T = [256, 128, 16, 32]$	$p(\theta_T) = \mathrm{N}\left(0, \frac{1}{16}\right)$
$a_{ij}^{(k)}$	Intrinsic connectivity from population j to population i in region k	$\exp(\theta_S) \cdot a$ $\quad a = \begin{bmatrix} -8 & -4 & -4 & 0 \\ 4 & -8 & -2 & 0 \\ 4 & 2 & -4 & 2 \\ 0 & 1 & -2 & -4 \end{bmatrix}$ where S is the intrinsic gain	$p(\theta_S) = \mathrm{N}\left(0, \frac{1}{16}\right)$
$G_i^{(k)}$	Self-inhibition of the i-th neuronal population in region k	$\exp(\theta_G) \cdot a_{ii}$	$p(\theta_G) = \mathrm{N}\left(0, \frac{1}{16}\right)$
$A_{kl}^{(ij)}$	Extrinsic connectivity from population j in region l to population i in region k	$\exp(\theta_A) \cdot A^{(12)}, \; A^{(12)} = 1024\exp(\theta_A) \cdot A^{(42)}, \; A^{(42)} = 512$ $\exp(\theta_A) \cdot A^{(24)}, \; A^{(24)} = 512\exp(\theta_A) \cdot A^{(34)}, \; A^{(34)} = 512$	$p(\theta_A) = \mathrm{N}\left(0, \frac{1}{8}\right)$
B_{klm}	Change in connectivity caused by the m-th input	θ_B	$p(\theta_B) = \mathrm{N}\left(0, \frac{1}{8}\right)$
$C_{im}^{(k)}$	Direct driving effect of the m-th input on population i in region k	$\exp(\theta_C)$	$p(\theta_C) = \mathrm{N}\left(0, \frac{1}{32}\right)$
D_{kl}	Conduction delay from region l to region k	$\exp(\theta_D)$	$p(\theta_D) = \mathrm{N}\left(0, \frac{1}{32}\right)$

Table 2
Haemodynamic parameters.

	Description	Parameterisation	Prior
J_{ij}	Neurovascular coupling of the i-th neuronal population (j=1,2,3)	θ_J	$p(\theta_J) = \mathrm{N}\left(0, \frac{1}{16}\right)$
$\eta^{(k)}$	Rate of vasodilatory signal decay per sec in region k	$0.64 \cdot \exp(\theta_\eta)$	$p(\theta_\eta) = \mathrm{N}\left(0, \frac{1}{256}\right)$
χ	Rate of flow-dependent elimination	$0.32 \cdot \exp(\theta_\chi)$	$p(\theta_\chi) = \mathrm{N}(0, 0)$
$\tau^{(k)}$	Rate of haemodynamic transit per sec in region k	$2.00 \cdot \exp(\theta_\tau)$	$p(\theta_\tau) = \mathrm{N}\left(0, \frac{1}{256}\right)$
α	Grubb's exponent	$0.32 \cdot \exp(\theta_\alpha)$	$p(\theta_\alpha) = \mathrm{N}(0, 0)$
ε	Intravascular: extravascular ratio	$1.00 \cdot \exp(\theta_\varepsilon)$	$p(\theta_\varepsilon) = \mathrm{N}\left(0, \frac{1}{256}\right)$
ϕ	Resting oxygen extraction fraction	$0.40 \cdot \exp(\theta_\phi)$	$p(\theta_\phi) = \mathrm{N}(0, 0)$

potentials of populations into firing rates $S(V)$, which are then weighted by intrinsic (a, G) and extrinsic (A, C) connectivity parameters to form the pre-synaptic inputs to each population. According to the categories of intrinsic and extrinsic connections, the pre-synaptic inputs at each population can be divided into three sorts, corresponding to intrinsic inhibitory afferents (denoted by p, pink colour), intrinsic excitatory afferents (q, light blue), and extrinsic afferents (r, dark blue). The subsequent dynamics (Jansen and Rit, 1995; Moran et al., 2013) effectively convolve pre-synaptic inputs with a synaptic kernel to depolarise post-synaptic populations, where T denotes population-specific post-synaptic rate constants. Conduction delays (denoted by D) are parameterised with values that correspond to the time taken for axonal propagation between regions (~16 ms) (Friston et al., 2017). All the above parameters, which we collectively refer to the *neuronal parameters* (see Table 1), are free parameters of the model that can be estimated during model inversions for both EEG and fMRI data.

The left panel of Fig. 1 shows the modality specific observation models corresponding to EEG and fMRI data. For EEG, the observation model is a spatial linear mapping that transforms the estimated source (neuronal) activities $V(t)$ into the EEG sensor (scalp) data $y(t)$. This mapping or matrix operator is called the lead field $L(\theta^L)$ – and is determined by some spatial parameters including three location and three orientation (or moment) parameters $\theta^L = (\theta^{pos}, \theta^{mom})$ (Kiebel et al., 2006, 2007a). A contribution matrix K is used to weight the depolarisation of all populations to the measurable (dendritic) signal from each source. The dominant contribution is usually from the pyramidal population, which is the prominent source that can be detected remotely on the scalp surface in EEG (because the spatial arrangement of pyramidal cell dendrites is perpendicular to the cortical surface) (David et al., 2006; Kiebel et al., 2006).

For fMRI, the observation model comprises neurovascular coupling and haemodynamics. Neurovascular coupling describes how neuronal

dynamics induce a vasoactive signal; here, we assume that each sort of pre-synaptic input to a neuronal population (i.e., intrinsic inhibitory, intrinsic excitatory and extrinsic) is accompanied by a collateral input of the same strength to nearby astrocytes. The neurovascular signal (denoted by s in Fig. 1) is then generated via a linear combination of all pre-synaptic drives in a region (scaled by neurovascular coupling parameters J_{ij}) (Friston et al., 2017). The subsequent haemodynamics describe how neurovascular signal drives a standard haemodynamic model to generate the BOLD response (Friston et al., 2000; Buxton et al., 2004; Stephan et al., 2007, 2015).

In brief, a neurovascular signal s enters the haemodynamic model and drives a vasodilatory signal h_1, blood flow h_2, which responds to the vasodilatory signal and causes changes in blood volume h_3 and deoxyhemoglobin h_4. Finally, the observed BOLD signal $y(t)$ is generated by a nonlinear function of volume and deoxyhemoglobin. The parameters in the haemodynamic state equations and neurovascular coupling parameters – which we collectively refer to the *haemodynamic parameters* – are described in detail in Table 2. These modality-specific parameters of the observation models can be estimated with the corresponding data modality.

Note that this haemodynamic model is the standard haemodynamic model that is released as part of the SPM software. There are a number of variants of this model in the literature that may be appropriate for different imaging modalities at different field strengths. This remains active area of research; e.g., (Kim et al., 2007; Fabiani et al., 2014; Anenberg et al., 2015; Havlicek et al., 2015; Poplawsky et al., 2015; Heinzle et al., 2016; Puckett et al., 2016).

2.2. Multimodal Bayesian fusion

Having a generative model for multimodal fusion means that one can use multiple data modalities to inform the parameters of the same model.

Generally, the data modalities should have a complementary nature; namely, one modality would provide more precise constraints on the parameters that are estimated less efficiently using the other modality – and *vice versa*. Therefore, in EEG-fMRI fusion, we take advantage of high spatial specificity of fMRI and high temporal resolution of EEG, respectively, to provide complementary and informative constraints on model parameters. These observations license the following pipeline for fusing EEG and fMRI data (i.e., Bayesian fusion) with canonical microcircuit DCM:

1. Identify regions significantly activated by an experimental paradigm, using whole brain (SPM) analysis of fMRI data. These regions constitute the network architecture for subsequent DCM analysis.
2. Conduct a DCM analysis of EEG data by setting the prior location of sources (i.e., θ^{pos} in Fig. 1) to the location of the activated regions from step 1, to obtain the posterior densities over neuronal parameters.
3. Conduct a subsequent DCM analysis of fMRI data by replacing the uninformative priors over neuronal parameters (Gaussian shrinkage priors; see Table 1) with the posterior means and covariances from step 2 (i.e., Bayesian belief updating), to inform posterior estimates of haemodynamic parameters.

Note that the above Bayesian fusion enables the acquisition and inversion of EEG and fMRI data separately through a process of Bayesian belief updating. Therefore, the form of fusion – afforded by using the same neuronal model across modalities – frees one from the overhead of having to acquire simultaneous EEG-fMRI measurements. This means that the range of applications for this technique goes beyond simultaneous data acquisition; provided the multimodal data are generated under a common experimental paradigm. This ensures the assumption – that the underlying neuronal dynamics are the same – remains valid.

In the following sections, we perform systematic numerical analyses to quantify any increase in the efficiency of parameter estimation by using Bayesian fusion of EEG and fMRI relative to unimodal inversion of fMRI data alone. In brief, we first simulate synthetic EEG and fMRI timeseries under the same task and neuronal model. We then implement two inversion schemes for synthetic multimodal data, corresponding to Bayesian fusion (following the pipeline above) and inversion using only fMRI data. Finally, we introduce information gain to quantify the relative utility of each modality and the data assimilation enabled by Bayesian fusion.

2.3. Simulations: multimodal data for mismatch negativity paradigm

We simulated EEG and fMRI timeseries for a mismatch negativity (or auditory oddball) paradigm that has been extensively explored using DCM under the mechanism hypothesis of predictive coding (Garrido et al., 2007, 2009a, 2009b; Kiebel et al., 2007a, 2009). To ensure the synthetic data were biologically plausible, we first conducted a standard DCM analysis using empirical, open access data (https://www.fil.ion.ucl.ac.uk/spm/data/eeg_mmn/). We used the ensuing estimates of effective connectivity and synaptic (and observation noise) parameters to generate synthetic data.

Early DCM studies of the mismatch negativity identified a hierarchical network of five cortical sources to explain evoked responses elicited by standard and oddball stimuli. These sources included left and right primary auditory cortex (left A1, right A1), left and right superior temporal gyrus (left STG, right STG), and right inferior frontal gyrus (right IFG) (Garrido et al., 2007, 2009a; Kiebel et al., 2007a, 2009). In our simulations, we defined a hierarchical architecture comprising these five sources: the sources were hierarchically connected via extrinsic forward and backward connectivity. The extrinsic forward connections arose from superficial pyramidal cells of left A1, right A1 and right STG, and targeted spiny stellate cells and deep pyramidal cells of left STG, right STG and right IFG, respectively. The extrinsic backward connections arose from deep pyramidal cells of right IFG, right STG and left STG,

to superficial pyramidal cells and inhibitory interneurons of right STG, right A1 and left A1, respectively (see the dark blue connections in Fig. 2B).
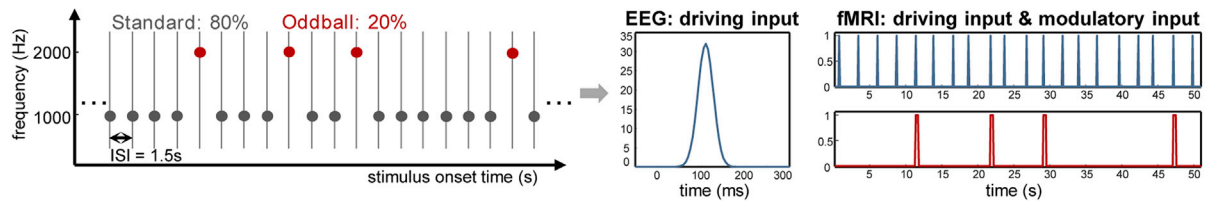
In addition to these extrinsic connections, there were intrinsic inhibitory and excitatory connections within and among the four populations that constituted the canonical microcircuit for each source (the pink and light blue connections in Fig. 2B). External experimental stimuli (i.e., driving inputs) arrived at the spiny stellate populations in bilateral auditory sources. Modulatory effects (i.e., modulatory inputs), which were responsible for explaining the differences in responses to standard and oddball conditions, were specified by changes in: (i) extrinsic forward connectivity from superficial pyramidal cells of left A1, right A1 and right STG, to spiny stellate cells of left STG, right STG and right IFG, respectively; (ii) extrinsic backward connectivity from deep pyramidal cells of right IFG, right STG and left STG, to superficial pyramidal cells of right STG, right A1 and left A1, respectively; and (iii) self-inhibitions of superficial pyramidal cells in bilateral auditory sources (see Fig. 2B). The precise details of this model of the mismatch negativity are not crucial for the purposes of this paper; however, we ensured that the model generating these synthetic data was as close as possible to real neuronal networks – and their context or condition sensitive connectivity.

After specifying this model, we conducted a DCM analysis using the empirical EEG data, and recovered a total of 74 biologically plausible parameter estimates, which included 12 extrinsic connectivity parameters in extrinsic connectivity **A** matrices, 8 modulating effect parameters in **B** matrices, 2 driving effect parameters in vector **C**, an intrinsic gain parameter in **S**, 20 self-inhibition parameters in **G**, 20 synaptic rate constant parameters in **T**, and 11 conduction delay parameters in **D**. The posterior expectations of these empirically determined neuronal parameters were then used as the parameter values generating the neuronal dynamics in the subsequent generation of synthetic EEG and fMRI data. The above connectivity architecture and the parameter values are shown in Fig. 2B.

For the generation of synthetic EEG, in addition to the neuronal parameters, we specified EEG spatial parameters using the lead field employed in the above DCM analysis of empirical data. These included three location and three orientation parameters for each source (the values of these spatial parameters are shown in Fig. 2B). The experimental (auditory) driving input was modelled as a Gaussian shape function of peristimulus time (see Fig. 2A). Finally, we used the forward model of the canonical microcircuit DCM for EEG to generate sensor data for standard and oddball conditions, respectively. The simulated sensor signals with observation noise served as the synthetic EEG data in subsequent Bayesian fusion analyses. A plausible level of observation noise was created by convolving Gaussian noise (with a standard deviation of one half of the generated response amplitude) with a smoothing kernel of eight time bins (4 ms). The resulting synthetic EEG data are shown in Fig. 3A.

Synthetic fMRI data were generated using the same neuronal parameters by specifying the timing of events in a simulated event-related fMRI of the auditory oddball paradigm, as well as a set of appropriate haemodynamic parameters. To ensure the synthetic multimodal data corresponded to the same experimental paradigm (and the two data modalities were generated by the same neuronal responses), we used the EEG mismatch negativity paradigm to derive an event-related auditory oddball paradigm for fMRI. More precisely, we read the trial definition file of the empirical EEG data to recover the stimulus onset times (SOTs) of all events (80% of standard events and 20% of oddball events in a pseudo-random sequence, with inter-stimulus interval of 1.5 s, see Fig. 2A). We then added a random jitter (of −0.25 to 0.25 s, uniformly distributed) to define the SOTs for the event-related fMRI paradigm (Kiebel et al., 2007b). Given that the neuronal responses were driven by all auditory stimuli, the modulating effects were restricted to oddball events. We modelled the driving input as a stick function encoding the SOTs of all events, and encoded the modulatory input as a box-car function (a box-car function allowed the changes in connectivity to

(A) Mismatch negativity paradigm and DCM input signals

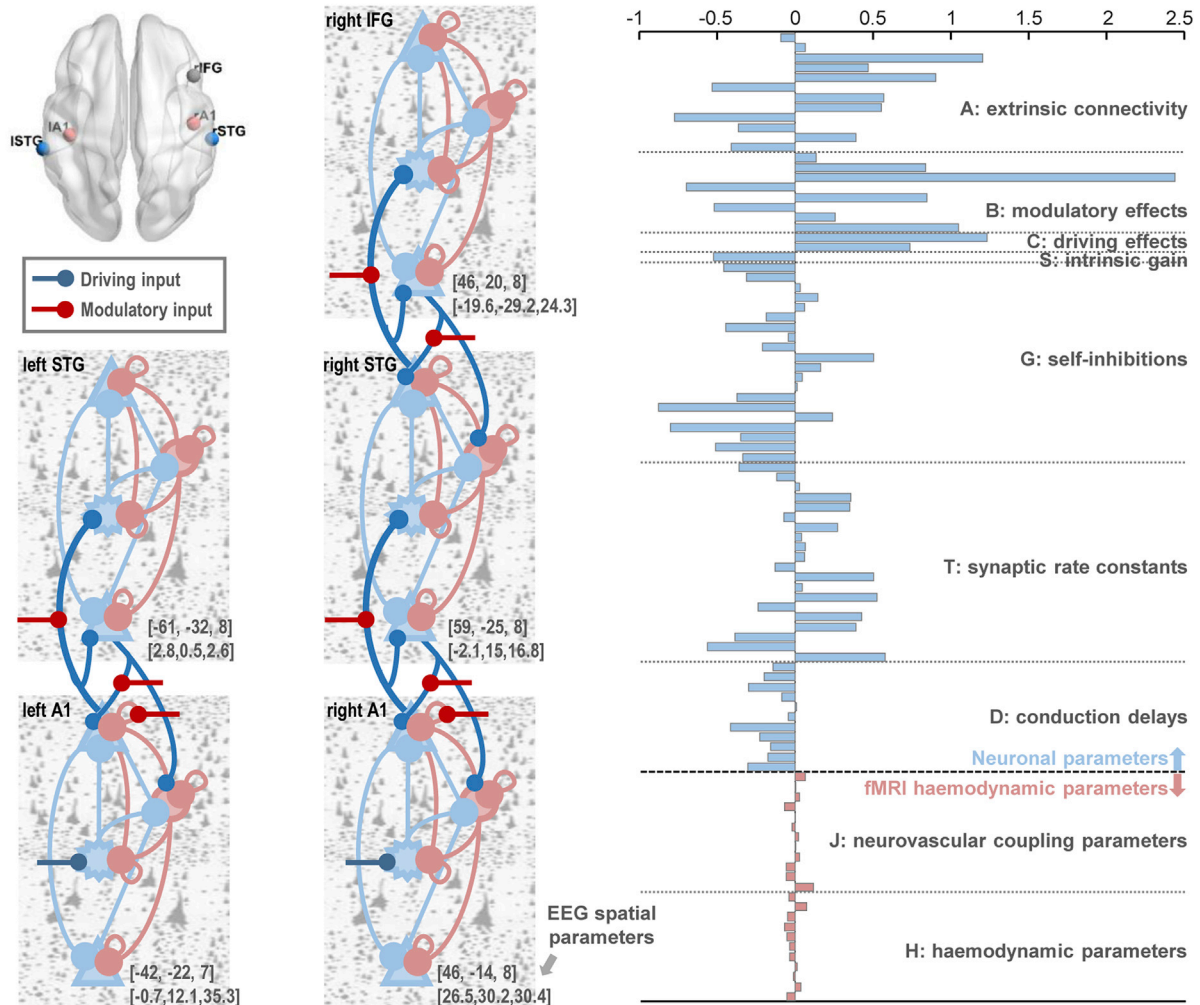## (B) Connectivity architecture and true parameter values

Fig. 2. Experimental inputs, connectivity architecture and parameter values used for simulations. (A) Mismatch negativity paradigm and DCM input signals. The mismatch negativity paradigm contains standard (1000 Hz) and oddball tones (2000 Hz), occurring 80% and 20% respectively in a pseudo-random sequence, with an inter-stimulus interval (ISI) of 1.5 s. The DCM driving input for EEG data is modelled as a Gaussian bump function of peristimulus time, while – for fMRI – the driving input is a stick function encoding stimulus onset times (SOTs) of all tones, where the modulatory input is a box-car function following the SOTs of oddball events. (B) Connectivity architecture and parameter values. This architecture comprises five regions; including left and right A1 (A1 = primary auditory cortex), left and right STG (STG = superior temporal gyrus) and right IFG (IFG = inferior frontal gyrus), visualised using the BrainNet Viewer (Xia et al., 2013). Each region is a canonical microcircuit with inhibitory (pink) and excitatory (light blue) intrinsic connections, where the five regions are coupled via extrinsic forward and backward connections (dark blue). The driving inputs (cyan) target spiny stellate cells in bilateral A1 regions, and the modulatory inputs (red) are specified by changes in either intrinsic or extrinsic connections. Values of neuronal and haemodynamic parameters used for generating synthetic multimodal data are shown using a bar chart.

persist for the duration of neuronal transients) following the SOTs of oddball events (see Fig. 2A). The TR was set to 1.7 s to emulate a typical repetition time. The haemodynamic parameters were sampled from their prior densities in Table 2, which included 12 neurovascular coupling parameters denoted by **J** and 11 haemodynamic parameters denoted by **H** for the five regions (the values are shown in Fig. 2B). After specifying the experimental inputs, the neuronal parameters and the haemodynamic parameters, we generated BOLD responses in the five regions. We

then added appropriate level of observation noise (with a typical signal-to-noise ratio of four) to the simulated BOLD responses (Kiebel et al., 2007b; Heinzle et al., 2016). The amplitude of observation noise is estimated as part of model inversion in the same way that connectivity parameters estimated. In other words, approximate posterior densities over noise covariance parameters are optimised with respect to the variational free energy (i.e., marginal likelihood) (Friston et al., 2007). However, we did not account for physiological noise such as cardiac and
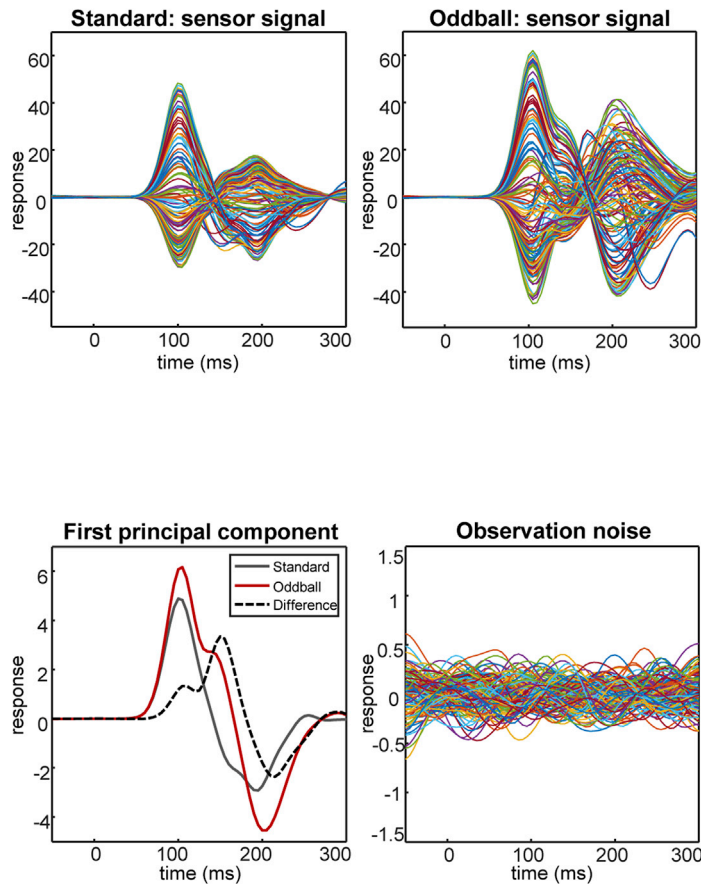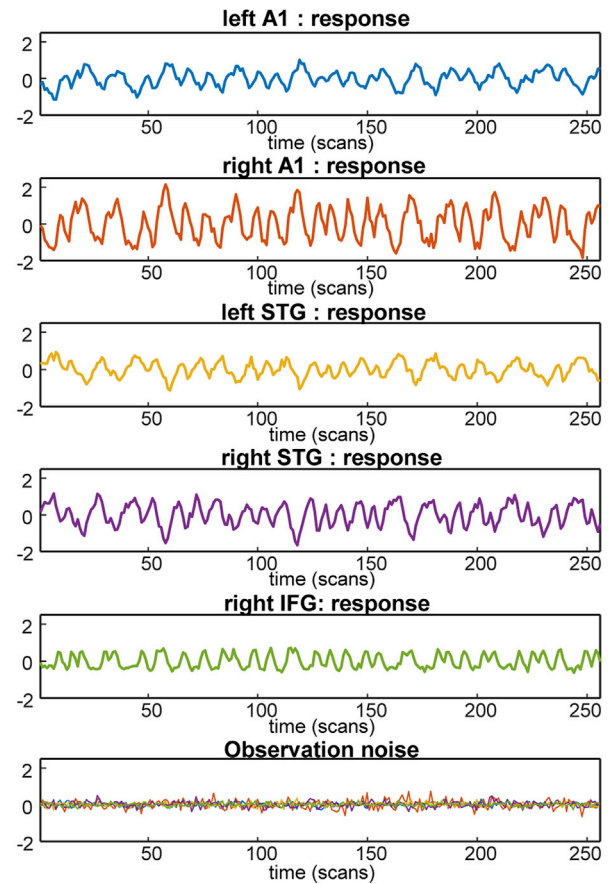
(A) Synthetic EEG

(B) Synthetic fMRI

Fig. 3. The simulated multimodal data. (A) The synthetic EEG data. The upper panel shows the synthetic EEG sensor data (127 channels, with observation noise) for the standard and oddball conditions respectively. The lower panel shows the first principal components of the sensor data for the two conditions and their difference (i.e., mismatch negativity), and the observation noise added to the sensor data. (B) The synthetic fMRI data. The simulated BOLD response (with observation noise) of each of the five regions is shown in this panel.

respiratory artefacts. We finally took a section of resulting timeseries – with length of 256 scans – from the noisy BOLD responses to be the synthetic fMRI data used in subsequent analyses. The ensuing synthetic fMRI timeseries are shown in Fig. 3B.

2.4. Model inversion schemes: Bayesian fusion vs. fMRI only inversion

In this section, we focused on the comparison of fMRI inversion with and without Bayesian fusion. We first performed Bayesian fusion using both synthetic EEG and fMRI data. Our procedure comprised the following two steps:

1. Specification of priors for EEG: we specified the connectivity model (the connectivity architecture shown in Fig. 2B), and the uninformative priors over neuronal parameters according to Table 1. In addition, we placed precise priors over the EEG spatial (lead field) parameters, based upon the spatial parameters (see Fig. 2B) used to generate the synthetic EEG data. This ensured the neuronal posteriors from the EEG analysis were not confounded by conditional dependencies between EEG spatial and neuronal parameter estimates (see Kiebel et al. (2006) for an evaluation of the effects of lead field specification on the recovery of model parameter estimates).
   Model inversion of EEG: we performed a DCM analysis for EEG using the above priors, to obtain the posterior means and covariances for neuronal parameters (using the SPM routine **spm_nlsi_N**.m – and ensuing posterior moments: Ep and Cp). This inversion uses a mean

field approximation to provide marginal (approximate) posterior densities over spatial and neuronal parameters. The neuronal posteriors were retained for subsequent Bayesian belief updating.

2. Specification of priors for fMRI: we specified the connectivity model (Fig. 2B), and the uninformative priors over haemodynamic parameters (Table 2), but replaced the uninformative neuronal priors (Table 1) with the posterior distribution over neuronal parameters (i.e., Ep and Cp) from the above EEG inversion (i.e., **Bayesian belief updating**).
   Model inversion of fMRI: we conducted a DCM analysis for fMRI using the above priors, to obtain the posterior beliefs over all free parameters informed by both EEG and fMRI data (inverted using **spm_nlsi_GN**.m).

We also performed fMRI only inversion for comparison purposes. The only difference between the two fMRI inversions concerned the priors over neuronal parameters: the fMRI only inversion retained the uninformative neuronal priors (Table 1), as opposed to the empirical neuronal priors derived from the EEG analysis. Using different priors causes the variational Laplace scheme to start from different initialisation points (i.e., from prior expectations) during parameter estimation. To ensure the results of the two inversion schemes would not be explained by different initialisations, we used the starting points of Bayesian fusion to initialise the fMRI only inversion. In summary, the fMRI only inversion was as follows:

Specification of priors for fMRI: we specified the connectivity model (the connectivity architecture shown in Fig. 2B), and the uninformative priors over neuronal and haemodynamic parameters according to Tables 1 and 2, respectively.

Model inversion of fMRI: we conducted a DCM analysis for fMRI using the above priors and the initialisation point of Bayesian fusion (by setting **DCM.options.P** in **spm_dcm_fmri_nmm**.m), to obtain the posterior beliefs over all free parameters (inverted using **spm_nlsi_GN**.m).

After implementing the two inversion schemes, we evaluated the quality of model inversions and the efficiency of parameter estimation from the following three perspectives: (i) we conducted Bayesian model comparison based on the (negative) free energy (as a lower bound approximation to log model evidence), (ii) checked the consistency between the predicted and true (synthetic) BOLD responses, (iii) compared the reduction in posterior variances of the parameter estimates, which scores the shrinkage of uncertainty. The free energy and the parameter estimation results are shown in Figs. 5 and 6.

2.5. Quantification of information gain

To further quantify the relative utility of EEG and fMRI in making inferences about various model parameters – and to evaluate any increase in the efficiency of parameter estimation attributable to Bayesian fusion – we calculated the information gain. This measure of data quality is the Kullback-Leibler (KL) divergence between prior and posterior beliefs, after observing one or more data modalities. Under the Laplace assumption employed in DCM; i.e., the prior and posterior probabilities are Gaussians, we used the KL divergence between multivariate Gaussians:

$$D(N_1 \parallel N_0) = \frac{1}{2}\left(\text{tr}\left(\Sigma_0^{-1}\Sigma_1\right) + (\mu_0 - \mu_1)^{\text{T}}\Sigma_0^{-1}(\mu_0 - \mu_1) - k + \ln\frac{\det\ \Sigma_0}{\det\ \Sigma_1}\right)$$

Where $N_0 = \text{N}(\mu_0, \Sigma_0)$ denotes the prior density over parameters with mean μ_0 and covariance Σ_0, $N_1 = \text{N}(\mu_1, \Sigma_1)$ denotes the corresponding posterior density, and $k = \text{rank}(\Sigma_0)$ indicates the number of free parameters. The divergence $D(N_1 \parallel N_0)$ measures the difference (in units of nats) from the prior density N_0 to posterior probability density N_1, which in the context of Bayesian inference reflects the information gained when we revise our beliefs about parameters (by observing data). In other words, it scores the degree of belief updating afforded by any given data (Kullback and Leibler, 1951; Burnham and Anderson, 2003; Duchi, 2007; Zeidman et al., 2018).

Therefore, based on the original priors ($p(\theta|m)$), the unimodal posteriors from EEG ($p(\theta|Y_{EEG})$), the unimodal posteriors from fMRI ($p(\theta|Y_{MRI})$), and the Bayesian fusion posteriors ($p(\theta|Y_{MRI}, Y_{EEG})$), one can quantify the information gain pertaining to neuronal and haemodynamic parameters $\theta = [\theta_N, \theta_H]$. Given that these prior and posterior distributions are joint multivariate Gaussians, the marginal distribution of neuronal and haemodynamic parameters can be obtained easily from the mean and the covariance of the joint distribution. Based on the marginalised prior and posterior distributions, we can then quantify the information gain with respect to the neuronal and haemodynamic parameters separately. In more detail, we evaluated the following information gain:

1 $D[p(\theta|Y_{MRI}) \parallel p(\theta|m)]$, $D[p(\theta_N|Y_{MRI}) \parallel p(\theta_N|m)]$, and $D[p(\theta_H|Y_{MRI}) \parallel p(\theta_H|m)]$ – The KL divergence from original priors to posteriors (joint, marginalised neuronal, and marginalised haemodynamic, respectively) given only fMRI data, to assess to what extent the fMRI data could inform parameter estimation. For convenience, we denoted this set of KL divergences as **D1**, **D1$_N$** and **D1$_H$**.

2 $D[p(\theta|Y_{EEG}) \parallel p(\theta|m)]$, $D[p(\theta_N|Y_{EEG}) \parallel p(\theta_N|m)]$, and $D[p(\theta_H|Y_{EEG}) \parallel p(\theta_H|m)]$ – The KL divergence from original priors to posteriors given

only EEG data, to quantify the amount of information that EEG data contributed in Bayesian fusion analysis. We denoted this group of KL divergences as **D2**, **D2$_N$** and **D2$_H$**.

3 $D[p(\theta|Y_{MRI}, Y_{EEG}) \parallel p(\theta|Y_{EEG})]$, $D[p(\theta_N|Y_{MRI}, Y_{EEG}) \parallel p(\theta_N|Y_{EEG})]$, and $D[p(\theta_H|Y_{MRI}, Y_{EEG}) \parallel p(\theta_H|Y_{EEG})]$ – The KL divergence from posteriors given only EEG data to posteriors given both modalities, to quantify the amount of information that the subsequent fMRI data could provide after observing EEG data. We denoted this group of KL divergences as **D3**, **D3$_N$** and **D3$_H$**.

4 $D[p(\theta|Y_{MRI}, Y_{EEG}) \parallel p(\theta|m)]$, $D[p(\theta_N|Y_{MRI}, Y_{EEG}) \parallel p(\theta_N|m)]$, and $D[p(\theta_H|Y_{MRI}, Y_{EEG}) \parallel p(\theta_H|m)]$ – The KL divergence from original priors to posteriors given both EEG and fMRI data, to quantify the total information gain afforded by fusing both modalities. We denoted this set of KL divergences as **D4**, **D4$_N$** and **D4$_H$**.

We used these information gain expressions to quantify the amount of information contained in fMRI, EEG and their fusion. One expects **D4** to be greater than **D1**, reflecting the fact that the fusion of both modalities should be informative than a single modality. Moreover, based on our previous assumption that EEG would be more informative than fMRI – with respect to neuronal parameters – we would expect **D2$_N$** to be greater than **D1$_N$**. If Bayesian fusion accounts for conditional dependencies between neuronal and haemodynamic parameters, a better estimate of haemodynamic parameters would be obtained – as reflected here by an expected greater value of **D3$_H$** than **D1$_H$**. We report the information gain using bar charts, as shown in Fig. 7.

3. Results

3.1. The simulated multimodal data

The simulated multimodal data are shown in Fig. 3, which reproduce plausible responses to the mismatch negativity paradigm in EEG research and the auditory oddball design in fMRI. The upper panel of Fig. 3A shows the synthetic EEG sensor data (127 channels, with observation noise) for the standard and oddball conditions respectively. The lower panel shows the first principal component of the sensor data for the two conditions and their difference, as well as the observation noise added to the sensor responses. The accompanying simulated BOLD response (with observation noise) of each of the five regions is respectively shown in Fig. 3B. These synthetic fMRI timeseries reflect a succession of frequent standard stimuli and occasional oddball stimuli and show that different regions exhibit functionally selective responses to oddball events.

To provide insight into the (predicted) neuronal responses measured with fMRI, we show an example in Fig. 4 of the simulated pre-synaptic inputs and the neurovascular signal of one brain region (right A1). The upper panel of Fig. 4 shows the three sorts of pre-synaptic inputs (i.e., intrinsic inhibitory afferents, intrinsic excitatory afferents and extrinsic afferents) to each of the four populations, where the 12 pre-synaptic inputs exhibit different amplitudes (some of them are zero indicating no such afferent) but similar patterns. The consequent neurovascular signal – shown in the lower panel – is a linear mixture (scaled by the 12 neurovascular coupling parameters) of these pre-synaptic inputs. The signal is then magnified to show how experimental effects (i.e., driving and modulatory inputs) are expressed at the synaptic level: here, we see that every event onset triggers a neuronal response, while the oddball events cause much stronger fluctuations in comparison with standard events. The final BOLD signal is generated by a haemodynamic convolution of the neurovascular signal. The resulting BOLD responses correspond to oddball events, because responses to standard events are smoothed over time by the haemodynamics. This also suggests that the neuronal parameters that can be estimated relatively more efficiently, using fMRI, are the changes in connectivity that generate oddball responses.
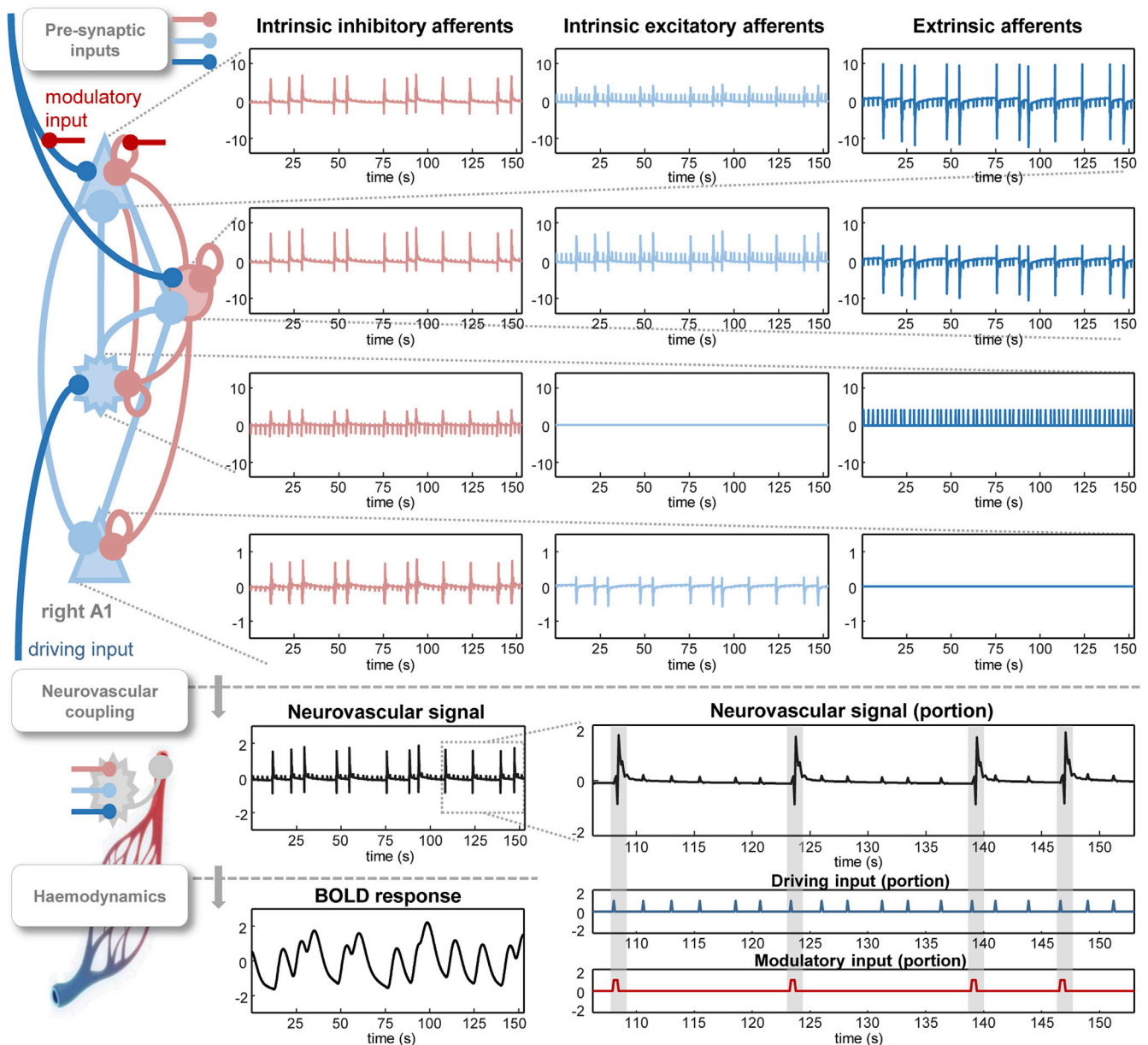
Fig. 4. The predicted neuronal responses and neurovascular signal (of one brain region: right A1) when modelling fMRI with the canonical microcircuit DCM. Each of the four populations receives three sorts of pre-synaptic inputs, corresponding to intrinsic inhibitory afferents (pink), intrinsic excitatory afferents (light blue), and extrinsic afferents (dark blue). The subsequent neurovascular signal is the linear mixture (scaled by the neurovascular coupling parameters) of these pre-synaptic inputs. This signal is then magnified to show in detail about how experimental effects (i.e., driving and modulatory inputs) are expressed at the synaptic level. The final BOLD signal is generated by the haemodynamic convolution of the neurovascular signal, where the BOLD response fluctuations correspond to oddball events.
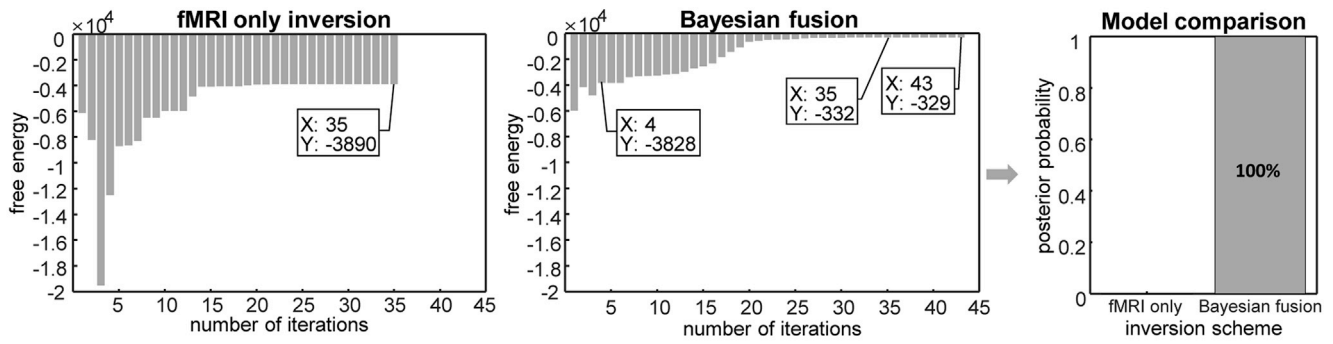
3.2. Model inversion results

Fig. 5 compares the free energy and the consistence between real and predicted BOLD responses under the two inversion schemes; namely, the fMRI only inversion and Bayesian fusion. In Fig. 5A one can see that, in general, the free energy (i.e., log model evidence) increases with increasing number of iterations in both inversion schemes, endorsing the variational Laplace scheme used to estimate model parameters and evidence. In the fMRI only inversion, the scheme converges after 35 iterations (to a free energy of −3890 nats); while in Bayesian fusion, the free energy reaches a plateau after 27 iterations, and converges at the 43rd iteration (to a much larger free energy of −329 nats). The subsequent Bayesian model comparison indicates, as expected, that Bayesian fusion is better than the fMRI only inversion with posterior probability of 100%. Fig. 5B shows the real and predicted BOLD responses under the two inversion schemes; in which the real BOLD signals correspond to the grey dotted lines, and predicted BOLD signals are denoted by pink solid lines. These results show that, with Bayesian fusion, the predicted BOLD signals of all the five regions match the real signals remarkably well, but for the fMRI only inversion, we see obvious inconsistencies between the predicted and real BOLD in the right A1, right STG and right IFG, indicating a relatively poor parameter estimates.

Fig. 6 shows the prior and posterior beliefs about the parameters (for both neuronal and haemodynamic parameters) as well as the conditional correlations among these parameters; for the fMRI only inversion (Fig. 6A) and Bayesian fusion (Fig. 6B). We first compare the prior beliefs and correlation matrices of the two inversion schemes. For fMRI only inversion, the default priors for both neuronal and haemodynamic parameters are uninformative (zero means and large variances) and independent of each other (i.e., a diagonal form for the prior correlation

(A) Free energy comparison
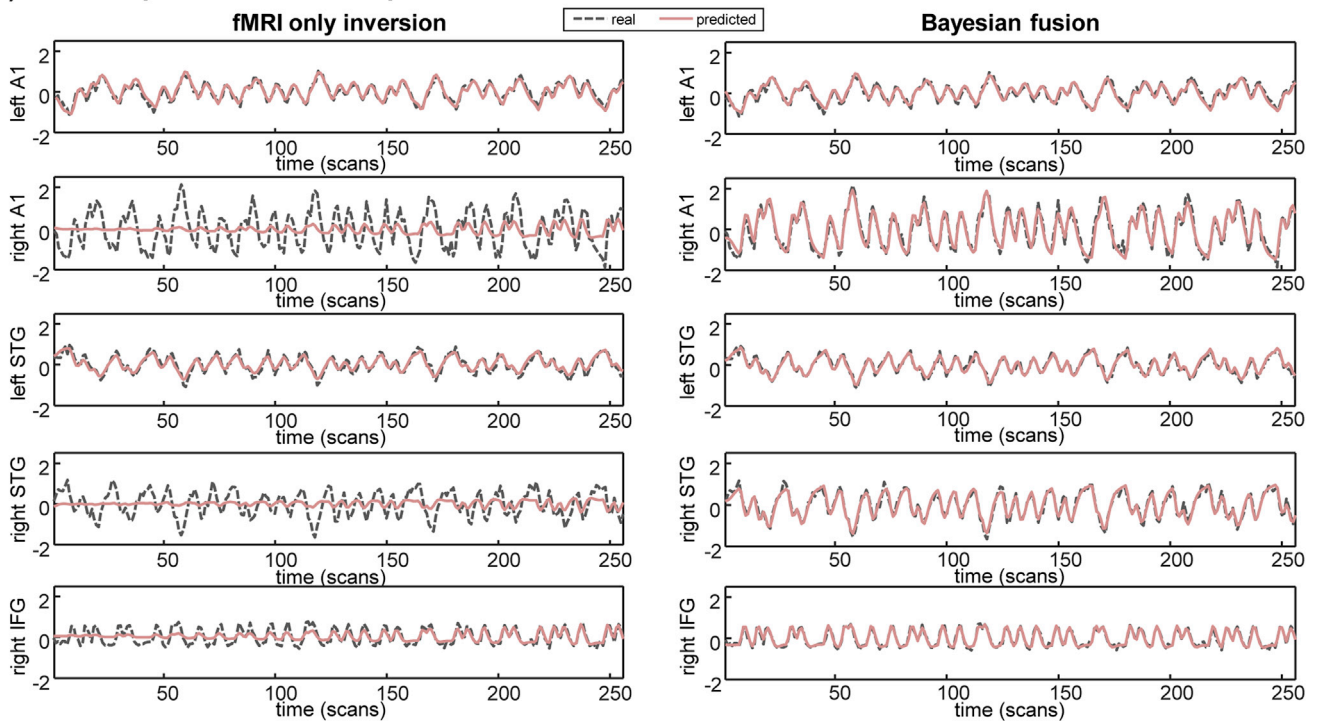

(B) Real and predicted BOLD responses

Fig. 5. Comparison of model inversion results between fMRI only inversion and Bayesian fusion. (A) The free energy changes over iterations of a variational Laplace scheme, and the ensuing Bayesian model comparison of the two inversion schemes. (B) The real (grey dotted lines) and predicted (pink solid lines) BOLD responses correspond to the two inversion schemes.

matrix). After Bayesian belief updating, the EEG modality moves the neuronal priors away from a mean of zero, and reduces the prior variance of most neuronal parameters; these reduced values reflect a reduction in uncertainty about neuronal parameters after observing EEG data. In addition, the (empirical) prior correlation matrix over the neuronal parameters is no longer diagonal, reflecting conditional dependencies given informative electromagnetic data.

The subsequent posterior beliefs and posterior correlation matrices for fMRI only inversion and Bayesian fusion, respectively, reveal the informativeness of fMRI data in estimating neuronal and haemodynamic parameters before and after seeing the EEG modality. Here, we are particularly interested in the efficiency with which haemodynamic parameters could be estimated after observing EEG data. Clearly, we obtain no direct information about haemodynamic parameters following inversion of EEG data: this is reflected in the fact that the prior beliefs over haemodynamic parameters are the same under the two inversion schemes. However, it is possible that EEG data can provide additional information about haemodynamic parameters, in virtue of conditional dependencies. In other words, EEG data can resolve uncertainty about

neuronal parameters – after seeing fMRI data – if, and only if, neuronal and haemodynamic parameter estimates depend on each other. Indeed, by inspecting the posterior correlation matrices in both Fig. 6B and A, one can see clear conditional correlations between neuronal and haemodynamic parameters. This means that reducing uncertainty about neuronal parameters will affect uncertainty about haemodynamic parameters – allowing for a synergy between EEG and fMRI data during model inversion. We now quantify this kind of synergetic interaction using the following information gain results.

3.3. Information gain

Fig. 7 shows the information gain results. The upper panel shows the amount of information we have obtained by conducting the fMRI only inversion. The lower panel illustrates the information accumulation from default priors to EEG posteriors (i.e., Bayesian belief updating), then from EEG posteriors to bimodal posteriors in the Bayesian fusion of EEG and fMRI; and the information gain from default priors to bimodal posteriors, reflecting the total information gain afforded by Bayesian fusion. Overall,
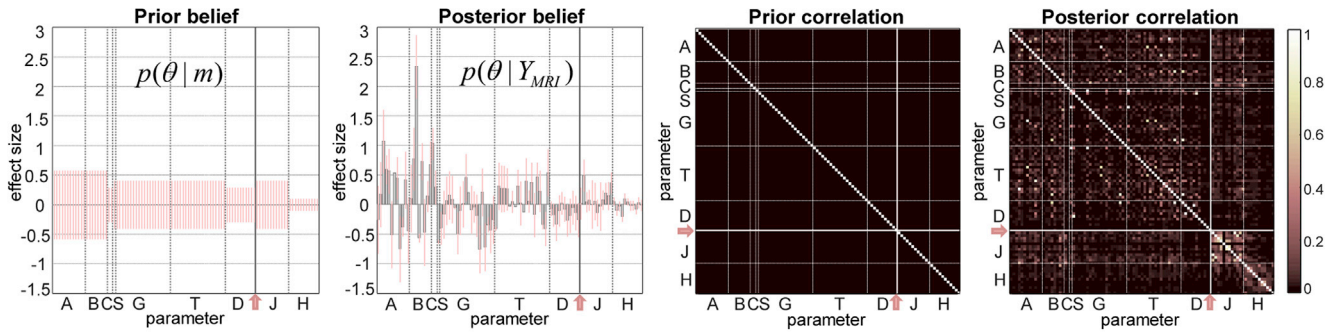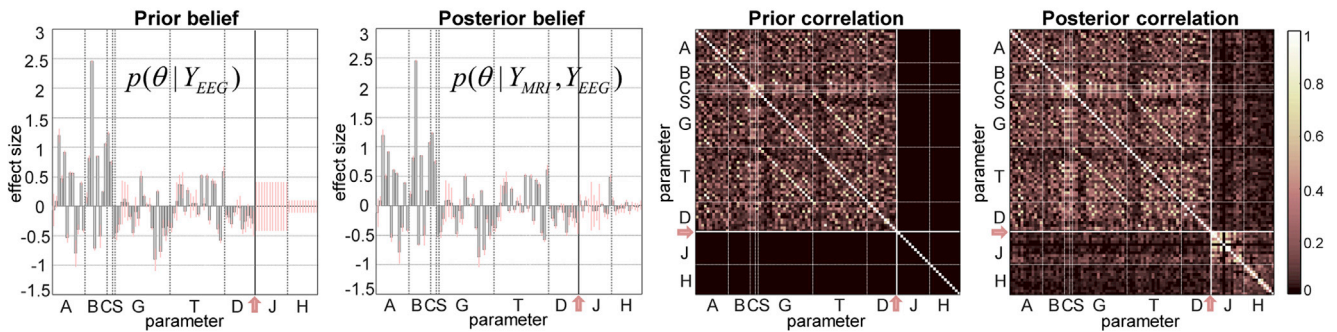
(A) fMRI only inversion

(B) Bayesian fusion

Fig. 6. The prior and posterior beliefs of the model parameters (see Fig. 2, Tables 1 and 2 for a detailed description of each model parameter), as well as the conditional correlations among these parameters, corresponding to (A) fMRI only inversion and (B) Bayesian fusion of EEG and fMRI.

as expected, Bayesian fusion reveals more information than just inverting fMRI data (i.e. $\mathbf{D4} > \mathbf{D1}$), and this increase in information gain is reflected in both neuronal and haemodynamic decompositions. More specifically, in terms of the neuronal parameters, we obtain greater value of $\mathbf{D2_N}$ than $\mathbf{D1_N}$ suggesting that EEG is more informative than fMRI. This endorses our understanding of the results in Fig. 6; namely, that EEG data contains much more information about neuronal parameters. For the haemodynamic parameters, we obtain no information following inversion of EEG data (i.e., $\mathbf{D2_H} = 0$). However, using the empirical priors from EEG, the fMRI data provides more information about the parameters that mediate neurovascular coupling, relative to just observing fMRI (i.e., $\mathbf{D3_H} > \mathbf{D1_H}$). This increase in information gain furnishes quantitative evidence that Bayesian fusion of EEG and fMRI can resolve conditional dependencies between neuronal and haemodynamic parameters; thereby improving the efficiency of parameter estimation based on non-invasive fMRI data and neural mass models of the canonical microcircuit.

4. Discussion

In this paper, we evaluated the contribution of multimodal EEG and fMRI data for estimating neuronal architectures using the canonical microcircuit DCM. In summary: 1) we created biologically plausible (based on empirically determined parameters) simulations of multimodal neuroimaging data (i.e., EEG and BOLD responses, see Fig. 3) under the same paradigm (i.e., the mismatch negativity paradigm, see Fig. 2). Crucially, these multimodal data were generated using a common neuronal architecture (Fig. 1). 2) We simulated neuronal responses (i.e., pre-synaptic inputs) and ensuing neurovascular signal, which encoded how experimental effects (i.e., driving and modulatory inputs) were probably expressed at the synaptic level (see Fig. 4). 3) Comparative analyses showed that Bayesian fusion furnished better model parameter estimation, which was reflected by increases in free energy (i.e., log model evidence), a better match between real and predicted BOLD signals (see Fig. 5), and a reduction of posterior variance, indicating a

shrinkage of uncertainty about model parameters (see Fig. 6). 4) Information gain provided quantitative evidence that Bayesian fusion could leverage conditional dependencies between neuronal and haemodynamic parameters; thereby evincing a synergetic resolution of uncertainty about model parameters (see Fig. 7).

This paper is the first formal demonstration that Bayesian fusion of EEG and fMRI supports inferences about detailed changes in microcircuitry. Based on this modelling framework and Bayesian fusion, one could propose and test many hypotheses pertaining to cortical hierarchies, such as exploring microcircuit models of specific regions or hierarchical networks, with distinct forward and backward extrinsic connections. One could also evaluate experimental effects (e.g., attention, visual perception, pharmacologic, etc.) on changes in either extrinsic (long-range) or intrinsic (short-range) connections at the level of specific neuronal populations: e.g., superficial or deep pyramidal cells (Heinzle et al., 2007; Brown and Friston, 2012; Auksztulewicz and Friston, 2015; Tsvetanov et al., 2016; Friston et al., 2017; Havlicek et al., 2017; Lawrence et al., 2018). These sorts of questions shift our focus away from simply localising regional responses (using fMRI) towards an ever more detailed characterisation of functional integration and physiologically grounded research that exploits the increasing temporal (M/EEG) and spatial (fMRI) fidelity of neuroimaging data.

4.1. Technical considerations

The multimodal Bayesian fusion proposed in this paper – particularly in reference to using posterior estimates from DCM analysis of EEG as empirical priors for subsequent DCM analysis of fMRI – rests upon a common neuronal model; namely, the canonical microcircuit neural mass model. This neuronal model was developed for DCM for M/EEG then applied to DCM for fMRI (Brown and Friston, 2012; Bhatt et al., 2016; Friston et al., 2017; Pinotsis et al., 2017). The thought of adopting the same neuronal model to explain both data modalities was discussed in the foundational paper describing DCM for M/EEG (David et al.,
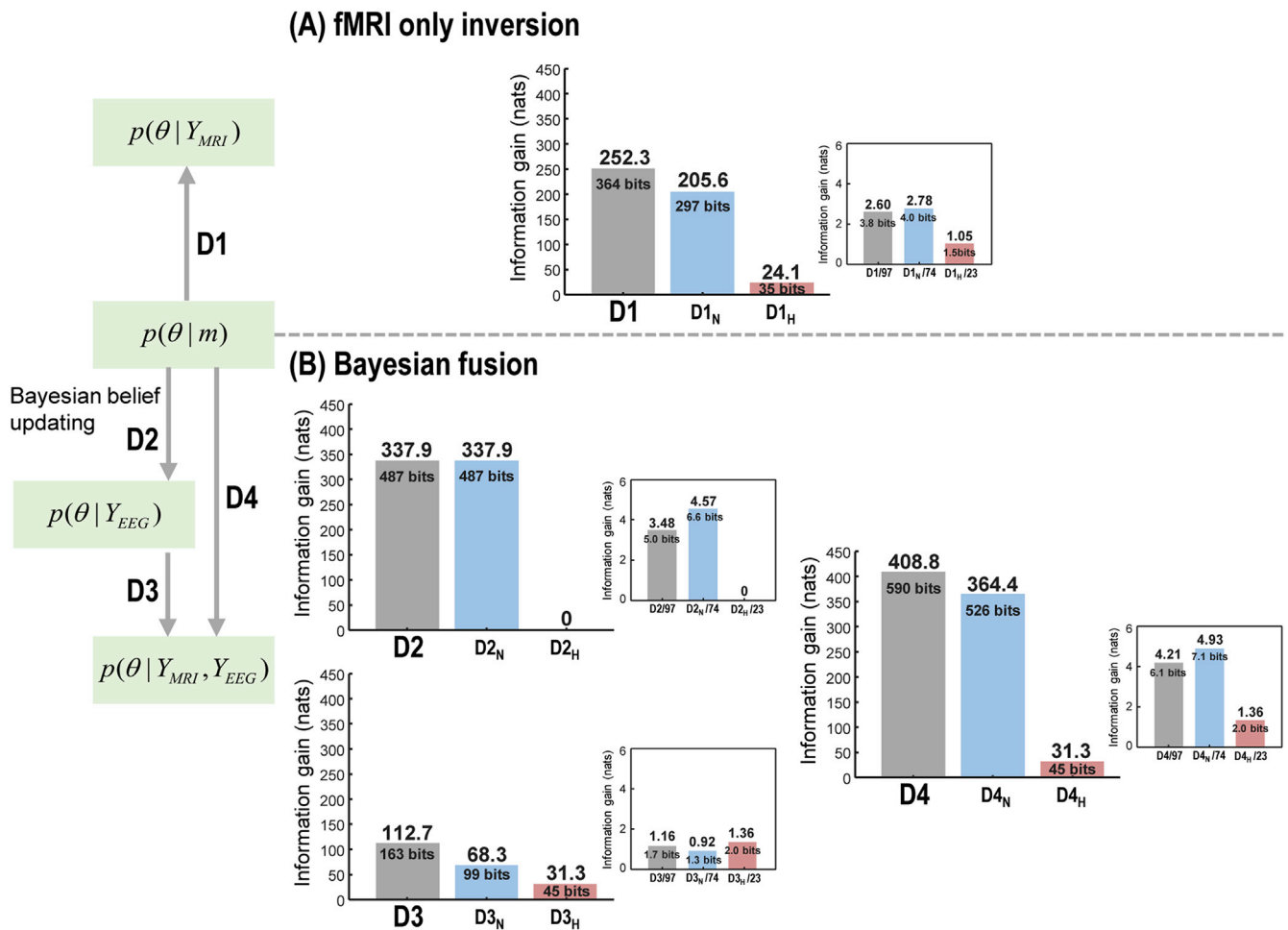
Fig. 7. Information gain (in units of nats or bits, 1 nat $=$ $\log_2(e)$ bits) associated with fMRI and EEG data separately – and their fusion. (A) The KL divergences from default priors to (joint, marginalised neuronal, marginalised haemodynamic) posteriors, given only fMRI data (i.e., D1, D1$_N$, and D1$_H$, respectively). This set of KL divergences speak to the information gain obtained in fMRI only inversion. (B) The KL divergences from default priors to empirical priors, given only EEG data (i.e., D2, D2$_N$, and D2$_H$, respectively), then from empirical priors given EEG to posteriors, given both EEG and fMRI (i.e., D3, D3$_N$, and D3$_H$, respectively), and a total measure, from default priors to posteriors, given both EEG and fMRI (i.e., D4, D4$_N$, and D4$_H$, respectively). These KL divergences quantify the information gain under multimodal Bayesian fusion. The KL divergences are then divided by the number of parameters (i.e., 97 model parameters, including 74 neuronal parameters and 23 haemodynamic parameters), to measure the information gain afforded by each parameter.

2006), in which the authors pointed out that the most significant challenge of employing over-parameterised neural mass models for fMRI data was guaranteeing an efficient model inversion. In DCM, model inversion generally uses standard variational Laplace procedures to estimate model parameters and evidence for inferences about specific connections and network structure; where the model evidence corresponds to accuracy minus complexity (Penny et al., 2004; Friston et al., 2007). In other words, a model with too many free parameters is a model whose complexity cost exceeds the increase in accuracy or goodness of fit afforded by extra parameters. In this study, we used two model inversion schemes; namely, the Bayesian fusion of EEG and fMRI, and the fMRI only inversion, and illustrated clearly that the fMRI only inversion indeed encountered the model inversion inefficiency in relation to Bayesian fusion. Some 'flat line' predictions of BOLD responses in fMRI only inversion, as well as the convergence of free energy to a much lower value compared with that in Bayesian fusion (see Fig. 5), suggest that in some cases the parameter estimation of canonical microcircuit DCM is inefficient given merely fMRI data and without the empirical priors afforded by EEG. Interestingly, the complexity part of (log) evidence is exactly the information gain used to assess the contribution from different data modalities above. Heuristically, informative data 'pulls' the posterior density – over unknown parameters – away from the prior density to provide an accurate account of the data (i.e., maximise model

fit). In other words, the accuracy 'pays for' a complexity cost, which is the information gain. Model evidence is therefore the difference between the accuracy and complexity or information gain.

The information gain afforded by different modalities have a useful quantitative interpretation. They are measured in natural units (nats) that can be converted into bits by multiplying with $\log_2(e)$. One bit of information allows us to say whether a particular parameter is high or low. For example, the information gain about the neuronal parameters, given the EEG data (i.e., **D2$_N$**) is about 487 bits. If we distribute this information over the 74 neuronal parameters, this suggests that we gain about six bits of information about each parameter. In other words, we could place the strength of each connection or synaptic rate constant in one of six 'bins'. The information gain – afforded by Bayesian fusion – about the haemodynamic parameters (i.e., **D3$_H$**) is about 31 nats or 45 bits. Given we used 23 haemodynamic parameters; this corresponds to about two bits per parameter (see Fig. 7).

Some readers may wonder why inverting the canonical microcircuit DCM using just fMRI (without any information from EEG) was possible in Friston et al. (2017)? In that paper, the authors fixed some neuronal parameters (including the synaptic rate constants, self-inhibitions, and conduction delays) to their prior mean when conducting model inversion. This contrasts with our fMRI only inversion, where we inverted the synthetic fMRI data without fixing any neuronal parameters – by doing

this we leveraged the information gain afforded by Bayesian belief updating. In Friston et al. (2017), the neuronal parameters were fixed to deflate the model complexity to ensure more efficient balance between accuracy and complexity. This illustrates the fact that priors can have an important role in determining the efficiency of model inversion (Kiebel et al., 2006); especially in the context of canonical microcircuit DCM for fMRI. These suggest that the main limitation of the current model is an over-parameterisation for modelling fMRI data – speaking to further refinement of the current generative model (for example, using appropriate empirical data to refine parameterisation via Bayesian model reduction). However, this inefficiency emphasises the necessity of conducting Bayesian fusion of EEG and fMRI, to doubly inform the parameter estimation. Having said these, in practice, canonical microcircuit DCM analyses with fMRI therefore require a careful consideration of model complexity – and the provision of precise constraints on model parameters; such as the empirical priors afforded by Bayesian fusion.

The Bayesian fusion approach adopted in this paper rests on the fact that "both EEG and BOLD measurements are observable consequences of the same (hidden) neuronal activity" (Friston et al., 2017). At first glance, using a generative model that can generate predictions in two modalities speaks to a 'symmetric' approach – in which both modalities are inverted at the same time (Daunizeau et al., 2007; Valdes-Sosa et al., 2009; Rosa et al., 2010; Huster et al., 2012; Lei et al., 2012). However, we have proposed a Bayesian belief updating scheme, in which the posteriors following inversion of the electromagnetic modality are used as priors for the fMRI inversion. So does the implicit asymmetry of belief updating come at any cost? In principle, under exact Bayesian inference, the answer is no. In other words, if both the data modalities informed the same set of parameters, then the order of belief updating should not matter. However, in practice, the order matters in virtue of using approximate Bayesian inference and nonlinear observation or generative models. For example, had we tried to estimate the neuronal parameters from the fMRI data, the conditional dependencies between neuronal and haemodynamic parameters might have confounded the inversion of the EEG data, by placing empirical priors over local minima. Unfortunately, one cannot use Bayesian model comparison to detect local minima, because a failure to converge to the global optimum precludes the use of variational free energy as an approximation to model evidence. Having said this, the pipeline described in this paper is sufficiently robust to violations of the Laplace assumption for our purposes; namely, to quantify the information gain afforded by Bayesian fusion.

4.2. Potential applications

This paper has focused on the technical procedures entailed by this kind of Bayesian fusion; however, potential domains of application deserve some comments. In one sense, people already use Bayesian fusion when they use (empirical) priors from the location of fMRI activations for source reconstruction in EEG. This is particularly the case in dynamic causal modelling of EEG data where the prior location of equivalent current dipoles (ECDs), modelling each electromagnetic source, requires location (i.e., spatial) priors – that are often based on fMRI (Henson et al., 2010; Huster et al., 2012). The Bayesian fusion described in this paper completes the synergetic use of both modalities by enabling the posterior estimates from EEG – that inherit spatial information from fMRI – to inform the modelling of fMRI – so that it inherits temporal information from EEG. One of the key benefits of this is that one can resolve conditional dependencies between neuronal and haemodynamic parameters. This is an important observation because it speaks to key questions about neurovascular coupling and attributing differences in evoked or induced responses to changes in neuronal connectivity or haemodynamics. This allows one to investigate, for example, the physiology of aging using multimodal DCM and Bayesian fusion. These include changes in neurovascular coupling strengths, stiffness of vessels (Grubb's exponent in the haemodynamic model) due to aging that cannot be

measured using physiological recordings (D'Esposito et al., 2003). In addition to questions about pathophysiology, other issues pertaining to haemodynamics and biophysics could also be addressed using Bayesian model comparison. For example, haemodynamic models usually consider changes in cerebral blood volume in venous compartments (as in this paper). Based on optical imaging, CBV changes in the arterial side may contribute to the BOLD signal. Although this contribution constitutes less than 33% of capillaries and venous contributions, the contribution of arterial components to the BOLD signal can, in principle, be assessed with models that do and do not include an arterial component. One can also ask, is the effect of aging on fMRI responses attributable to changes in neuronal architectures, changes in the elasticity of blood vessels – or both? e.g., Fabiani et al. (2014). In principle, the scheme described in this paper should provide an optimal estimate of changes in neuronal and haemodynamic parameters to address this sort of question.

Another potential application is to incorporate depth-dependent haemodynamic models within the current DCM framework, which requires high-resolution fMRI to resolve cortical layers. Please note in this paper, we are not concerned with laminar fMRI and therefore model the measurable fMRI as a mixture of contributions from distinct cortical layers. The basic principle of multimodal DCM – that both EEG and fMRI measurements are observable consequences of the same (hidden) neuronal activity – enables the laminar designation of canonical microcircuitry to be inherited from the EEG and MEG modelling literature. In this setting, laminar specificity is written into the model, which has been continuously refined over the past years (Brown and Friston, 2012; Dima et al., 2012; Pinotsis et al., 2014, 2017; Moran et al., 2015; Bhatt et al., 2016; Friston et al., 2017; Shaw et al., 2017; Symmonds et al., 2018). The particular (canonical microcircuit) model used here was based upon the statistics of inter- and intralaminar connectivity from empirical neurophysiology and microcircuit studies; e.g., (Douglas and Martin, 1991; Haeusler and Maass, 2007; Sohal et al., 2009; Pi et al., 2013). A discussion of this model in relation to the computational architectures implied by predictive coding (Bastos et al., 2012; Shipp, 2016; Keller and Mrsic-Flogel, 2018) – and empirical ECoG data – can be found in (Auksztulewicz and Friston, 2015; Bastos et al., 2015). Clearly, different mappings between predictive coding and the physiological/anatomical data on canonical microcircuits are conceivable; e.g., (Markov et al., 2013; Shipp, 2016). In addition, it is worth mentioning that some patterns of connectivity in the current canonical microcircuit model are introduced to ensure mathematical efficiency (for more detail, please see Bastos et al. (2015)), which means that some connections would be loosely related to anatomy. However, the canonical microcircuit DCM applied here has been validated, and is sufficient to enable the illustration of the Bayesian fusion procedure for our purposes.

In short, although the canonical microcircuit model is now the standard model for most DCM studies, there are several variants – that speak to the ability of different data to support efficient model comparison – or the hypotheses at hand (Moran et al., 2015; Muthukumaraswamy et al., 2015; Symmonds et al., 2018; Shaw et al., 2019). Therefore, there is no commitment to the particular canonical microcircuit used in this paper, other microcircuit models could have been used within the current framework. As noted above, it will be an interesting endeavour to extend the current formalism to deal with high field fMRI that can resolve different layers. This may call for more refined haemodynamic models; e.g. (Havlicek et al., 2015; Heinzle et al., 2016), that can be optimised using Bayesian model comparison. However, potential modelling challenges include the difficulty of modelling a full six-layer haemodynamics given the present spatial resolution of non-invasive fMRI, and regional differences in cortical thickness. In short, the mechanisms and modelling of depth-dependent haemodynamics remain an outstanding challenge that may require spatiotemporal models, with explicit laminar depth parameters: e.g., (Kiebel and Friston, 2002; Drysdale et al., 2010; Lawrence et al., 2018; Self et al., 2019).

4.3. Limitations and future directions

As with any Bayesian modelling, one should interpret any (model) evidence-based conclusions in light of the models considered. Formally, this means that the models chosen for comparison have been selected *a priori*. The standard canonical microcircuit and haemodynamic models we used were chosen to showcase the methodology – not to make any assertions that would necessarily generalise to other kinds of data or models. In this technical note, we have chosen to illustrate multimodal fusion using the original formulation of haemodynamics implicit in the balloon model. This choice was motivated by simplicity and consistency with previous technical descriptions of DCM. However, it does not imply a commitment to the balloon model. It is worth mentioning that the key idea behind the balloon model was that an increase in arteriolar blood flow leads to an increase in blood volume in the post-capillary vessels. In this model, the post stimulus undershoot is explained by a delayed return of the post-capillary volume to baseline, with a time constant of seconds (i.e., delayed vascular compliance). This could be one explanation of the post stimulus undershoot. However, plausible alternatives are that there is an undershoot in cerebral blood flow (Chen and Pike, 2009), or that there is an elevated oxygen metabolism with cerebral blood flow and cerebral blood volume at baseline (Frahm et al., 1996; van Zijl et al., 2012). This speaks to the biophysical interpretability of the parameters of the balloon model – and the opportunity to repeat the kind of analysis described above with more recent formulations of the balloon model; e.g., Havlicek et al. (2017). In short, alternative proposals could be evaluated through Bayesian model comparison in the usual way. As more refined empirical priors based upon advances in quantification of intrinsic and extrinsic circuitry is gathered, and as the biophysical mechanisms generating the BOLD signal become increasingly elucidated, this knowledge could be incorporated into future DCMs.

Several lines of research suggest themselves for future work. First, it will be useful to establish which applications would benefit from inverting a single generative model of EEG and fMRI data, as presented here, and which applications could be addressed by fitting separate neuronal and haemodynamic models to each modality, for example, using neuronal responses from a DCM analysis of EEG to drive a haemodynamic model of fMRI responses (Jafarian et al., 2019). In principle, this question can be addressed using Bayesian model comparison. Second, the utility of the canonical microcircuit DCM and multimodal Bayesian fusion as a clinical computational assay or phenotyping to characterise pharmacological, pathophysiological and cognitive dysfunctions using patient data (Boly et al., 2011; Campo et al., 2012; Uludag and Roebroeck, 2014; Benetti et al., 2015; Mechelli et al., 2017). Additionally, when dealing with multi-subject and multi-model Bayesian fusion, the parametric empirical Bayes (PEB) procedures may be useful (Henson et al., 2011; Friston et al., 2015, 2016; Litvak et al., 2015; Wakeman and Henson, 2015). Third, the refinement of the current generative model, especially optimising the parameterisation of neurovascular coupling, may benefit from carefully selected empirical data; while using Bayesian fusion to resolve conditional dependencies between neuronal and haemodynamic parameters. Finally, the expansion of the current task-based analysis to the corresponding resting-state methodology (Friston et al., 2014; Razi et al., 2017), where an equivalent canonical microcircuit formulation for cross spectral data features, e.g., Razi et al. (2015), will be needed.

Declaration of interest

None.

Ethics statement

This technical note makes use of existing neuroimaging data which is freely available online (https://www.fil.ion.ucl.ac.uk/spm/data/eeg_mmn/). The data set has been published online, with details about the associated ethical procedures here: Garrido et al. (2007).

Data and code availability statement

Functions to implement DCM model inversion for EEG, fMRI, and the Bayesian fusion are available as part of the SPM12 (https://www.fil.ion.ucl.ac.uk/spm/software/spm12/), or custom MATLAB code that will be made available as a demo routine through the SPM software package.

CRediT authorship contribution statement

Huilin Wei: Methodology, Software, Formal analysis, Visualization, Writing - original draft. **Amirhossein Jafarian:** Methodology, Validation, Writing - review & editing. **Peter Zeidman:** Methodology, Validation, Writing - review & editing. **Vladimir Litvak:** Software, Validation. **Adeel Razi:** Visualization, Writing - review & editing. **Dewen Hu:** Writing - review & editing, Supervision, Funding acquisition. **Karl J. Friston:** Conceptualization, Methodology, Validation, Writing - review & editing, Supervision.

Acknowledgements

References

Anenberg, E., Chan, A.W., Xie, Y., LeDue, J.M., Murphy, T.H., 2015. Optogenetic stimulation of GABA neurons can decrease local neuronal activity while increasing cortical blood flow. J. Cerebr. Blood Flow Metabol. 35, 1579–1586.

Auksztulewicz, R., Friston, K., 2015. Attentional enhancement of auditory mismatch responses: a DCM/MEG study. Cerebr. Cortex 25, 4273–4283.

Bastos, A.M., Litvak, V., Moran, R., Bosman, C.A., Fries, P., Friston, K.J., 2015. A DCM study of spectral asymmetries in feedforward and feedback connections between visual areas V1 and V4 in the monkey. Neuroimage 108, 460–475.

Bastos, A.M., Usrey, W.M., Adams, R.A., Mangun, G.R., Fries, P., Friston, K.J., 2012. Canonical microcircuits for predictive coding. Neuron 76, 695–711.

Benetti, S., Pettersson-Yeo, W., Allen, P., Catani, M., Williams, S., Barsaglini, A., Kambeitz-Ilankovic, L.M., McGuire, P., Mechelli, A., 2015. Auditory verbal hallucinations and brain dysconnectivity in the perisylvian language network: a multimodal investigation. Schizophr. Bull. 41, 192–200.

Bhatt, M.B., Bowen, S., Rossiter, H.E., Dupont-Hadwen, J., Moran, R.J., Friston, K.J., Ward, N.S., 2016. Computational modelling of movement-related beta-oscillatory dynamics in human motor cortex. Neuroimage 133, 224–232.

Boly, M., Garrido, M.I., Gosseries, O., Bruno, M.A., Boveroux, P., Schnakers, C., Massimini, M., Litvak, V., Laureys, S., Friston, K., 2011. Preserved feedforward but impaired top-down processes in the vegetative state. Science 332, 858–862.

Brown, H.R., Friston, K.J., 2012. Dynamic causal modelling of precision and synaptic gain in visual perception - an EEG study. Neuroimage 63, 223–231.

Burnham, K.P., Anderson, D.R., 2003. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. Springer, New York.

Buxton, R.B., Uludag, K., Dubowitz, D.J., Liu, T.T., 2004. Modeling the hemodynamic response to brain activation. Neuroimage 23 (Suppl. 1), S220–S233.

Campo, P., Garrido, M.I., Moran, R.J., Maestu, F., Garcia-Morales, I., Gil-Nagel, A., del Pozo, F., Dolan, R.J., Friston, K.J., 2012. Remote effects of hippocampal sclerosis on effective connectivity during working memory encoding: a case of connectional diaschisis? Cerebr. Cortex 22, 1225–1236.

Chen, C.C., Kiebel, S.J., Friston, K.J., 2008. Dynamic causal modelling of induced responses. Neuroimage 41, 1293–1312.

Chen, J.J., Pike, G.B., 2009. Origins of the BOLD post-stimulus undershoot. Neuroimage 46, 559–568.

D'Esposito, M., Deouell, L.Y., Gazzaley, A., 2003. Alterations in the BOLD fMRI signal with ageing and disease: a challenge for neuroimaging. Nat. Rev. Neurosci. 4, 863–872.

Daunizeau, J., Grova, C., Marrelec, G., Mattout, J., Jbabdi, S., Pelegrini-Issac, M., Lina, J.M., Benali, H., 2007. Symmetrical event-related EEG/fMRI information fusion in a variational Bayesian framework. Neuroimage 36, 69–87.

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

Dima, D., Frangou, S., Burge, L., Braeutigam, S., James, A.C., 2012. Abnormal intrinsic and extrinsic connectivity within the magnetic mismatch negativity brain network in schizophrenia: a preliminary study. Schizophr. Res. 135, 23–27.

Douglas, R.J., Martin, K.A., 1991. A functional microcircuit for cat visual cortex. J. Physiol. 440, 735–769.

Drysdale, P.M., Huber, J.P., Robinson, P.A., Aquino, K.M., 2010. Spatiotemporal BOLD dynamics from a poroelastic hemodynamic model. J. Theor. Biol. 265, 524–534.

Duchi, J., 2007. Derivations for Linear Algebra and Optimization, p. 3. Berkeley, California.

Fabiani, M., Gordon, B.A., Maclin, E.L., Pearson, M.A., Brumback-Peltz, C.R., Low, K.A., McAuley, E., Sutton, B.P., Kramer, A.F., Gratton, G., 2014. Neurovascular coupling in normal aging: a combined optical, ERP and fMRI study. Neuroimage 85 Pt 1, 592–607.

Frahm, J., Kruger, G., Merboldt, K.D., Kleinschmidt, A., 1996. Dynamic uncoupling and recoupling of perfusion and oxidative metabolism during focal brain activation in man. Magn. Reson. Med. 35, 143–148.

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

Friston, K., Mattout, J., Trujillo-Barreto, N., Ashburner, J., Penny, W., 2007. Variational free energy and the Laplace approximation. Neuroimage 34, 220–234.

Friston, K., Zeidman, P., Litvak, V., 2015. Empirical Bayes for dcm: a group inversion scheme. Front. Syst. Neurosci. 9, 164.

Friston, K.J., Bastos, A., Litvak, V., Stephan, K.E., Fries, P., Moran, R.J., 2012. DCM for complex-valued data: cross-spectra, coherence and phase-delays. Neuroimage 59, 439–455.

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

Friston, K.J., Kahan, J., Biswal, B., Razi, A., 2014. A DCM for resting state fMRI. Neuroimage 94, 396–407.

Friston, K.J., Litvak, V., Oswal, A., Razi, A., Stephan, K.E., van Wijk, B.C.M., Ziegler, G., Zeidman, P., 2016. Bayesian model reduction and empirical Bayes for group (DCM) studies. Neuroimage 128, 413–431.

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., Preller, K.H., Mathys, C., Cagnan, H., Heinzle, J., Razi, A., Zeidman, P., 2017. Dynamic causal modelling revisited. Neuroimage.

Garrido, M.I., Kilner, J.M., Kiebel, S.J., Friston, K.J., 2009a. Dynamic causal modeling of the response to frequency deviants. J. Neurophysiol. 101, 2620–2631.

Garrido, M.I., Kilner, J.M., Kiebel, S.J., Stephan, K.E., Friston, K.J., 2007. Dynamic causal modelling of evoked potentials: a reproducibility study. Neuroimage 36, 571–580.

Garrido, M.I., Kilner, J.M., Stephan, K.E., Friston, K.J., 2009b. The mismatch negativity: a review of underlying mechanisms. Clin. Neurophysiol. 120, 453–463.

Haeusler, S., Maass, W., 2007. A statistical analysis of information-processing properties of lamina-specific cortical microcircuit models. Cerebr. Cortex 17, 149–162.

Havlicek, M., Ivanov, D., Roebroeck, A., Uludag, K., 2017. Determining excitatory and inhibitory neuronal activity from multimodal fMRI data using a generative hemodynamic model. Front. Neurosci. 11, 616.

Havlicek, M., Roebroeck, A., Friston, K., Gardumi, A., Ivanov, D., Uludag, K., 2015. Physiologically informed dynamic causal modeling of fMRI data. Neuroimage 122, 355–372.

Heinzle, J., Hepp, K., Martin, K.A., 2007. A microcircuit model of the frontal eye fields. J. Neurosci. 27, 9341–9353.

Heinzle, J., Koopmans, P.J., den Ouden, H.E.M., Raman, S., Stephan, K.E., 2016. A hemodynamic model for layered BOLD signals. Neuroimage 125, 556–570.

Henson, R.N., Flandin, G., Friston, K.J., Mattout, J., 2010. A parametric empirical Bayesian framework for fMRI-constrained MEG/EEG source reconstruction. Hum. Brain Mapp. 31, 1512–1531.

Henson, R.N., Wakeman, D.G., Litvak, V., Friston, K.J., 2011. A parametric empirical Bayesian framework for the EEG/MEG inverse problem: generative models for multi-subject and multi-modal integration. Front. Hum. Neurosci. 5, 76.

Huber, L., Tse, D.H.Y., Wiggins, C.J., Uludag, K., Kashyap, S., Jangraw, D.C., Bandettini, P.A., Poser, B.A., Ivanov, D., 2018. Ultra-high resolution blood volume fMRI and BOLD fMRI in humans at 9.4T: capabilities and challenges. Neuroimage 178, 769–779.

Huster, R.J., Debener, S., Eichele, T., Herrmann, C.S., 2012. Methods for simultaneous EEG-fMRI: an introductory review. J. Neurosci. 32, 6053–6060.

Jafarian, A., Litvak, V., Cagnan, H., Friston, K.J., Zeidman, P., 2019. Neurovascular Coupling: Insights from Multi-Modal Dynamic Causal Modelling of fMRI and MEG arXiv preprint arXiv:1903.07478.

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

Kashyap, S., Ivanov, D., Havlicek, M., Sengupta, S., Poser, B.A., Uludag, K., 2018. Resolving laminar activation in human V1 using ultra-high spatial resolution fMRI at 7T. Sci. Rep. 8, 17063.

Katzel, D., Zemelman, B.V., Buetfering, C., Wolfel, M., Miesenbock, G., 2011. The columnar and laminar organization of inhibitory connections to neocortical excitatory cells. Nat. Neurosci. 14, 100–107.

Keller, G.B., Mrsic-Flogel, T.D., 2018. Predictive processing: a canonical cortical computation. Neuron 100, 424–435.

Kiebel, S., Friston, K.J., 2002. Anatomically informed basis functions in multisubject studies. Hum. Brain Mapp. 16, 36–46.

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

Kiebel, S.J., Garrido, M.I., Friston, K.J., 2007a. Dynamic causal modelling of evoked responses: the role of intrinsic connections. Neuroimage 36, 332–345.

Kiebel, S.J., Garrido, M.I., Moran, R., Chen, C.C., Friston, K.J., 2009. Dynamic causal modeling for EEG and MEG. Hum. Brain Mapp. 30, 1866–1876.

Kiebel, S.J., Kloppel, S., Weiskopf, N., Friston, K.J., 2007b. Dynamic causal modeling: a generative model of slice timing in fMRI. Neuroimage 34, 1487–1496.

Kim, T., Hendrich, K.S., Masamoto, K., Kim, S.G., 2007. Arterial versus total blood volume changes during neural activity-induced cerebral blood flow change: implication for BOLD fMRI. J. Cerebr. Blood Flow Metabol. 27, 1235–1247.

Kullback, S., Leibler, R.A., 1951. On information and sufficiency. Ann. Math. Stat. 22, 79–86.

Lawrence, S.J.D., van Mourik, T., Kok, P., Koopmans, P.J., Norris, D.G., de Lange, F.P., 2018. Laminar organization of working memory signals in human visual cortex. Curr. Biol. 28, 3435–3440 e3434.

Lei, X., Valdes-Sosa, P.A., Yao, D., 2012. EEG/fMRI fusion based on independent component analysis: integration of data-driven and model-driven methods. J. Integr. Neurosci. 11, 313–337.

Li, B., Daunizeau, J., Stephan, K.E., Penny, W., Hu, D., Friston, K., 2011. Generalised filtering and stochastic DCM for fMRI. Neuroimage 58, 442–457.

Litvak, V., Garrido, M., Zeidman, P., Friston, K., 2015. Empirical Bayes for group (dcm) studies: a reproducibility study. Front. Hum. Neurosci. 9, 670.

Markov, N.T., Ercsey-Ravasz, M., Van Essen, D.C., Knoblauch, K., Toroczkai, Z., Kennedy, H., 2013. Cortical high-density counterstream architectures. Science 342, 1238406.

Markuerkiaga, I., Barth, M., Norris, D.G., 2016. A cortical vascular model for examining the specificity of the laminar BOLD signal. Neuroimage 132, 491–498.

Mechelli, A., Lin, A., Wood, S., McGorry, P., Amminger, P., Tognin, S., McGuire, P., Young, J., Nelson, B., Yung, A., 2017. Using clinical information to make individualized prognostic predictions in people at ultra high risk for psychosis. Schizophr. Res. 184, 32–38.

Moran, R., Pinotsis, D.A., Friston, K., 2013. Neural masses and fields in dynamic causal modeling. Front. Comput. Neurosci. 7, 57.

Moran, R.J., Jones, M.W., Blockeel, A.J., Adams, R.A., Stephan, K.E., Friston, K.J., 2015. Losing control under ketamine: suppressed cortico-hippocampal drive following acute ketamine in rats. Neuropsychopharmacology 40, 268–277.

Muthukumaraswamy, S.D., Shaw, A.D., Jackson, L.E., Hall, J., Moran, R., Saxena, N., 2015. Evidence that subanesthetic doses of ketamine cause sustained disruptions of NMDA and AMPA-mediated frontoparietal connectivity in humans. J. Neurosci. 35, 11694–11706.

Ninomiya, T., Sawamura, H., Inoue, K., Takada, M., 2012. Segregated pathways carrying frontally derived top-down signals to visual areas MT and V4 in macaques. J. Neurosci. 32, 6851–6858.

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

Pi, H.J., Hangya, B., Kvitsiani, D., Sanders, J.I., Huang, Z.J., Kepecs, A., 2013. Cortical interneurons that specialize in disinhibitory control. Nature 503, 521–524.

Pinotsis, D.A., Brunet, N., Bastos, A., Bosman, C.A., Litvak, V., Fries, P., Friston, K.J., 2014. Contrast gain control and horizontal interactions in V1: a DCM study. Neuroimage 92, 143–155.

Pinotsis, D.A., Geerts, J.P., Pinto, L., FitzGerald, T.H.B., Litvak, V., Auksztulewicz, R., Friston, K.J., 2017. Linking canonical microcircuits and neuronal activity: dynamic causal modelling of laminar recordings. Neuroimage 146, 355–366.

Pinotsis, D.A., Moran, R.J., Friston, K.J., 2012. Dynamic causal modeling with neural fields. Neuroimage 59, 1261–1274.

Poplawsky, A.J., Fukuda, M., Murphy, M., Kim, S.G., 2015. Layer-specific fMRI responses to excitatory and inhibitory neuronal activities in the olfactory bulb. J. Neurosci. 35, 15263–15275.

Puckett, A.M., Aquino, K.M., Robinson, P.A., Breakspear, M., Schira, M.M., 2016. The spatiotemporal hemodynamic response function for depth-dependent functional imaging of human cortex. Neuroimage 139, 240–248.

Razi, A., Kahan, J., Rees, G., Friston, K.J., 2015. Construct validation of a DCM for resting state fMRI. Neuroimage 106, 1–14.

Razi, A., Seghier, M.L., Zhou, Y., McColgan, P., Zeidman, P., Park, H.J., Sporns, O., Rees, G., Friston, K.J., 2017. Large-scale DCMs for resting-state fMRI. Netw Neurosci 1, 222–241.

Rosa, M.J., Daunizeau, J., Friston, K.J., 2010. EEG-fMRI integration: a critical review of biophysical modeling and data analysis approaches. J. Integr. Neurosci. 9, 453–476.

Rosa, M.J., Kilner, J.M., Penny, W.D., 2011. Bayesian comparison of neurovascular coupling models using EEG-fMRI. PLoS Comput. Biol. 7, e1002070.

Scheeringa, R., Koopmans, P.J., Van, M.T., Jensen, O., Norris, D.G., 2016. The relationship between oscillatory EEG activity and the laminar-specific BOLD signal. Proc. Natl. Acad. Sci. U. S. A. 113, 201522577.

Self, M.W., van Kerkoerle, T., Goebel, R., Roelfsema, P.R., 2019. Benchmarking laminar fMRI: neuronal spiking and synaptic activity during top-down and bottom-up processing in the different layers of cortex. Neuroimage 197, 806–817.

Shaw, A.D., Hughes, L.E., Moran, R., Coyle-Gilchrist, I., Rittman, T., Rowe, J.B., 2019. In vivo assay of cortical microcircuitry in frontotemporal dementia: a platform for experimental medicine studies. Cerebr. Cortex.

Shaw, A.D., Moran, R.J., Muthukumaraswamy, S.D., Brealy, J., Linden, D.E., Friston, K.J., Singh, K.D., 2017. Neurophysiologically-informed markers of individual variability and pharmacological manipulation of human cortical gamma. Neuroimage 161, 19–31.

Shipp, S., 2016. Neural elements for predictive coding. Front. Psychol. 7, 1792.

Sohal, V.S., Zhang, F., Yizhar, O., Deisseroth, K., 2009. Parvalbumin neurons and gamma rhythms enhance cortical circuit performance. Nature 459, 698–702.

Stephan, K.E., Iglesias, S., Heinzle, J., Diaconescu, A.O., 2015. Translational perspectives for computational neuroimaging. Neuron 87, 716–732.

Stephan, K.E., Weiskopf, N., Drysdale, P.M., Robinson, P.A., Friston, K.J., 2007. Comparing hemodynamic models with DCM. Neuroimage 38, 387–401.

Symmonds, M., Moran, C.H., Leite, M.I., Buckley, C., Irani, S.R., Stephan, K.E., Friston, K.J., Moran, R.J., 2018. Ion channels in EEG: isolating channel dysfunction in NMDA receptor antibody encephalitis. Brain 141, 1691–1702.

Thomson, A.M., Bannister, A.P., 2003. Interlaminar connections in the neocortex. Cerebr. Cortex 13, 5–14.

Tsvetanov, K.A., Henson, R.N., Tyler, L.K., Razi, A., Geerligs, L., Ham, T.E., Rowe, J.B., Cambridge Centre for, A., Neuroscience, 2016. Extrinsic and intrinsic brain network connectivity maintains cognition across the lifespan despite accelerated decay of regional brain activation. J. Neurosci. 36, 3115–3126.

Uludag, K., Blinder, P., 2018. Linking brain vascular physiology to hemodynamic response in ultra-high field MRI. Neuroimage 168, 279–295.

Uludag, K., Roebroeck, A., 2014. General overview on the merits of multimodal neuroimaging data fusion. Neuroimage 102 (Pt 1), 3–10.

Valdes-Sosa, P.A., Sanchez-Bornot, J.M., Sotero, R.C., Iturria-Medina, Y., Aleman-Gomez, Y., Bosch-Bayard, J., Carbonell, F., Ozaki, T., 2009. Model driven EEG/fMRI fusion of brain oscillations. Hum. Brain Mapp. 30, 2701–2721.

van Zijl, P.C., Hua, J., Lu, H., 2012. The BOLD post-stimulus undershoot, one of the most debated issues in fMRI. Neuroimage 62, 1092–1102.

Wakeman, D.G., Henson, R.N., 2015. A multi-subject, multi-modal human neuroimaging dataset. Sci Data 2, 150001.

Xia, M., Wang, J., He, Y., 2013. BrainNet Viewer: a network visualization tool for human brain connectomics. PloS One 8, e68910.

Zeidman, P., Kazan, S.M., Todd, N., Weiskopf, N., Friston, K.J., Callaghan, M.F., 2018. Optimising data for modelling neuronal responses. Front. Neurosci. 12, 986.