Dynamic effective connectivity

Tahereh S. Zarghami [a,*], Karl J. Friston [b]

[a] Bio-Electric Department, School of Electrical and Computer Engineering, University of Tehran, Amirabad, Tehran, Iran
[b] The Wellcome Centre for Human Neuroimaging, University College London, Queen Square, London, WC1N 3AR, UK

ARTICLE INFO

ABSTRACT

Metastability is a key source of itinerant dynamics in the brain; namely, spontaneous spatiotemporal reorganization of neuronal activity. This itinerancy has been the focus of numerous dynamic functional connectivity (DFC) analyses – developed to characterize the formation and dissolution of distributed functional patterns over time, using resting state fMRI. However, aside from technical and practical controversies, these approaches cannot recover the neuronal mechanisms that underwrite itinerant (e.g., metastable) dynamics—due to their descriptive, model-free nature. We argue that effective connectivity (EC) analyses are more apt for investigating the neuronal basis of metastability. To this end, we appeal to biologically-grounded models (i.e., dynamic causal modelling, DCM) and dynamical systems theory (i.e., heteroclinic sequential dynamics) to create a probabilistic, generative model of haemodynamic fluctuations. This model generates trajectories in the parametric space of EC modes (i.e., states of connectivity) that characterize functional brain architectures. In brief, it extends an established spectral DCM, to generate functional connectivity data features that change over time. This foundational paper tries to establish the model's face validity by simulating non-stationary fMRI time series and recovering key model parameters (i.e., transition probabilities among connectivity states and the parametric nature of these states) using variational Bayes. These data are further characterized using Bayesian model comparison (within and between subjects). Finally, we consider practical issues that attend applications and extensions of this scheme. Importantly, the scheme operates within a generic Bayesian framework – that can be adapted to study metastability and itinerant dynamics in any non-stationary time series.

1. Introduction

The brain continuously expresses transient patterns of coordinated activity that emerge and dissolve in response to internal and external perturbations. The emergence and evolution of such metastable coordination dynamics (in self-organizing complex systems, such as the brain) has been studied extensively in nonlinear dynamics systems theory. Furthermore, characterising itinerancy in the brain has attracted increasing interest with advances in functional neuroimaging (Deco et al., 2017; Roberts et al., 2019; Tognoli and Kelso, 2014). There is still much to learn about the mechanisms of spontaneous spatiotemporal reorganization in the brain, its neuronal underpinnings, and its relation to cognition and behaviour. In particular, metastable dynamics are closely tied to notions of functional integration and functional segregation (specialization) in the brain (Friston, 2002)—and how one contextualises the other.

The tendency of segregated neuronal populations to communicate over a distributed network is known as *functional integration*. Although

neuroanatomy (i.e. *structural connectivity*) provides the infrastructure for cross-talk between regions, there need not necessarily be a direct (mono) synaptic, anatomical link between two areas to mediate functional integration. Hence, to get a rough picture of functional integration, one can quantify statistical dependency between distinct (observable) regional responses, which is generally known as *functional connectivity* (FC). Examples of such statistical relationships include correlations, coherence and transfer entropy. Conversely, estimating how remote neuronal dynamics influence (i.e., cause) each other—in different contexts—is the topic of *effective connectivity* (EC) research (Friston, 2011).

In past years, the functional neuroimaging community has focused on characterizing the spontaneous reorganization of functional connectivity patterns at rest (i.e. during task-free periods). Considerable effort has been invested into developing methods and procedures that can track changes in FC patterns through time. This field is known as *dynamic functional connectivity* (DFC). Among the non-invasive functional neuroimaging modalities that have been used for this purpose, functional magnetic resonance imaging (fMRI) remains particularly popular, given

---

its high spatial resolution and inherent circumvention of the inverse problem (i.e., source localization).

Since the inception of DFC, the brain connectivity community has been contending with a number of theoretical and practical issues. These include ongoing debates over: the most reliable (sensitive and robust) FC measures and techniques, definitions of (non)stationarity, the pertinent time scales of changes, the cognitive relevance of the fluctuations, the extent of artefactual (e.g. motion related) effects, interpretational issues, and so on; see (Hutchison et al., 2013; Preti et al., 2017) for review and references. However, the more fundamental limitation of DFC (for studying functional integration and metastability) inherits from the *descriptive* nature of FC. In other words, statistical relationships (or changes thereof) between measured brain signals do not disclose the underlying neuronal dynamics and their interactions. In the case of fMRI, the problem is further compounded by the neurovascular coupling of neuronal dynamics to measurable hemodynamics—rendering fMRI a slow and indirect measure of neuronal activity. Therefore, as useful as (D)FC results can be – as potential biomarkers for classification and prediction purposes – they are not apt for disclosing the mechanisms of functional integration (and endogenous or autonomous dynamics) in the brain. Put succinctly, this is because "[d]ynamics occur in the underlying neuronal system … then they cause nontrivial statistics in time-resolved data samples" (Heitmann and Breakspear, 2018); the latter is what DFC analyses can quantify, without a straightforward connection to the former.

The closest the community has come to grounding DFC fluctuations in neuronal phenomena has been by (1) establishing correlations with power envelopes of certain band-limited electrophysiological and electroencephalography (EEG) recordings (Chang et al., 2013; Tagliazucchi et al., 2012; Thompson et al., 2013) and (2) by showing overall accordance with emergent dynamics from large-scale computational models (Deco et al., 2017, 2011; Rabinovich and Varona, 2011). These initiatives acknowledge that fluctuations in functional connectivity have neuronal correlates – and are not simply artefactual, despite physiological and measurement noise. However, one still needs to solve an inverse problem to interpret functional connectivity in terms of function integration mediated by effective connectivity: i.e., a procedure to map these (observable and) evolving spatiotemporal patterns back to the (unobservable and) evolving neuronal networks that generate them. This can only be achieved by constructing generative models of the observations, and adopting an inference procedure to estimate time varying effective connectivity.

Effective connectivity is – by definition – model-based. That is, it relies on generative models of coupled neuronal systems that can explain empirical observations. The goal is to identify the model (network) that has the greatest evidence for any given data, and to infer the corresponding (context-sensitive) coupling or effective connectivity parameters. So how can effective connectivity analysis contribute to our understating of functional integration (particularly at rest), and its presumed transient dynamics?

Among the existing EC approaches, we focus on the most established: dynamic causal modelling (DCM) (Friston et al., 2003). Specifically, we opt for a spectral dynamic causal model (spDCM) for task-free fMRI (Friston et al., 2014b). This spectral DCM is an efficient biologically-grounded model that can explain (i.e., predict) the complex cross spectra of the blood oxygenation level dependent (BOLD) signals generated by endogenous neuronal fluctuations. Based on some local linearity assumptions (on the state space equations of neuronal dynamics, and the observation equation for the measured hemodynamic signal) – together with parametrized power law distributions for the cross spectra (of endogenous fluctuations and observation noise) – spectral DCM effectively converts a (nonlinear) state space model into a (linear) deterministic model, which can be inverted quickly and efficiently (Razi et al., 2015). This model has been validated elsewhere (Friston et al., 2014a; Razi et al., 2015) and has been used extensively in imaging neuroscience over the past few years. Importantly, since cross spectra are

the Fourier counterparts of cross-correlations, and a zero-lag cross-correlation is the most prevalent FC measure, spDCM is essentially an EC model of how FC is generated.

To date, spectral DCM has been applied largely to whole sessions of resting state (rs)fMRI data, inferring networks and couplings that explain the session-average cross-spectra (i.e. the Fourier generalisation of FC). To track network reconfigurations through time, a few studies have applied spDCM to successive epochs of data. For instance, in the context of epilepsy research, spDCM (in a more physiologically detailed form) has been applied to successive epochs of (invasive and non-invasive) EEG recordings and electrophysiological data to follow the evolution of the coupling parameters through time (Cooray et al., 2015; Papadopoulou et al., 2017, 2015).

Recently, spDCM has been used to estimate the fluctuations of coupling parameters in successive epochs of rsfMRI data (Park et al., 2017). To this end, the authors added a linear probabilistic model to the generative model of spDCM, which allowed for parameter variations (over epochs) to be explained as a weighted average of a number of (constant and oscillatory) regressors. This augmented – parametric empirical Bayesian (PEB) (Friston et al., 2016) – scheme is essentially a Bayesian general linear model (GLM), with the advantage that the posterior expectations *and* uncertainty about EC parameters are accounted for; furnishing a Bayesian treatment of between-epoch effects. For a concise mathematical description of spDCM and PEB see (Park et al., 2017). The key finding of this research was that the subject-specific baseline effective connectivity is more consistent, across sessions, when dynamic components are modelled explicitly. However, interestingly, the authors also noted that the first (subject-specific) eigenvariate of temporal fluctuations in effective connectivity was also conserved across sessions. This speaks to a reproducible connectivity dynamics that we pursue in the present paper, through hierarchical Bayesian modelling.

The generative model that we propose is a mixed (discrete and continuous) hierarchical model that calls on constructs from nonlinear dynamical systems theory to explain metastable dynamics in the brain. To this end, we associate the itinerancy in configurations or modes of EC with reproducible transient dynamics (characteristic of complex self-organizing systems). Under minimal assumptions, this connectivity dynamics can be cast as a latent variable or hidden Markov model—that we adopt and implement. With such a generative model at hand, the outstanding issue is to establish and validate a procedure for efficient model inversion (i.e., mapping from a set of observations to the underlying model parameters and hidden states). It is worth reiterating that generative models of EC dynamics can explain the temporal evolution of functional connectivity patterns (of the sort investigated in DFC). In other words, in the same sense that conventional DCM for CSD provides a complete explanation of functional connectivity, the current scheme explains dynamic functional connectivity.

In what follows, we describe the fundaments of our generative model, outline the accompanying (variational Bayesian) inference procedure and establish the face-validity of the approach (through simulations and numerical analyses). We conclude with a discussion of the implications of the analyses, along with some potential applications.

2. Methods and theory

2.1. The generative model

Our generative model underlying itinerant brain states is based on the notion that macroscopic (slow) dynamical modes (Haken, 1983; Jirsa et al., 1994) visit a succession of unstable fixed points in the parameter space of directed (effective) connectivity. In other words, we suppose that neuronal dynamics are generated by patterns of intrinsic (within-region) and extrinsic (between-region) connectivity changes over time. In particular, we assume that patterns of connectivity trace out a heteroclinic orbit (i.e., a stable heteroclinic cycle or SHC) in parameter space; visiting a discrete number of unstable fixed points to produce a

winnerless competition among different states or modes of brain activity[1] (Afraimovich et al., 2008; Deco and Jirsa, 2012; Friston et al., 2014b). We further assume that the transitions from one unstable fixed point to the next are fast in relation to the amount of time connectivity remains in the neighbourhood of a fixed point. This assumption is licensed by two features of SHC. These include: (i) the origin of the structural stability of the SHC, and (ii) the long passage time in the vicinity of saddles in the presence of moderate noise (Rabinovich and Varona, 2018). The structural stability speaks to a discrete approximation to any SHC in terms of a discrete number of connectivity modes, while the long passage time endorses a model of discrete transitions from one connectivity state to the next.

These characteristics of stable heteroclinic channels licence a hidden Markov model (HMM) of transitions among connectivity states (i.e., unstable fixed points in parameter space), where each connectivity state generates fast neuronal fluctuations that can be observed over epochs, in terms of their cross spectral density (Daunizeau and Friston, 2007). This follows in the spirit of (Rabinovich et al., 2008): who note that "The reproducible transient dynamics based on SHC that we have discussed contains two different time scales, i.e., a slow time scale in the vicinity of the saddles and a fast time scale in the transitions between them. Taking this into account, it is possible to build a dynamical model based not on ODEs but on a Poincare map, which can be computationally very efficient for modelling a complex system". See also (Krupa, 1997) for a discussion of robust heteroclinic cycles.[2]

Fig. 1 illustrates the generative model we have in mind. In brief, this model uses a hidden Markov model of switching states that are associated with a particular pattern of connectivity. Each connectivity state then generates fast neuronal dynamics, driven by endogenous neuronal fluctuations within each brain region. More specifically, one can describe this model in terms of how data would be generated. First, we sample a state transition matrix from a Dirichlet distribution parameterised by its concentration parameters b. This transition matrix (B) is then used to select the current brain state given the previous state; assuming a small number (n) of hidden states (s). The current state then selects a state specific pattern of connectivity (β), sampled from a multivariate Gaussian distribution. A random Gaussian variate is then added to this pattern to generate the connectivity (θ) for the current epoch. This connectivity

defines the neuronal network's transfer functions (K) and cross spectra ($\mathbf{Y}(\omega)$), under local linearity assumptions – and parameterised, scale free, endogenous fluctuations in each brain region. The predicted or generated cross spectra are then used to generate sample cross spectra.

A technical aspect of this generative model is the form of the likelihood for the complex cross spectra that are generated. With ideal estimators, one could assume that these spectral data features had a Wishart distribution, with one degree of freedom for each frequency. However, we assume that the cross spectra constitute the average of estimates, with consistent and asymptotically normal estimates. In this setting, the variance of the difference between the predicted and observed spectral estimates is equal to the cross spectral density times the (effective) degrees of freedom, which we treat as an unknown parameter. See (Camba-Mendez and Kapetanios, 2005) for details.

2.2. Model inversion with variational message passing

Fig. 2 shows the same generative model but in the form of a normal or Forney style factor graph, as compared to the graphical model shown in Fig. 1. A factor graph is a useful schematic summary because it shows clearly the architecture of message passing implicit in model fitting or inversion. The graphical model in Fig. 1 associates nodes with random variables and edges with conditional dependencies. Conversely, the factor graph in Fig. 2 associates nodes with the factors or marginal probability distributions of the generative model (denoted by the numbers in green squares), while the edges now correspond to random variables. Once the architecture of the (Forney style) factor graph has been specified in this way, one can envisage model inversion in terms of variational message passing along the edges (in both directions).

Model inversion refers to inverting the generative model of the probabilistic mapping between causes and observable consequences, so that one can map from observations to (inferred) causes; here, the unknown or hidden parameters of the generative model. The key parameters of interest include the probability transition matrix among hidden states (B) and the patterns of connectivity (β) associated with each of n states. However, to evaluate the posterior over these quantities one also has to optimise the posterior over all other parameters (and hyperparameters). For example, one has to infer the amplitude of various random (Gaussian) fluctuations, the (scale free) form of endogenous neuronal fluctuations and the degrees of freedom with which the cross spectra were estimated.

The equations on the left of Fig. 2 summarise the principles of variational inversion. In brief, one constructs a variational free energy bound on model evidence (also known as an ELBO in machine learning). This lower evidence bound is a functional of the observed data and an approximate posterior density or Bayesian 'belief'. The data in this instance are the cross spectra estimated for a succession of epochs that constitute an fMRI timeseries. The approximate posterior corresponds to a mean field approximation over all unknown parameters and states and is denoted by Q. This approximate posterior is then optimized with respect to the free energy functional, thereby minimising the KL divergence between the approximate posterior belief and the true posterior (see equalities in Fig. 2). At the same time, this optimisation renders the free energy a lower bound on model evidence. This is important because it means we can use the free energy functional to compare different models in terms of the evidence or marginal likelihood (e.g., models with different numbers of hidden Markov states).

Fig. 2 also shows that the solutions for the approximate beliefs render the variation with respect to posterior beliefs (or derivative with respect to their sufficient statistics $\mathbf{q}$) zero. The solutions under the generative model in Fig. 1 are provided in the white panel. These are effectively the messages that are passed between model factors, until convergence and maximisation of free energy. These messages (denoted by the circled numbers) are shown on the factor graph on the right to illustrate the bidirectional and hierarchical message passing implicit in this scheme. For people who are familiar with fMRI data analysis, this message passing

---

[1] In general, winnerless competition need not be a cyclic process: "The participants of such a process can become winners periodically, or, especially when the number of participants is more than three, the process can be non-cyclic and even can be terminated following a stable sequence of transients and one participant becomes the ultimate winner (Busse and Heikes, 1980)" as cited in (Afraimovich et al., 2010). However, the spontaneous quasi-periodic recurrence of sequences of connectivity modes speaks to stable heteroclinic cycles (Kashyap and Keilholz, 2019; Majeed et al., 2011). In theory—and evidenced in vivo—these (periodic) cycles might be embedded at different levels of the temporal hierarchy of neuronal dynamics (Cabral et al., 2017; Handwerker et al., 2012; Vidaurre et al., 2017). We elaborate on the distinction between heteroclinic channels and cycles in the discussion.

[2] Who note: "A notable feature of robust cycles is that they can be asymptotically stable (or possess some weaker form of stability). Intuitively, stability can be expected when the stable eigenvalues of the equilibria on the cycle are on the average stronger than the unstable ones. A stable cycle defines a mechanism of intermittency—a solution approaching it spends long periods near equilibria and makes fast transitions from one equilibrium to the next. In a perfectly symmetric system, the return times increase monotonically and rapidly approach infinity, thus making the intermittent behaviour uninteresting. However, under small, symmetry-breaking perturbations, the cycling behaviour persists (even if there no longer is a cycle) and the transition times no longer converge to infinity. In many cases the transition times are either bounded or extremely long transition times are very infrequent. In a similar manner stochastic perturbations, or round-off errors in numerical computations, lead to boundedness of transition times. Hence, in applications, the existence of a stable heteroclinic cycle in the idealized model problem can be linked to the occurrence of intermittence."
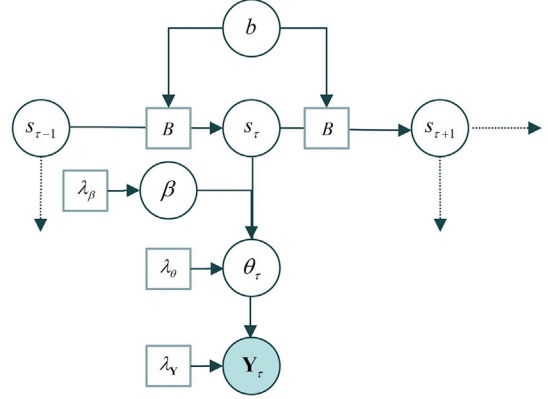
Fig. 1. The generative model in terms of its constituent factors (left) and corresponding graphical model (right). Specifically, B indicates the transition matrix, s denotes the hidden states, β stands for state-specific connectivity pattern, θ_τ reflects epoch-specific connectivity and $Y_\tau(\omega)$ represents the complex cross spectra of an epoch. For more details concerning the generative model, please refer to the main text.

Fig. 2. Variational message passing (left) and normal style factor graph (right) based upon the generative model in Fig. 1. Please note the key inferred model parameters: B indicates the posterior expectation of the transition matrix, s denotes the posterior expectation of hidden states, β stands for the posterior state-specific connectivity patterns, θ reflects the posterior epoch-specific connectivity and Y_τ represents the complex cross spectral data features of an epoch.

scheme can be divided into three parts: the lower part corresponds to a conventional *dynamic causal modelling* (DCM) analysis of complex cross spectra within each epoch. The parameter estimates from the DCM then constitute the evidence for a hierarchical model of changing connectivity over epochs; estimated using *parametric empirical Bayes* (PEB). Finally, the third part of this model corresponds to a *hidden Markov model* (HMM), which furnishes empirical priors for the PEB scheme.

In summary, changes in connectivity are constrained by the prior belief that there are transitions among a small number of brain connectivity states, while connectivity estimates from the PEB level provides evidence for the hidden Markov model that the brain is in one state or another. This evidence is summarised as L in the figure. Interestingly, under this formulation of model inversion, one can regard the hidden Markov model as providing Bayesian model averages for the PEB (intermediate) level that in turn provide empirical priors for each DCM epoch. The ascending messages about the current connectivity – that provide evidence for connectivity states in the hidden Markov model – are shown in green while the descending messages – that constitute Bayesian model averages of connectivity – are shown in red.

Crucially, the variational message passing that emerges from free energy minimisation has, unsurprisingly, formal similarities with many established schemes; including the Baum-Walsh algorithm, expectation maximisation and nonlinear system identification in timeseries analysis. In what follows, we try to establish the face validity of this inversion scheme by generating synthetic data and ensuring that we can recover the underlying state transitions. A focus of this validation will be to show how optimising variational free energy can also be used for Bayesian model selection; here, to estimate the number of hidden states and the nature of structured transitions.

2.3. Modelling latent states

The inversion problem we are dealing with presents many challenges. This is because we are trying to infer an unknown number of hidden or latent states, the probability transitions among those states and the likelihood mapping to the connectivity parameters at a lower level of a hierarchical model. Furthermore, this inversion accommodates uncertainty about the connectivity parameters at the lower level, conditioned upon the data. This is a difficult problem for several reasons.

First, there is an indeterminacy or degeneracy when trying to estimate both the probability transitions and the likelihood mapping from latent states. This follows because these mappings are only defined up to an unknown order or labelling of hidden states. In other words, permuting the states will change the transition and likelihood parameters without changing the data that are generated. We will see below that this degeneracy can be finessed by a systematic relabelling of the hidden states.

However, we are then still left with an ill-conditioned information geometry of the posterior parameter space. In other words, because the number of parameters (i.e., Dirichlet concentration parameters) of the probability transition matrices can be large, the free energy functional of this space is not convex. This means that the gradient ascent – inherent in the optimisation of variational free energy – will encounter numerous local maxima. The usual solution to this problem is to consider re-parameterisations of the generative model, to ensure that the free energy functional is well-behaved. An example of this can be found in (Friston et al., 2014b), where connectivity matrices are re-parameterised in terms of their eigenvalues and eigenvectors. However, here, we do not have the latitude to reparametrize the probability transitions, due to their Dirichlet parameterisation.

An alternative approach to the local maxima problem is to initialise the variational inversion at different starting points. Normally, this would be computationally intractable, due to the high dimensional problem at hand. However, we can use the computational efficiency afforded by Bayesian model reduction (Friston et al., 2016) to perform an exhaustive search of all plausible initialisations. This requires us to define a plausible initialisation in terms of prior beliefs about the probability

transitions. As noted in the introduction, the use of a hidden Markov model is motivated by the notion of orbits[3] through connectivity parameter space that visit a succession of unstable fixed points (Rabinovich et al., 2008; Rabinovich et al., 2012). This assumption tells us immediately that there are certain attributes that the probability transition matrices must possess. These can be summarised as follows:

- First, the dwell time at any unstable fixed point must be small. In other words, the probability of staying in a particular latent state of connectivity – at the time scale imposed by the choice of epoch length – is small. We will refer to this as a **dwell prior.**
- Second, there will be a systematic structure to the orbit among latent states; such that the transition from one unstable fixed point to the next is relatively predictable. We will call this an **orbit prior**
- Finally, if a certain number of hidden states are necessary to explain the data, then they must all be occupied for a nontrivial proportion of the time over which the data were observed. We will refer to this as an equilibrium or **equidistribution prior.**

It is relatively straightforward to formalise these three prior beliefs in terms of corresponding prior energies as follows, where a · denotes an inner or dot product:

$$
-\ln p(B) = \underbrace{\ln\left(\frac{4}{n} \cdot trace(B) + e^{-\alpha}\right)}_{\text{Dwell potential}} \underbrace{-\beta \ln n \cdot \frac{1}{n} trace(B \cdot \ln B)}_{\text{Orbit potential}} \\
\underbrace{+\beta \ln n \cdot eig(B) \cdot \ln eig(B)}_{\text{Equilibrium potential}}
\tag{1}
$$

The first term (*dwell prior*) is simply the log of the trace of the probability transition matrix. This means that if the probability of staying in any state is low, the trace or average probability will also be low, leading to a small prior potential. The second term is the average conditional entropy of transitions, such that a deterministic transition from one state to another state results in a minimal conditional entropy (of zero). The final term is the entropy over the equilibrium distribution (obtained from the principal eigenvector of the density dynamics; i.e., the probability transition minus the identity matrix). If all n states are occupied with equal probability over time, then this entropy (i.e., equilibrium potential) attains its minimum of zero. The hyperparameters (α, β) ensure that the prior potentials scale naturally with beliefs about what is likely and unlikely. For example, the parameter $\beta = 8$ means the entropy terms have a range of eight. This means one can interpret potential differences as log Bayes factors; namely, a strong prior belief would correspond to a prior potential difference of three, while a very strong prior belief would correspond to a potential difference of five (Kass and Raftery, 1995).

Equipped with this formal (**itinerancy**) prior over probability transition matrices, we can now perform an exhaustive search over all plausible probability transition matrices – by using samples from the prior distribution to initialise a gradient ascent. Furthermore, the log prior can also be entered into the gradient ascent; thereby destroying a large number of local maxima. Fig. 3 shows an example of plausible probability transition matrices – and some typical trajectories.

Equipped with the **itinerant prior** above, we can identify plausible prior transition matrices within an Occam's window of 3 (nats), for a given number of hidden states. In other words, we can define a plausible prior transition as being at most $\exp(3) \approx 20$ times less likely than the most probable. To finesse degeneracy in the state labels, we consider all permutations of each transition matrix and treat the column entries as

---

[3] We use the term *orbit* as a convenient shorthand for a stable heteroclinic cycle (SHC) – this is closely related to a heteroclinic orbit (which is the technical term for the path that connects different equilibria, in dynamical systems theory). In SHC, the fixed points are unstable (i.e., saddle points) and connected via saddle-sink connections.
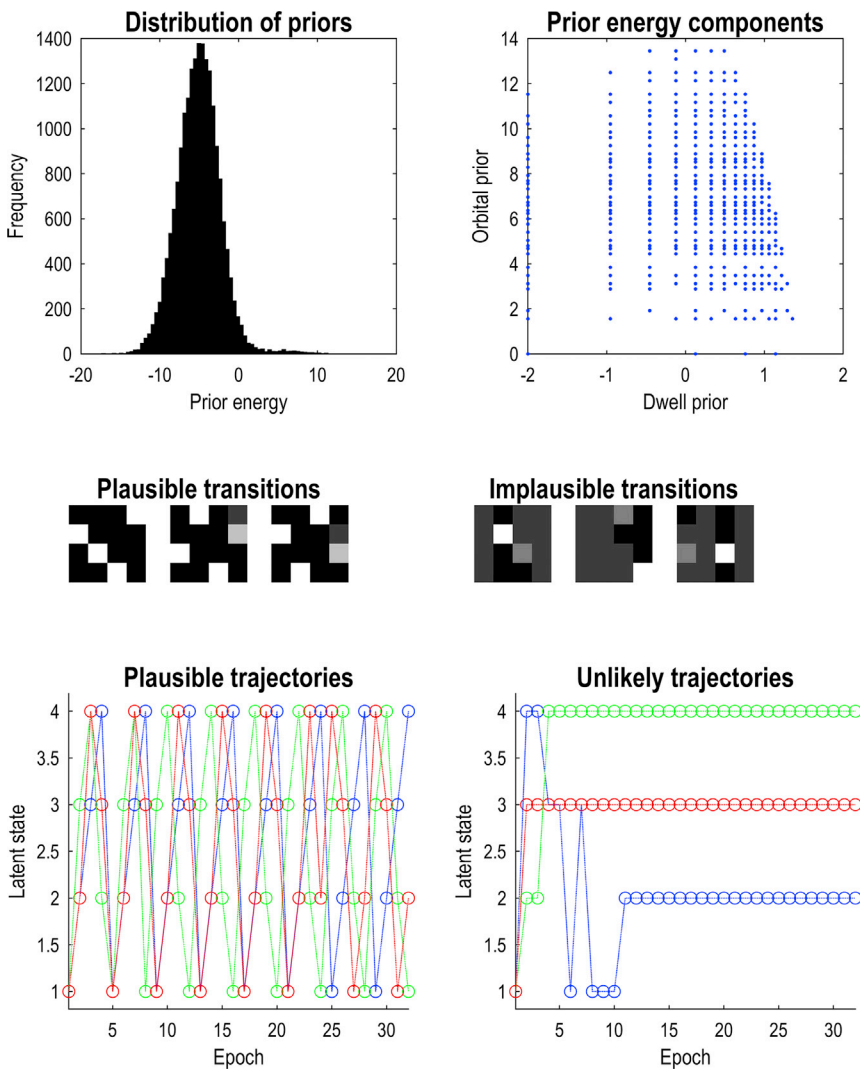
T.S. Zarghami, K.J. Friston

Fig. 3. This figure illustrates the distinction between *a priori* plausible and unlikely probability transition matrices, where the prior probability or potential of a probability transition matrix is defined above with three potential terms (**dwell**, **orbit** and **equilibrium** potentials). In this example, we created several thousand probability transition matrices by assigning four Dirichlet parameters at random to each column of transition matrices over four hidden or latent states. These matrices were then scored in terms of their prior potential. Unique transition matrices were then identified – on the basis of the three terms of the prior potential. The resulting distribution of prior potentials is shown on the upper left. For interest, the orbit potential is plotted against the dwell potential (for this sample) on the upper right. The middle row shows the three most likely (left) and least likely (right) transition matrices according to this specification of prior beliefs. Note the degeneracy in the state labels: the first and second transitions from the left appear distinct, whereas upon rearranging the labels, one can see that the second transition is also an orbit—with a slight uncertainty over one of the transitions. Exemplar trajectories for these transition matrices are shown in the corresponding panels below, starting from the first hidden state (in cyan). The key thing to observe is that trajectories that could be plausibly associated with metastable dynamics are generated by, and only by, those probability transition matrices that have a high prior probability (or low prior potential). In this example, the most likely prior probability transition matrix is a deterministic orbit among four states that would correspond to a stable heteroclinic cycle. The trajectories generated by unlikely matrices converge upon a single state, after a short period of time.

concentration parameters of a Dirichlet distribution. This furnishes the KL-divergence between each permuted transition and that of a canonical orbit of the same size (cycling through states 1 to n). This allows one to retain only the permuted (transition) matrix with the smallest divergence from the canonical orbit. Furthermore, all the hidden state sequences are presumed to start in state 1. This consistent reordering or alignment of the states resolves the degenerate labelling and enables one to perform an exhaustive search over (unique and aligned) transition priors – that also serve as the initialisation of the posterior and associated trajectory of states.

Having inverted the hidden Markov model under all plausible priors (using Bayesian model reduction), one can then perform a Bayesian model average by treating each prior as a model. This entails weighting the posterior Dirichlet parameters of the transition matrix by the relative evidence for each prior (based upon the associated free energy). Crucially, the itinerant prior potential is included during this model averaging; transforming the evidence into a posterior probability of each model of connectivity state transitions.

Practically, the model inversion scheme takes about a second to run for each prior transition. For example, with $(\alpha, \beta) = (2, 8)$ and with three hidden states, there are 9 plausible (unique) transition priors – using prior concentration parameters that sum to 3 for each transition; hence, initialisation for each subject takes about 9 s. By including only plausible priors within Occam's window, we are effectively excluding all the other possible transition matrices from the final Bayesian model average (under the assumption that their evidence would not be sufficiently high to compensate for their prior implausibility).

2.4. Bayesian model comparison

As noted above, the inference problem that we are contending with is extremely ill posed. This follows because there are an enormous number of connectivity state transitions that could explain the same data. However, one can finesse this problem by a careful specification of the inferences that are – and are not possible – based upon the sort of questions people want to ask of their data. In brief, we focus on Bayesian model comparison at the group level as the ultimate objective.

For any given subject, we will be interested in comparing models that do and do not feature transitions among states of effective connectivity. Formally, this requires the evidence for models of multiple connectivity states to be greater than models with a single connectivity state (i.e., no transitions). Having established the optimum number of states is greater than one, one can ask whether there is evidence for an orbit by comparing models with and without orbital priors. Finally, if there is evidence for (or against) an orbit, we can meaningfully interpret the parametric nature of these transient dynamics, in terms of the probability transition matrices and associated states of connectivity. Using simulations, we will show that this greedy search procedure (using Bayesian model comparison) correctly disambiguates orbits from irregular or non-orbits, and correctly identifies the number of connectivity states. In terms of group comparisons, we can repeat the above procedure, with and without modelling group differences at the between-subject level. In other words, we can first identify the best model for all subjects and then ask whether there are any differences in the group specific parameters,
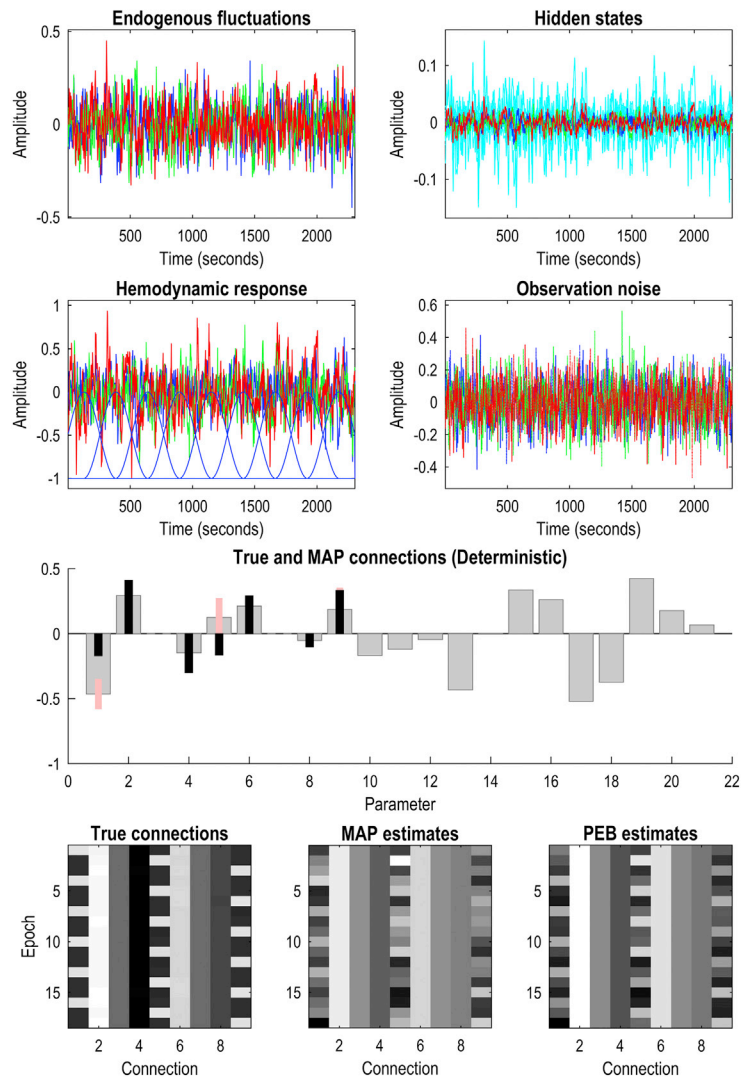
T.S. Zarghami, K.J. Friston
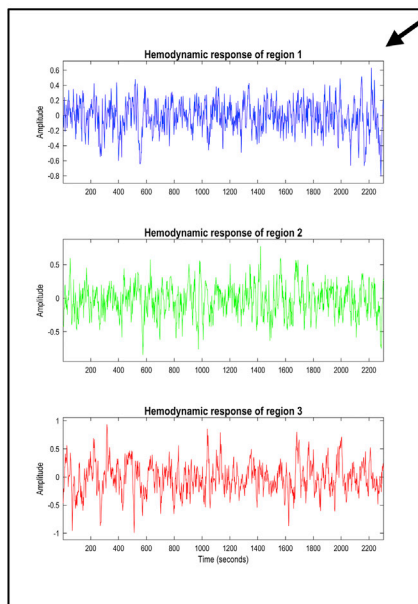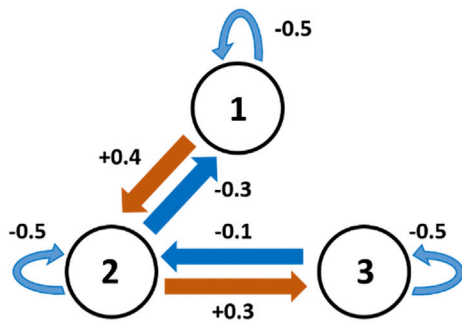
Fig. 4. *Synthetic data and parameter estimates (regular orbit group)*. This figure illustrates the synthetic timeseries and various connectivity estimates following the inversion of a hidden Markov model of state-dependent connectivity in fMRI. Panel A shows the causal network structure. In panel B, the first three plots show the simulated timeseries for a single subject. The first plot shows the endogenous fluctuations driving neuronal responses. The second plot (upper right) shows the neuronal responses for each of three nodes in the network (coloured lines) and the haemodynamic fluctuations they induce (cyan lines). These haemodynamic signals generate BOLD responses that, with observation noise, constitute the simulated signal – shown in the third and fourth plots. The three simulated regional BOLD signals have been separated in the inset figure, for clarity. The epochs (in blue) superimposed represent the (Hanning) windows applied to partition the timeseries. Each epoch was inverted using Bayesian parameter averages, from a standard spDCM inversion of the entire timeseries, as the prior expectation (with very precise shrinkage priors). A few selected connections were allowed to vary between epochs (by relaxing the shrinkage priors for these connections). In this example, the intrinsic (self-inhibition) connectivity parameters of each node were allowed to change from epoch to epoch. The posterior densities for the last epoch are shown in the fifth (third row) plot. The gray bars correspond to the posterior expectations and the pink bars denote the 90% Bayesian confidence intervals. The connectivity values used to simulate the data are shown as black bars. It is evident from these results that the overall profile of connectivity has been captured; however, there is a high degree of conditional uncertainty – as indicated by the confidence or Bayesian credible intervals. Note that the posterior estimates of other (conserved) connections are satisfactory and show a characteristic regression to the (prior) mean. The lower panels show the true connectivity parameters over all 18 epochs from both subjects. These nine parameters correspond to the adjacency (A) matrix coupling the three nodes. Note that only three connection parameters change, because we applied precise shrinkage priors to extrinsic (between-node) fluctuations. The middle panel reports the maximum *aposteriori* estimates before (middle) and after (right) applying empirical (PEB) priors from the hidden Markov model (HMM). The agreement between the final (PEB) estimates and the true pattern of state-dependent connection changes is not perfect but reasonable. These posterior estimates were based upon empirical priors afforded by an HMM, whose posterior expectations are shown in Fig. 6. This illustrative analysis shows how transitions among hidden or latent connectivity states can be estimated using PEB under the prior assumption of orbital transitions among a known number of states. However, we now have to establish whether this orbital model is the best explanation for the data, using Bayesian model comparison.
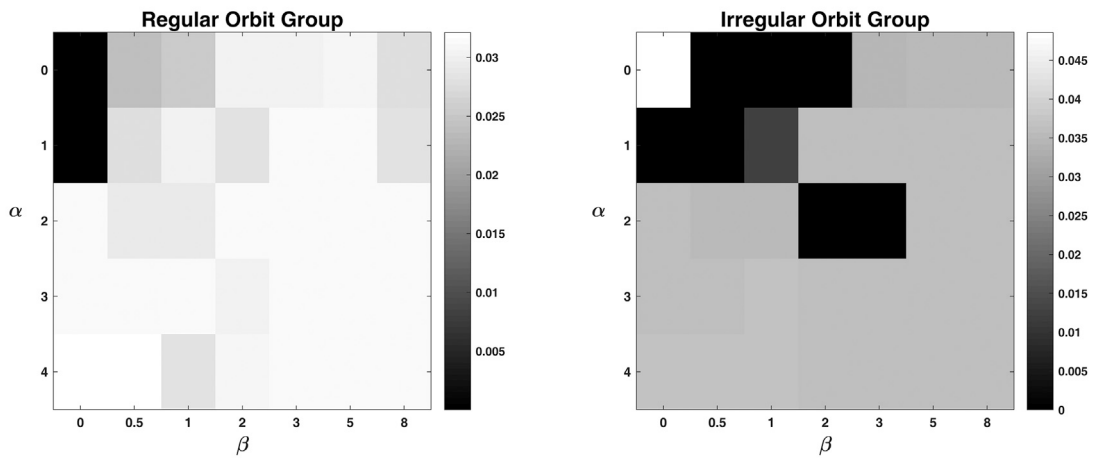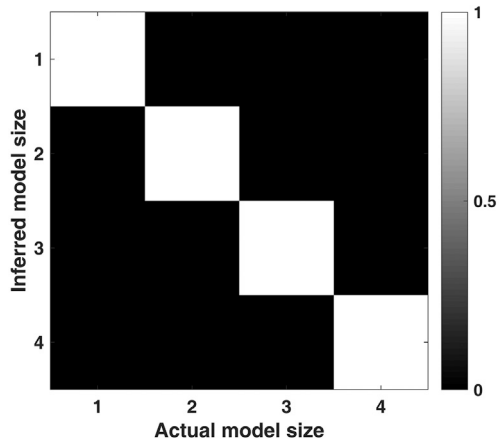
A. Regular orbit
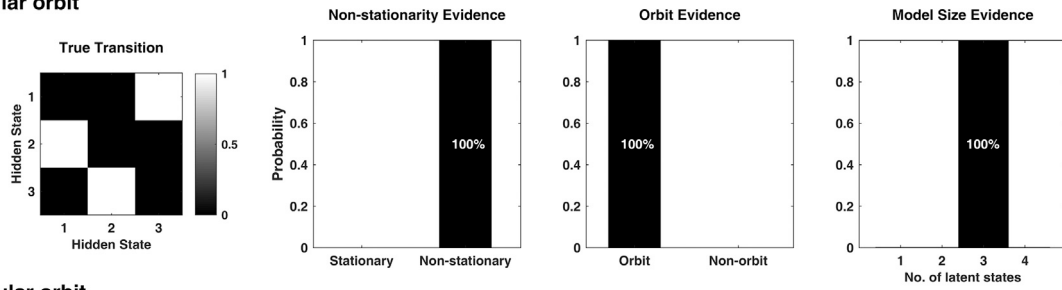
B. Irregular orbit
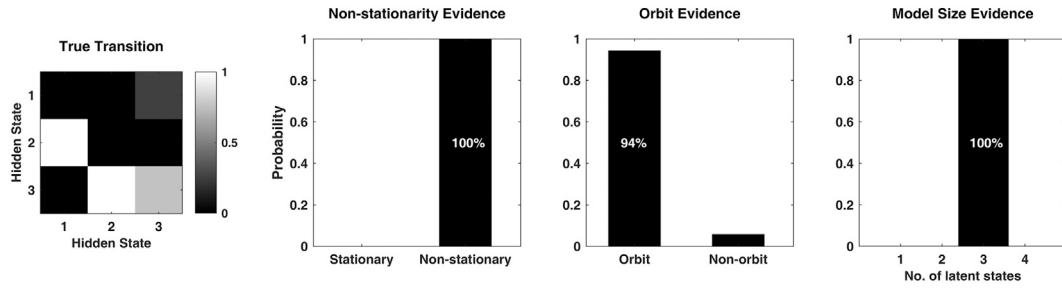
C. Marginal likelihood over hyperparameters (α, β), for n=3

D. Confusion matrix: marginal likelihood over the number of latent states (n)

(caption on next page)

Fig. 5. *Bayesian model comparison.* This figure shows how BMC can characterize the main attributes of the (simulated) data; that is, the presence of state transitions (i.e., non-stationarity), the optimal number of hidden states (i.e., model size), and the orbital nature of the transitions (i.e., orbit evidence). These were achieved by searching over a (prior) model space spanned by (n, α, β). After recording the free energy for each combination of these values ($n \in [1, 4]$, $\alpha \in [0, 4]$, $\beta \in [0, 8]$), the approximate log model evidences were passed through a softmax function to estimate the joint probability of all the combinations. Thereafter, marginalization over subsets of this space furnished the model likelihoods presented in the bar plots in the upper panels. Specifically, non-stationarity is the probability of models with $n > 1$, and was established for both groups with high posterior confidence. The evidence for orbits reflects the probability of models inverted under orbital priors (i.e. $\beta > 0$), as opposed to orbit-free priors (i.e. $\beta = 0$). As a result, the presence of the orbit has been identified confidently, for the regular orbit group – and BMC has (correctly) detected the existence of certain non-orbital attributes in the irregular orbit group. These marginal likelihoods were based on the marginal likelihoods conditioned upon each combination of (n, α, β), as evaluated by free energy. Panel C provides an example of the associated marginal likelihoods over the hyper parameters (α, β) for $n = 3$. Panel D marginalises over the hyperparameters to show the marginal likelihoods over the number of latent states n, in the form of a confusion matrix. This confusion matrix is based upon simulated data generated with different numbers of hidden or latent connectivity states (each column of the confusion matrix corresponds to a different dataset). Please see main text for details.

under the best model. In what follows, we illustrate these procedures using synthetic data. In subsequent work, we will apply the analyses described in this paper to empirical data to illustrate the sorts of questions that can be addressed.

3. Illustrative simulations

3.1. Simulation setup

In this section, we describe the setup for analysing synthetic data generated by transitions among hidden (Markovian) states. These simulations are presented to establish the face validity of the scheme; namely, to show it can recover structured fluctuations in connectivity when they are present. We considered two sorts of structured fluctuations; i.e. a *regular* orbit and an *irregular* orbit. For the former, we assume that the succession of connectivity states conforms to a SHC with a dwell time for each state that corresponds to epoch length.

The other form of state transition probability—which we call an irregular orbit—is essentially a slower SHC with a greater dwell time for a subset of the states. This sort of imbalance – in the period of a heteroclinic cycle – violates all three prior potential terms in Eq. (1). While this is obvious for the dwell prior, an increased dwell time induces uncertainty in moving to the next state from the current state; thereby increasing the entropy of the transition matrix. Moreover, this implicit transition uncertainty implies that the number of times each state is visited is not definitive for any sample trajectory or sequence. This affects the contribution of different states to the cycle and violates the equidistribution prior. In short, an irregular orbit possesses non-orbital attributes, which we hoped model comparison would identify. We will address the theoretical and empirical implications of these distinct transition patterns in the discussion.

In brief, we simulated fMRI timeseries for two pairs of subjects. This is the smallest number of subjects needed to illustrate how one can test for within and between group (i.e., pair) effects. Each subject's data comprised nine epochs of 128 scans with a TR of 2 s. The data were generated using scale-free random fluctuations for neuronal activity (with an exponent of one) and (region-specific) observation noise (with an exponent of one half). In these simulations, state-dependent connectivity was simply an increase in self-inhibition of about 50% in the first, second and third nodes of a hierarchically connected network or graph.

In the first subject pair, after each epoch of 128 scans, the connectivity switched to another (state dependent) pattern in a regular fashion, through three hidden states (i.e., 1,2,3,1,2,3, etc.). For the second pair, the same state-dependent connectivity profiles were used, but with irregular switching. Specifically, we simulated a slow orbit with increased dwell time in state 3. The state sequence generated from this transition (and used in the simulations) was: (1,2,3,3,3,1,2,3,3) for nine epochs. Predefined connectivity patterns (associated with each hidden state) were used to specify the corresponding spDCM kernels, which translate endogenous neuronal fluctuations into BOLD signals. Representative simulations of neuronal and hemodynamic series are provided in Fig. 4.

In summary, we simulated timeseries from two groups (i.e. pairs) of subjects with regular and irregular orbits. The differences between the two groups were subtle; they differed only in the form of itinerant

transitions between the same connectivity states. This difference was motivated by insights from dynamic functional connectivity analyses that are reviewed in the discussion. In what follows, we first focus on whether itinerant dynamics can be recovered at the within group level. We will then turn to between group comparisons, to see if we can identify differences between subjects with regular and irregular orbits.

3.2. Model inversion

Having generated fMRI data for all synthetic subjects, we performed model inversion to characterize the ensuing fluctuations in functional connectivity. In the middle left plot of Fig. 4, the superimposed epochs (in blue) represent the Hanning windows applied to epoch the timeseries. Each epoch from each subject was then inverted using Bayesian parameter averages (over subjects) following inversion of each subject's timeseries without epoching, as the prior expectation, with very precise shrinkage priors. A selected number of connections were allowed to vary between epochs (with uninformative shrinkage priors); namely, the intrinsic (self-inhibition) connectivity parameters of each node. The bottom middle panel of Fig. 4 reports the maximum aposteriori (MAP) estimates from this procedure (from the regular orbit subjects), before applying empirical (PEB) priors from the hidden Markov model (HMM).

The improvement in the posterior estimates (in relation to the values used to generate the data) afforded by the use of empirical priors from the HMM is self-evident. In other words, knowing that the states were generated by orbital transitions among latent states of connectivity allows us to recover the states actually generating the timeseries. In this example, the PEB scheme assumed (correctly) the hyperparameters were $(\alpha, \beta) = (2, 8)$.

Fig. 4 shows that one can recover the dynamical architecture of (connectivity) state transitions from the spectral density features of observed data. However, to do this, we needed to know the number of states and hyperparameters generating heteroclinic cycles; i.e., (n, α, β). In an empirical setting, one would not know these hyperparameters. In what follows, we consider a range of values for the number of latent states and the hyperparameters, such that each (n, α, β) combination furnishes a (prior) model—with an associated model evidence. One can then use model comparison to establish the existence of state transitions (i.e. non-stationarity), the number of latent states, and the strength of orbital attributes.

3.3. Bayesian model comparison

This section illustrates the of use Bayesian model comparison to evaluate the evidence for different models; namely, models of regular or irregular orbital transitions among an unknown number of hidden or latent connectivity states. In brief, we repeated the above inversion under different priors to evaluate the evidence for each prior assumption. In our case, the priors come into flavours; first, the number of connectivity states and, second, the strength of prior beliefs about the form of state transitions, as parameterised by the hyperparameters (α, β) in Equation (1).

To recap, the presence of state transitions requires $n > 1$. If $n > 1$, we can infer the presence of orbital transitions when $\beta > 0$. This follows
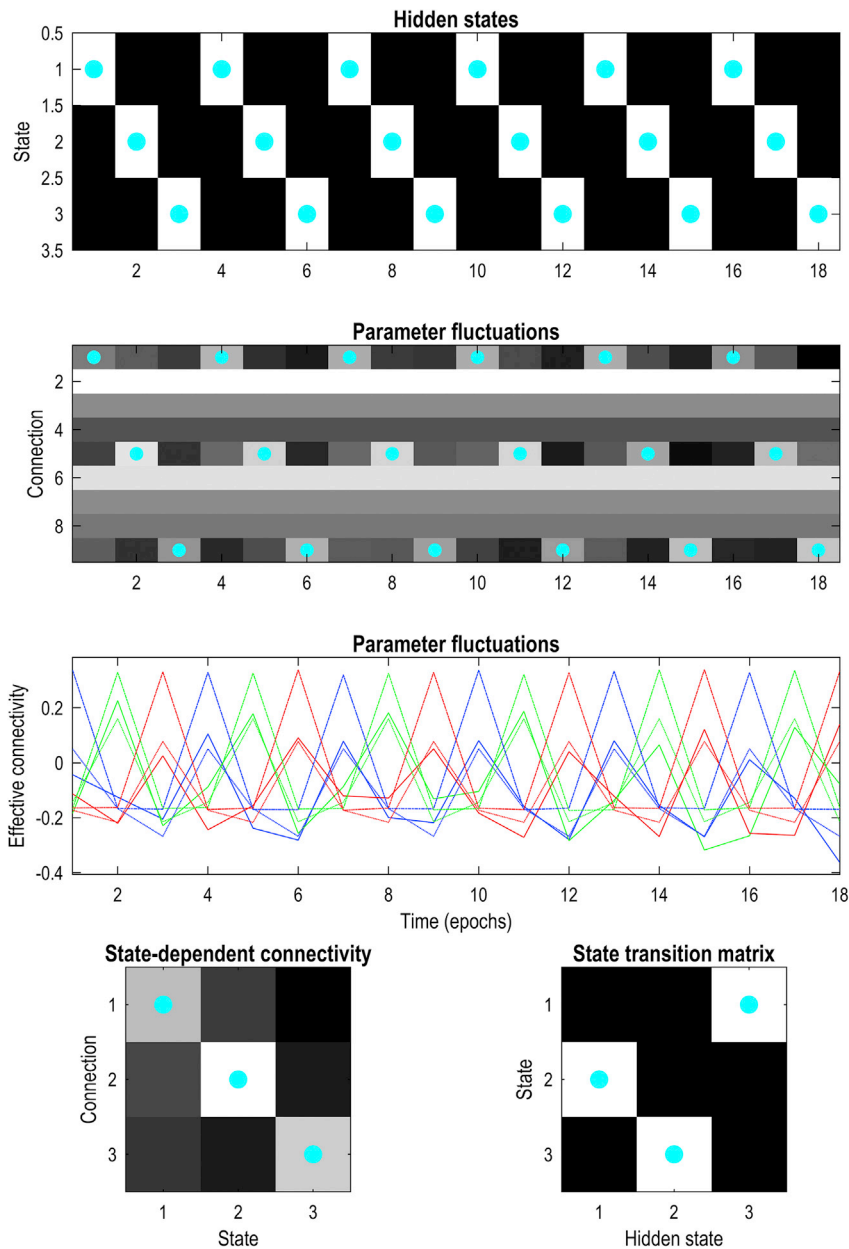
Fig. 6. *Hidden Markov modelling of fMRI data (regular orbit group).* This figure summarises the posterior expectations about hidden Markovian states (top panel) and associated fluctuations in connectivity parameters (shown in the middle two panels). The lower panels report the state-dependent connectivity patterns associated with each state (left) and the expected probability transition matrix describing transitions among those states (right). The cyan dots correspond to the true values of the hidden states (or the largest connection strengths, where appropriate). These results are shown for all 18 epochs (nine epochs from two subjects). With these simulated data, one can be nearly a hundred percent certain about the state of the (synthetic) brain at each epoch, with a faithful reproduction of the regular orbit through the three hidden states. In this example, the associated patterns of connectivity have been correctly identified in terms of their qualitative or relative sizes. The true values are shown as a function of time (epochs) in the third panel as broken lines. The posterior estimates of these parameter fluctuations are shown as dotted lines and their empirical prior expectation as a solid line. In this example, the posterior (within-epoch) estimates are remarkably similar to the empirical prior (between-epoch) estimates. This reflects the fact that the conditional uncertainty from the first level was sufficiently high to enable the empirical priors to dominate the posterior estimate. The true state-dependent changes in connectivity involved increases along the leading diagonal that have been roughly recovered in the posterior estimates of state-dependent connectivity (lower left). The posterior expectation of state transitions is correct and precise, in virtue of having six instances of each transition. These results were taken from a hidden Markov model inversion, assuming three hidden states. This was the model with the greatest evidence over all models tested; as illustrated in Fig. 5.

because when $\beta = 0$, the priors over the key terms in Equation (1) disappear. This means that we can search over a model space that spans different values of (n, α, β) and record the free energy for each combination of values (using different initialisations when necessary or Bayesian model reduction, when appropriate). By applying a softmax function to the resulting approximations to log-evidence, one can estimate the joint probability of all combinations of hyperparameters (i.e. $P(n, \alpha, \beta)$) and marginalise over subsets.

In what follows, we will compare the probability of models with $n = 1$ and $n > 1$ to establish the existence of transitions among connectivity states. We can then compare models with $\beta = 0$ and $\beta > 0$ to infer the presence of orbits; in both cases marginalizing over α. Specifically, the evidence for an orbit; i.e. $P_{orbit} = 1 - P_{nonorbit}$ was evaluated using $P_{nonorbit} = \sum_n \sum_{\alpha \leq 1} \sum_{\beta = 0} P(n, \alpha, \beta)$. The results of these model comparisons are shown in Fig. 5.

Effectively, a high orbit evidence indicates that the data were generated by a cycle through connectivity states. For example, our simulated regular orbit data favours the orbit hypothesis (as indicated by

100% orbit evidence, in Fig. 5A); hence, the evidence for these data was greatest when inverted with orbital priors, specifically $(\alpha, \beta) = (2, 8)$, using the correct number of latent states (i.e. n = 3 with 100% confidence). The ensuing posteriors are shown in Fig. 6. Conversely, for data generated by the irregular orbit, the evidence for orbital hyperparameters change profoundly (see Fig. 5C) and the posterior probability of an orbit fell to 94% (see Fig. 5B).

To further illustrate this kind of Bayesian model comparison, we repeated the above analysis using different numbers of states, to see if we could recover the number used to generate the data. In brief, we generated data with regular orbits among different numbers of hidden states; i.e., n = {1,2,3,4}. and evaluated the marginal posterior over n and for each simulated dataset. This allows one to construct a posterior *confusion matrix*, where each column of the matrix corresponds to the posterior (marginal) probability over the number of states; i.e. $P_n = \sum_\alpha \sum_\beta P(n, \alpha, \beta)$[4]. Ideally, this confusion matrix will concentrate its

---

[4] $P_n = \sum_\alpha \sum_{\beta \gg 0} P(n, \alpha, \beta)$ for noisier data (from shorter epochs).

posterior probability mass around the leading diagonal. In other words, the posterior probability should be greatest over the number of states used to generate each timeseries. Fig. 5D shows the results of this analysis and suggests that one can identify, precisely, the number of latent connectivity states using this scheme – and simulation setup. These results are presented to illustrate the use of confusion matrices, which can be useful when trying to establish how levels of noise (or the length of the timeseries) affects model identifiability.

The first column of the confusion matrix (Fig. 5D) has an important interpretation: in the absence of connectivity dynamics, model evidence correctly and precisely indicates that only $n = 1$ connectivity state is necessary to explain the data; in other words, there is no dynamic functional connectivity. Previous studies have highlighted the need for inference methods that can detect true (non)stationarity in FC fluctuations (Hindriks et al., 2015; Leonardi and Ville, 2015; Zalesky and Breakspear, 2015). However, the null-hypothesis is usually assessed (using surrogate data) by testing for nonstationarity in the time series per se—rather than changes in connectivity (Chang and Glover, 2010; Hindriks et al., 2015; Zalesky et al., 2014). The current Bayesian treatment provides explicit evidence for models of stationary ($n = 1$) versus non-stationary ($n > 1$) effective connectivity. This kind of (Bayesian) inference would also apply to the associated functional connectivity, since dynamic EC provides a full account of DFC (as noted in the introduction).

In summary, one can use the free energy bound on model evidence to optimise the parameters and hyperparameters of the model – to identify the best explanation for how the data were generated. In the foregoing, Bayesian model comparison was used to identify the number of connectivity states (and hyperparameters) generating the itinerancy of connectivity state transitions. In the next section, we illustrate how Bayesian model comparison can be used to identify *differences* in EC dynamics between subjects. In brief, this involves comparing models that allow for differences in parameters between groups with models that do not.

3.4. Comparing groups of subjects

In this section, we turn to the analysis of between subject effects; here, testing for differences between the two pairs of subjects of the simulations above. We first tested for any within-group differences (between the two subjects, in each group). This form of Bayesian model comparison can be implemented quickly and efficiently by inverting the hidden Markov model for all epochs from both subjects and comparing the ensuing log evidence with the sum of log evidences when each subject is inverted separately. The difference provides the evidence for a model in which the state transitions and associated state-dependent connectivity is allowed to differ between the two groups (or subjects). In this example, we used the same parameters (and hyperparameters) for both subjects in each pair and, unsurprisingly, found strong evidence for the model with no differences. The posterior estimates of state-dependent connections are shown when estimated for the two subjects separately on the lower right of Fig. 7. Although the absolute values differ, they have a similar profile. Similar results were obtained for the irregular orbit group (i.e. strong evidence for null within-group difference). The accompanying estimates of state transitions are shown in Fig. 8.

We then tested for a difference between the two groups (that is, regular versus irregular orbit group) using the same approach. The evidence for parametric differences between the two groups was positive, but not strong (log Bayes factor $\approx 2.9 < 3$) (Kass and Raftery, 1995). This inconclusive evidence may reflect the fact that the two simulated groups have identical state-dependent connectivity parameters and their transition patterns did not differ substantially. Note that we are not simply rejecting the null hypothesis of no difference. We are fairly confident that we do not know, given these data, whether there are differences or not. In other words, we can assert that these data do not allow us to disambiguate confidently between the two models or hypotheses. This begs the
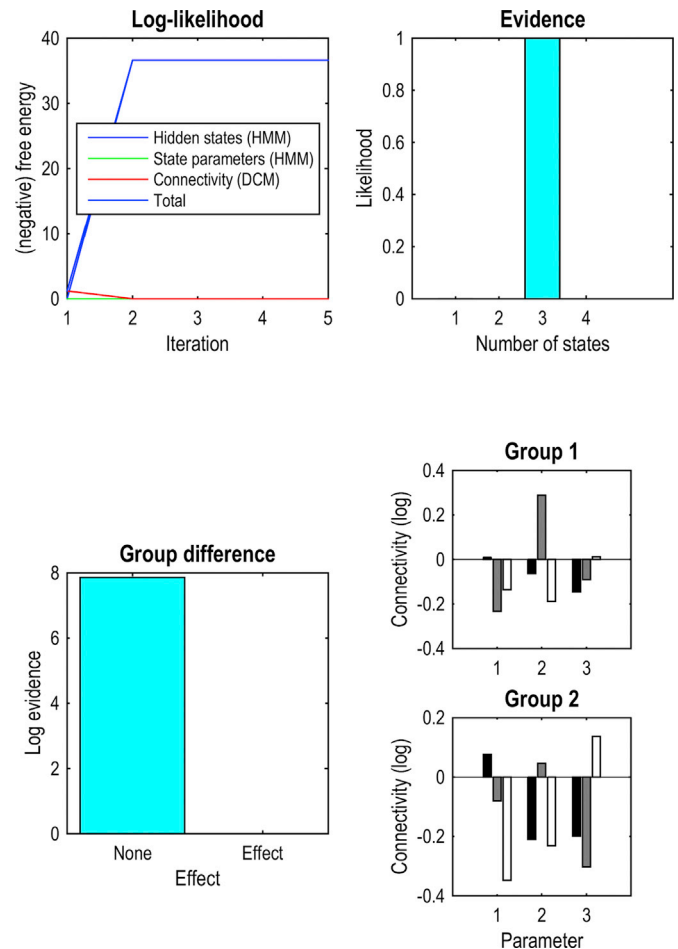

Fig. 7. *Bayesian model comparison of hidden Markov models (regular orbit group).* This figure illustrates the key aspects of Bayesian model inversion and comparison. The upper left panel shows the contributions to log evidence in terms of log likelihoods during five iterations of the HMM inversion, for a model with three hidden states. The coloured lines show the relative contribution to the total log evidence (i.e., free energy) shown as a cyan line. The blue line corresponds to the log likelihood afforded by posterior beliefs about hidden states. This increase in log likelihood more than compensates for the increasing complexity at the level of each DCM (shown as a red line). In this example, the contribution from the complexity due to parameters of the probability transition matrix (green line) makes an exceedingly small contribution. Note that the scheme converges in a handful of iterations. The total free energy after convergence was evaluated for hidden Markov models with 1, 2, 3 and four hidden states. The resulting posterior probability over the four models is shown on the upper right. In this case, we can be almost certain that there are three hidden states in play. The lower two panels illustrate model comparison in terms of testing for group differences – in this case the differences between the two simulated subjects. This form of Bayesian model comparison can be implemented quickly and efficiently by inverting the hidden Markov model by concatenating all epochs from both subjects and comparing the ensuing log evidence with the sum of log evidences when each subject data is inverted separately. The difference provides the evidence for a model in which the state transitions and associated state-dependent connectivity is allowed to differ between subjects. In this example, we generated data using the same parameters for both subjects and, unsurprisingly, find strong evidence for the model with no differences. The posterior estimates of state-dependent connections are shown when estimated for the two subjects separately on the lower right. Although the absolute values differ, they have a similar profile. Similar results were acquired for the irregular orbit group (i.e., quick convergence, correct model size and strong evidence for no differences). In the next section, we revisit the simulations above, while addressing some practical issues concerning network size and key data features.
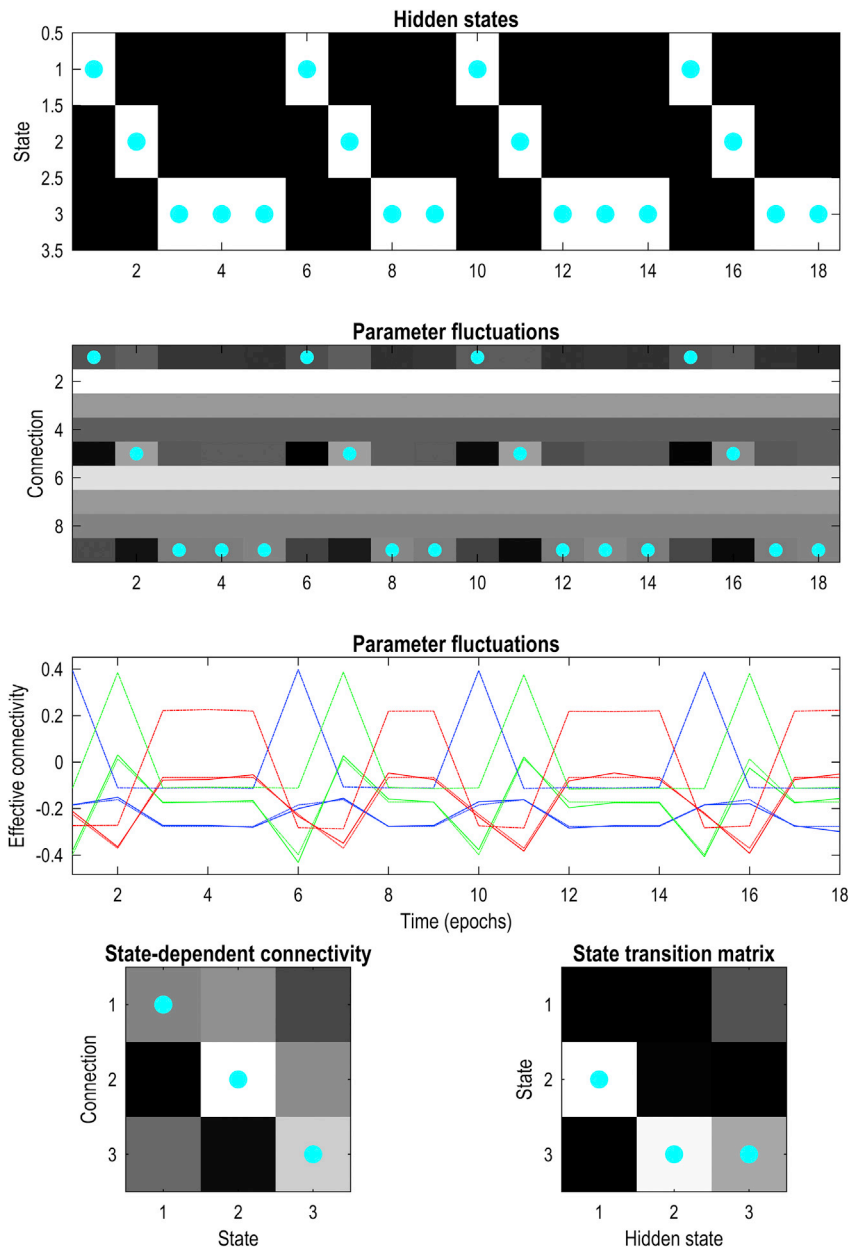
T.S. Zarghami, K.J. Friston

Fig. 8. *Hidden Markov modelling of fMRI data (irregular orbit group).* This figure summarises the posterior expectations of the parameters, after model inversion using data from the second simulated group. The cyan dots correspond to the true values of the hidden states (or the largest connection strengths where appropriate). The hidden state posteriors and the transition matrix are both accurate and precise. The parameter fluctuations track the simulated values reasonably well. Furthermore, the state-dependent connectivity parameters correctly reflect the dominant connection in each state, with moderate specificity. This (state to connection) mapping could potentially be improved by including more subjects in the analysis, furnishing more precise empirical priors for the DCMs. Please refer to the caption of Fig. 6 for additional details.

pragmatic question: how many subjects would be needed to make a definitive statement about group differences?

To show how this sort of question can be answered quantitatively, we repeated the above analyses using two, four and eight subjects. Fig. 9 shows the results in terms of differences in log evidence, as a function of the number of subjects for each group. The key thing to observe is that the difference in log evidence or log Bayes factor becomes more definitive (i.e., diverges from zero) as the number of subjects increases.

In short, one can use simulated data and Bayesian model comparison to assess the number of subjects or length of time series that might be needed to address key hypotheses about group effects. In practice, group differences might arise from divergence in the state transitions, the connectivity profiles, or both—which can be further explored using Bayesian model comparison and inspection of the corresponding posteriors.

3.5. More realistic simulations

So far, we have introduced a hierarchical generative model of dynamic connectivity, based on the notion of stable heteroclinic cycles; elaborated the accompanying variational inference scheme; outlined Bayesian model comparison for characterizing itinerant dynamics and have described a procedure for analyzing between-subject effects. We also presented illustrative simulations to show how these methods apply to neuroimaging data analysis—in this case, resting state fMRI signals.

These simulations were based on a simple causal architecture comprising three nodes (Fig. 4). These nodes had dynamic inhibitory self-connections that served to switch the EC mode (every 128 time points) according to cyclic (i.e., orbital) dynamics. The results suggest that the inference scheme and ensuing BMC can establish non-stationarity and orbital behavior (or lack thereof) and disambiguate regular from irregular orbits. It was further demonstrated that variational model inversion can characterize the dataset in terms of the posterior expectations of Markovian transitions among the (optimal number of) latent states, the parametric EC modes and the structured epoch to epoch fluctuations of the dynamic connections. This simulation serves to formalise the conceptual and methodological fundaments of the approach. Now we apply the same procedures to a more sophisticated setup—of the sort
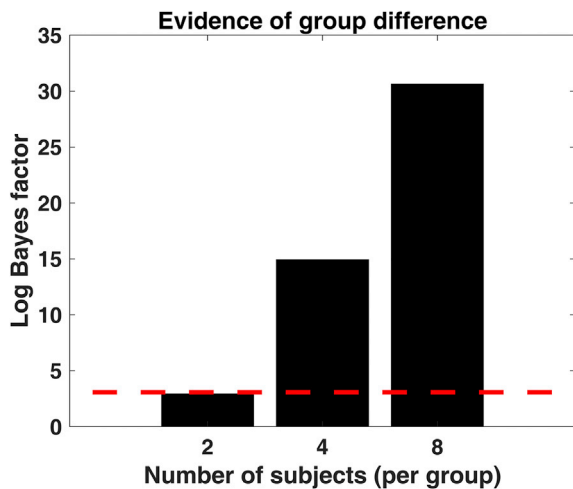
Fig. 9. *Evidence for group differences, as a function of subject number.* This graph shows the change in the log Bayes factor (i.e., differences in log evidence as scored by variational free energy) when comparing models with and without group specific differences. Evidence for group differences emerges with an increasing number of subjects. By convention, a log evidence difference of about three (dotted line) is usually considered strong evidence for one model over another.

encountered in practice.

Specifically, we extend the neuronal architecture to a network of five nodes[5] and show that Bayesian model inversion and comparison can still detect and characterize orbits. We will also simulate faster orbits (with shorter minimum dwell time) and investigate how well they can be identified using shorter epochs. Moreover, a model comparison procedure is shown to reveal the intrinsic timescale of the underlying orbital dynamics (by finding the optimum epoch length for model inversion).

3.5.1. Simulation setup

We first demonstrate the effect of scaling up the network; i.e. increasing the number of nodes or regions. We embedded the causal architecture and itinerant dynamics of the previous simulation (Fig. 4A) in an extended network of five regions (Fig. 10A). We do not presume transient connectivity dynamics on these new nodes, but rather let them influence the original regions, via excitatory and inhibitory (extrinsic) connections. The rest of the setup is similar to the regular orbit example presented before.

In brief, we simulated sessions in two subjects each comprising 9 epochs; where each epoch comprises T time samples with TR = 2 s. The 3 state-dependent connectivity modes pertain to a 50% increase in the inhibitory self-connections of the first three regions (one at a time) with orbital Markovian transitions. The data were generated using scale-free random fluctuations for neuronal activity (with an exponent of one) and region-specific observation noise (with an exponent of one half), using the generative model of spectral DCM for fMRI (Friston et al., 2014a). The network is depicted schematically in Fig. 10A.

3.6. Characterizing slow transitions

Initially, to focus on the role of network size, we maintained T at 128 time samples (per epoch). This translates to an orbit that lingers in the vicinity of a metastable state (i.e., EC mode) for about 4 min. Notably, with fMRI data from faster acquisitions,[6] the same number of samples can

---

correspond to an epoch of around 90 s. Such (relatively long) window sizes are appropriate for tracking slow connectivity dynamics (Chang and Glover, 2010; Handwerker et al., 2012; Leonardi and Ville, 2015).

Having generated the data, we followed the same Bayesian model inversion and comparison procedure described above. As previously, we assumed[7] that we know the best epoch length for dividing the data into shorter intervals (to which spectral DCM's are fitted): i.e., 9 epochs per session. The model evidence was recorded over a plausible range of hyperparameters and model sizes to furnish the joint probability distribution over the searched space; i.e., $P(n,\alpha,\beta)$, for $n \in [1,4]$, $\alpha \in [0,4]$ and $\beta \in [0,8]$. Thereafter, Bayesian model comparison characterized the main attributes of the connectivity orbits by marginalizing $P(n,\alpha,\beta)$ over subsets of the hyperparameters (α, β) and the model size (n)—as described above and in Fig. 5.

The results are summarised in Fig. 10. It is evident that the main attributes (of non-stationarity, orbitiness and model size) have been recovered perfectly (Fig. 10B). The posterior expectations of the parameters (i.e., the state sequence, transition pattern and epoch to epoch variations in the self-connections) are also consistent with the values used to generate the data. The between-subject analysis also returns strong evidence (>3 nats) for no difference between the two simulated subjects.

Note that the number of the estimated DCM connections has almost doubled in this scenario (see Fig. 10C top right plot, compared to Fig. 4). However, these simulations show that adding a few regions to the network has not affected the performance of the scheme when detecting and characterizing slow orbital dynamics. The reason is twofold: (1) the number of the dynamic connections did not increase; and (2) the **itinerant priors** that we use for parametrizing the (prior) model space are particularly efficient when detecting SHC-like behavior—despite the ill-posed nature of the inference problem. Next, we consider how robustly faster orbital dynamics can be inferred from the spectral data features, generated within the same network, using shorter epochs.
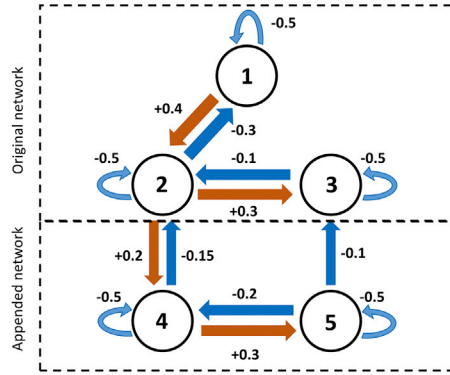
3.7. Characterizing faster transitions

We repeated the above experiment, but this time decreasing T (the number of samples per epoch) used for generating and inverting the non-stationary data. That is, we simulate faster dynamics and use correspondingly shorter epochs to track them. Specifically, four additional datasets were generated with $T \in \{96, 80, 64, 48\}$ samples per epoch. This decrease in epoch length speaks to SHC dynamics with shorter minimum passage time around the saddles (i.e., metastable EC modes). In other words, T corresponds to the smallest dwell time in a specific EC pattern, before switching to the next. The rest of the settings were identical to those used in the previous section.

The model inversion and subsequent BMC results show that the SHC dynamics of the simulated data is reflected clearly in the marginal likelihoods (Fig. 11). That is, the presence of non-stationarity, orbital itinerancy and the number of latent states are all correctly inferred—despite the extended network structure and the shorter epochs. A closer inspection of the parameter posteriors (Figs. 12 and 13) reveals reliable inference about the sequence of state occupations, transitions and connection variations, in all cases. Evidence for (null) group difference is either strong or positive. We showed earlier how the evidence for between-subject effects could be augmented by including more subjects (Fig. 9). Of note, only when T falls to 48 samples per epoch, the parameter fluctuations become smeared over adjacent epochs, compromising the temporal specificity of inferred states. In short, these simulations illustrate the operational range of dynamic EC analysis for inspecting orbital dynamics of different rates, when applied to idealized fMRI data.

---

[5] Networks of 5–6 nodes are prevalent in spectral DCM studies with resting state fMRI (Friston et al., 2014a).

[6] Such as those available from the Human Connectome Project (HCP) with TR = 0.7 s.

[7] We will demonstrate how the best epoch size can be inferred from the data below.

T.S. Zarghami, K.J. Friston
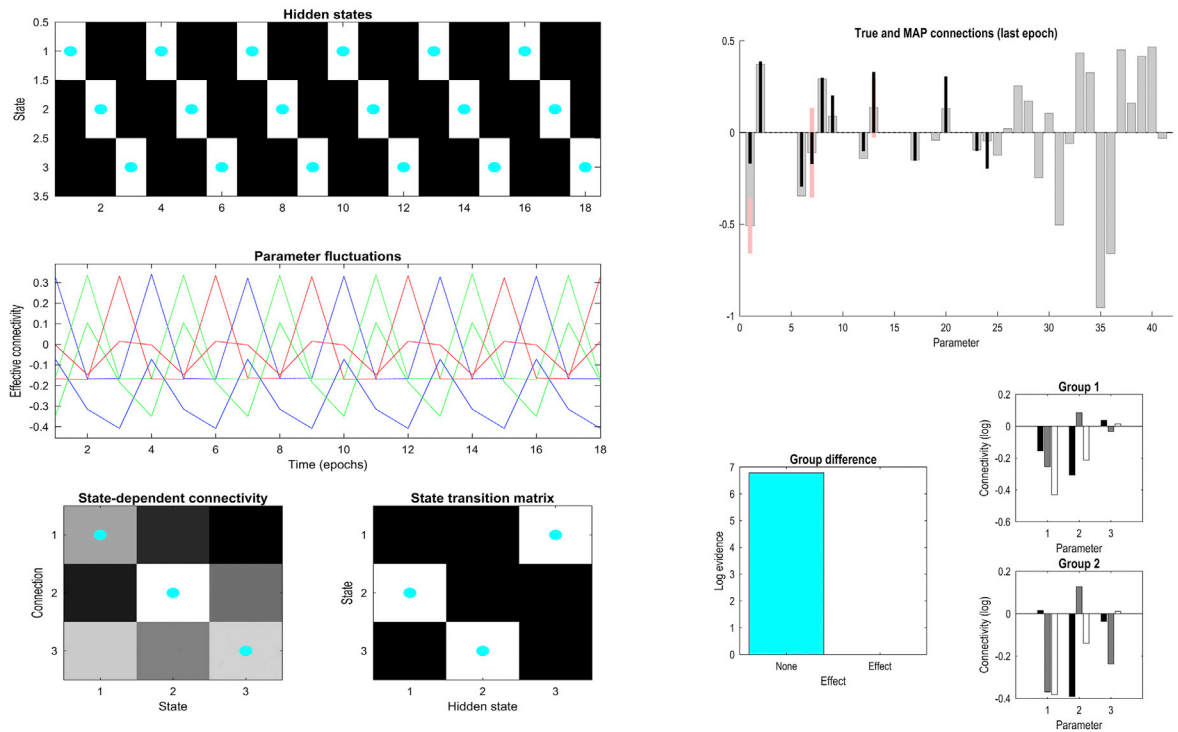
A. Extended network

B. Model evidence, for T=128 samples per epoch
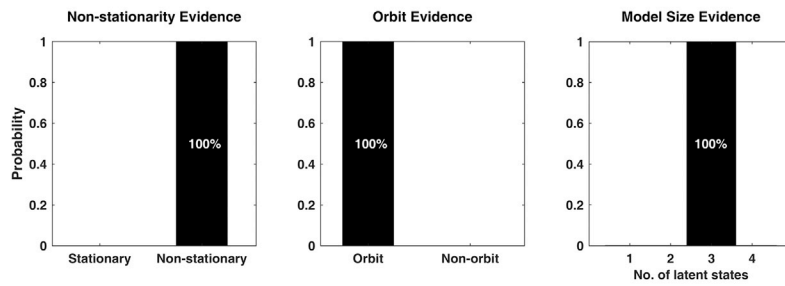
C. Parameter posteriors, for T=128 samples per epoch

(caption on next page)

Fig. 10. *Inference over an extended network.* This figure shows (A) a network of five regions, where the first three self-connections fluctuate over epochs. State-dependent connectivity patterns were constructed by inducing 50% increase in the self-inhibitions of regions 1–3 (one at a time) to generate 3 states in the parametric space of effective connectivity. Sessions from two subjects were simulated, each comprising 9 epochs, with 128 time samples per epoch and TR = 2 s. The connectivity pattern changed from epoch to epoch, traversing a regular orbit through the 3 parametric states. The resultant timeseries were divided into sub-intervals using 9 Hanning windows and a spectral DCM was inverted for each epoch (for details of this inversion using BPA and shrinkage priors, please refer to Fig. 4 and the main text). To characterize itinerant connectivity, we explored a (prior) model space spanned by the model size and the hyperparameters, i.e. (n, α, β) and recorded the associated approximate log model evidences (i.e., free energies); these values were softmaxed to yield the joint probability distribution over the search space; i.e. $P(n, \alpha, \beta)$, for $n \in [1, 4]$, $\alpha \in [0, 4]$ and $\beta \in [0, 8]$; (B) Bayesian model comparison identified the main attributes of the connectivity fluctuations by marginalizing $P(n, \alpha, \beta)$ over subsets of the hyperparameters (α, β) and the model size (n), as described in Fig. 5. In this example, the marginal likelihoods accurately and precisely identify the underlying nonstationary connectivity dynamics, the existence of an orbit and the correct model size (of three latent states); (C) posterior expectations of the key model parameters show perfect recovery of the state occupation sequence and the orbital transition pattern, as well as strong evidence for no difference between the two subjects (see Fig. 7 for the group analysis procedure). The posterior expectations of the parameters from a representative (final) epoch are plotted in panel C (top right) in gray bars; the superimposed black bars are the connection values used to generate data for that epoch, and the pink bars denote 90% confidence intervals. The posterior expectations of the dynamic self-connections (in panel C, middle left) vary in accord with the simulated variations. Following previous plots, the cyan dots correspond to the true values of the hidden states (or the largest connection strengths where appropriate).

Clearly, the epoch length limits the detectable rate of transitions and affects the quality of model inversion. In fact, the choice of window size has been long debated in DFC analyses (Hutchison et al., 2013; Leonardi and Ville, 2015). This is largely because DFC interpretations can be confounded by spurious FC fluctuations induced by very short windows. Hence, different rules of thumb have been proposed to guide exploratory DFC analyses for choosing (fixed or adaptive) window lengths – and assessing the FC fluctuations that they capture (Chang and Glover, 2010; Handwerker et al., 2012; Hindriks et al., 2015; Keilholz et al., 2013; Leonardi and Ville, 2015; Sakoğlu et al., 2010; Zalesky and Breakspear, 2015). Importantly, with a biophysically plausible model of effective connectivity (in particular DCM), the observation function (i.e. the neurovascular coupling) and the observation noise are modelled explicitly, hence separating observable fluctuations from the dynamics of their causes. But inferring these causes efficiently still requires sufficient data.

Aside from the signal processing, statistical and optimisation reservations that attend the choice of window size, a more fundamental issue pertains to the relevant timescale(s) of the underlying dynamics. To date, the most neurobiologically relevant timescales of connectivity fluctuations are still unclear and debated in neuroscience (Hutchison et al., 2013; Preti et al., 2017). In other words, identifying the most *functionally informative* epoch size for disclosing the dynamics of neuronal interactions is not trivial. This brings us to the next natural question: in an empirical setting, how can dynamic EC analysis reveal (and be informed by) the intrinsic timescale(s) of the underlying itinerant dynamics? The next section tries to address this question.

3.8. Optimum epoch length

The functionally relevant timescales of connectivity fluctuations (particularly at rest) are largely unknown – and an arbitrary window length (T) could miss important dynamics. Hence, it would be useful to establish the most plausible epoch length, for a given set of data, from which the presumed SHC dynamics of EC modes can be efficiently inferred. In this section, we present a realistic scenario in which one is agnostic to both model size and minimum dwell period of the underlying EC dynamics. We use this example to show how model evidence serves to guide the choice of window size, such that it discloses the most relevant orbital dynamics, when present.

The simulation was based on the extended network in Fig. 10, comprising 5 regions with three fluctuating self-connections (on regions 1–3), associated with 3 EC modes. The orbital transitions caused the EC pattern to change after every epoch of $T = 80$ time samples (with TR = 2 s). Data for each subject were generated from 9 such consecutive epochs. Here, two subjects were simulated with the same changes in connectivity.

These data were inverted under a model space spanned by a range of epoch numbers (w) and model sizes (n), assuming orbital dynamics. Specifically, the model space was defined by (w, n) pairs, with $w \in \{6 : 12\}$ and $n \in \{1 : 4\}$. The associated epoch lengths (for model inversion) were in the range of $T = 9 \times 80/w \in$

$\{120, 102, 90, 80, 72, 64, 60\}$ time samples; noting that epoch length decreases with epoch number for a session of the same generation. Variational inversion furnished approximate model evidences (i.e., free energies) for different combinations; i.e. $F(w, n)$. To infer the optimal number of latent states, we pooled evidence for each model size (n) over all window lengths and picked the model size with maximum posterior probability, (assuming *a priori* equally likely models):

$$P(n) = softmax\left(\sum_w F(w, n)\right) \tag{2}$$

$$n_{opt} = \arg\max_n P(n)$$

with the optimal model size (n_{opt}) at hand, the outstanding task was to infer the optimal epoch length. One can score different epoch lengths by examining the relative log evidence under the best model size (n_{opt}) compared to the stationary model (n = 1), as a function of epoch length.[8] This is simply the log Bayes factor (BF) or log odds ratio (Kass and Raftery, 1995). The rationale here is that the best epoch length reveals the greatest evidence for orbits. In short:

$$Log\,BF(w) = \log\frac{P(w|n = n_{opt})}{P(w|n = 1)} = F(w|n = n_{opt}) - F(w|n = 1) \tag{3}$$

$$w_{opt} = \arg\max_w Log\,BF(w)$$

The outcome of this procedure is demonstrated in Fig. 14. On the simulated dataset, the best model size has been inferred accurately (96% evidence in favor of n = 3) and the log odds ratio points to w = 9 windows as the best choice, which coincides with the window length used to generate the data.

These simulations were performed on a typically encountered network size (of 5 nodes) with relatively fast orbital dynamics. Adopting the inferred model size and the best epoch length furnishes the most relevant dynamic EC analysis and reveals the intrinsic timescale of the underlying dynamics, assuming orbital itinerancy. The posteriors from this EC analysis (assuming 9 epochs) are illustrated in the previous section (Fig. 12B). Importantly, the model size inferred herein (Fig. 14) did not assume knowledge of the best window size, in contrast to the previous examples. Instead, both the model size and epoch length were inferred from the data using a greedy search, as described above.

4. Discussion

Generative models of coupled neuronal systems – that can explain

---

[8] Note that this eschews the problem of directly comparing the evidence for different epoch lengths, which would entail comparing the evidence for different datasets.

A. Parameter posteriors, for T=96 samples per epoch
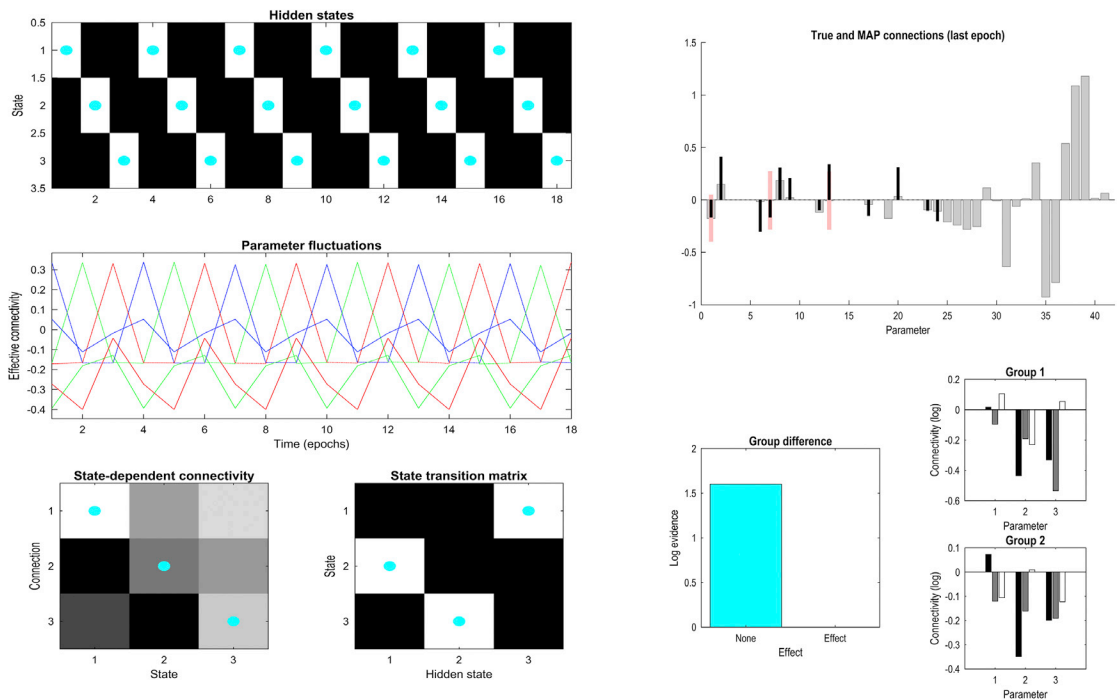
B. Parameter posteriors, for T=80 samples per epoch
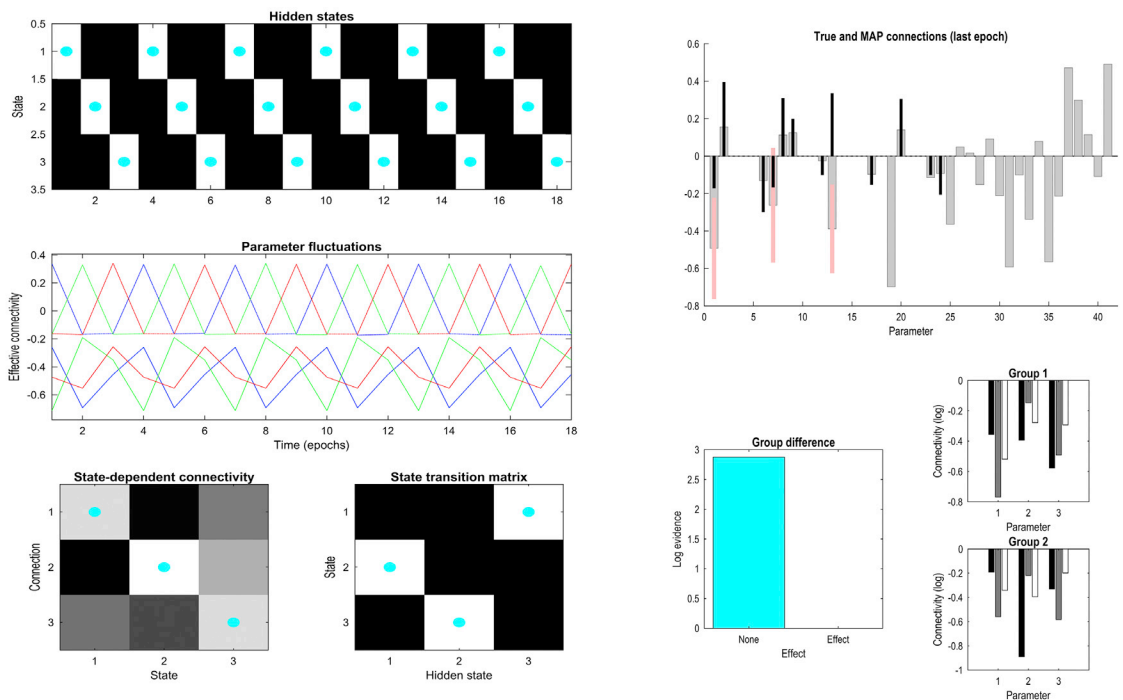
Fig. 11. *Bayesian model comparison results for the extended network with faster dynamics.* This figure shows the BMC results with different transition rates. Each dataset was generated using an orbital pattern of transitions among three latent states (i.e., EC modes), based on the network in Fig. 10. Notably, the connectivity pattern switched after a designated period of 128, 96, 80, 64 or 48 samples. The horizontal axis values reflect both the time (in samples) over which effective connectivity remains stationary, and the window size used to divide the data into epochs. The evidence is conclusive about the non-stationary orbital nature of the data and reliably indicates the correct size of three, used to generate these transient dynamics. We will next inspect the posterior expectations of the model parameters in each case (in Figs. 12 and 13).

A. Parameter posteriors, for T=64 samples per epoch
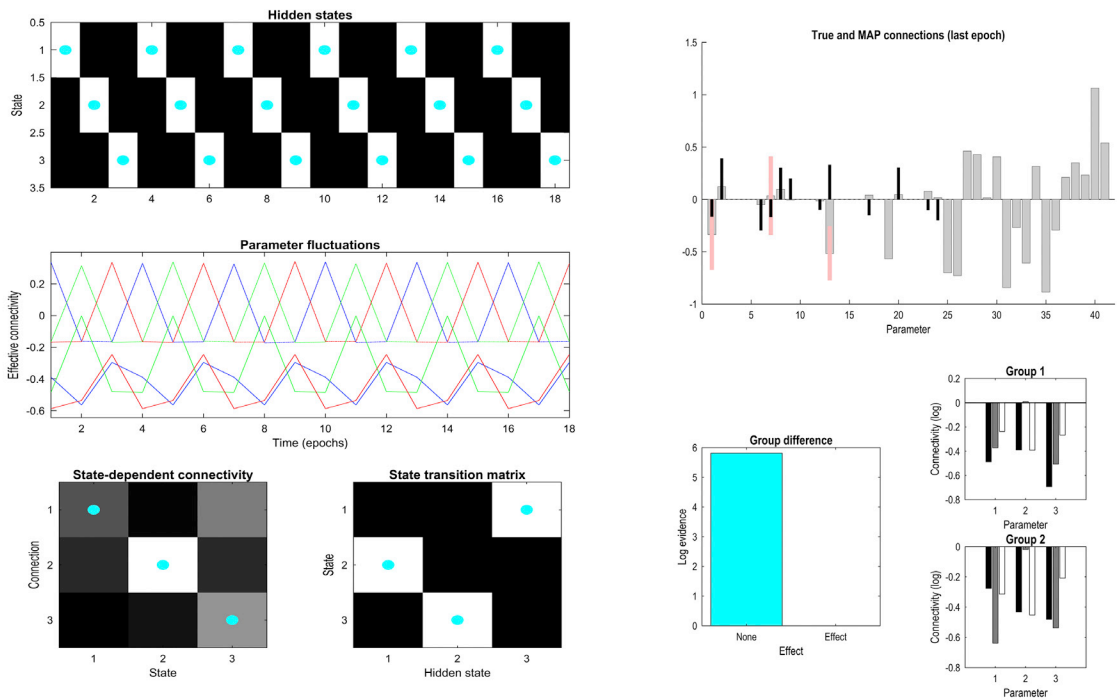
B. Parameter posteriors, for T=48 samples per epoch

Fig. 12. *Posterior expectations of the model parameters inferred from the extended network with faster dynamics.* Supplementing the model comparison results in Fig. 11, this figure shows the performance of model fitting (in terms of the posterior expectation of the parameters) as a function of epoch length, which denotes both the minimum dwell time of the orbital dynamics and the window length used to divide the data into epochs. (A) Using epochs containing T = 96 samples, the parametric itinerant traits (i.e., the state sequence and transition profile) are correctly inferred and the evidence for null group difference (between two subjects) is positive. (B) For T = 80 samples per epoch, the key parameter posteriors are recovered fully and the connectivity fluctuations closely follow the simulated changes.
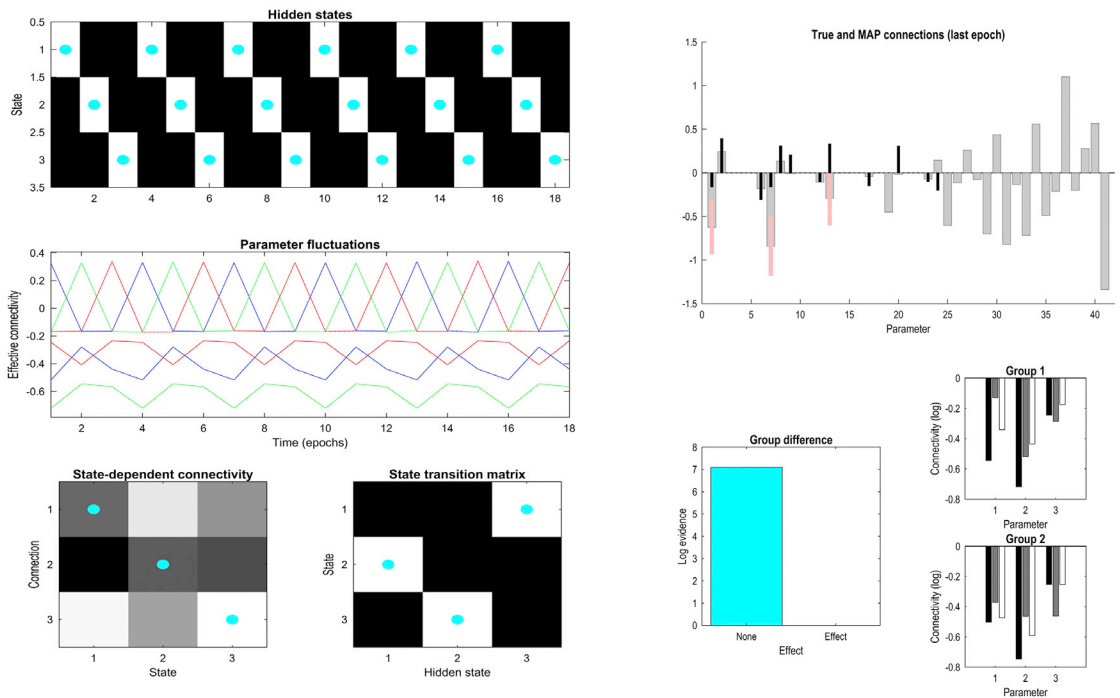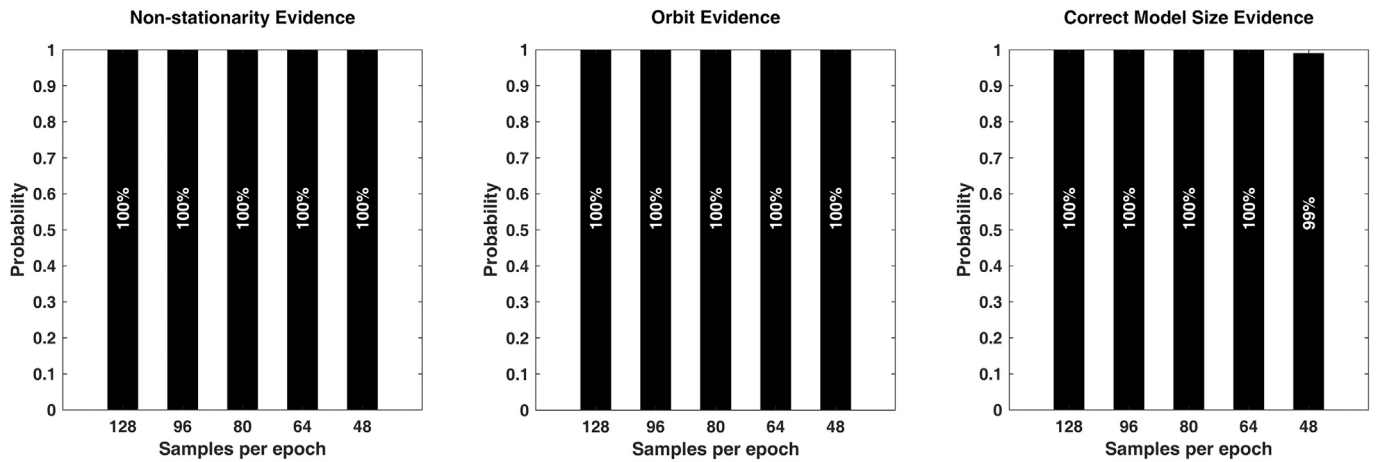
Fig. 13. *Posterior expectations of the model parameters inferred from the extended network with faster dynamics.* Supplementing the model evidences in Fig. 11, this figure shows the posterior expectation of the parameters when the inference scheme operates over shorter epochs of data, capturing faster orbital transitions (A) For $T = 64$ time samples per epoch, the itinerant characteristics are precisely captured in the posteriors, with strong evidence for a null group difference. The parameter fluctuations are partially correlated for two of the connections, but the strongest connection per state is quite indicative. (B) For $T = 48$ samples per epoch, the state sequence, transition pattern and group evidence are accurate, but the parameter fluctuations are smeared over the epochs; thereby losing state specificity.
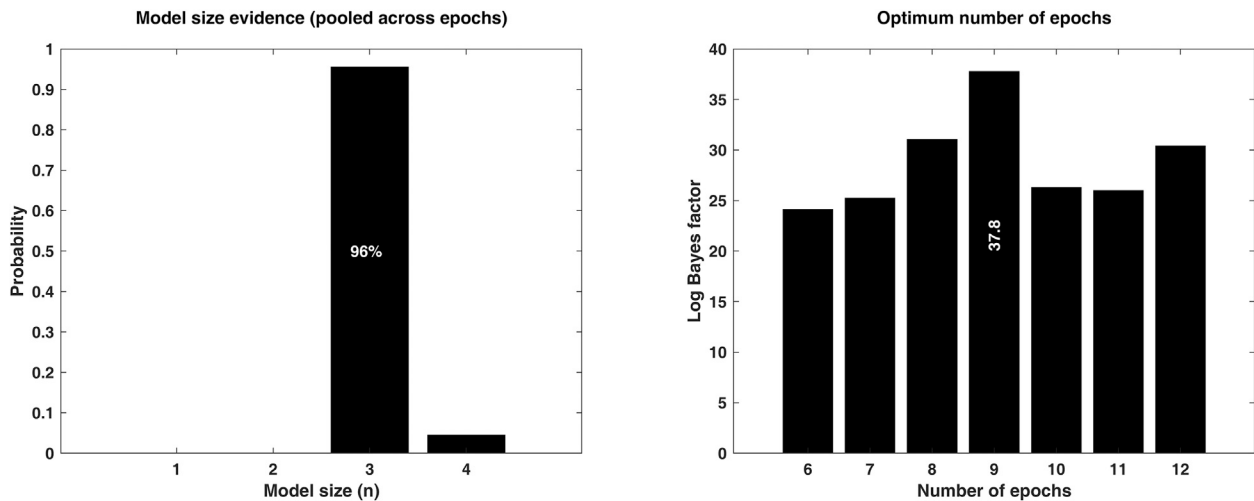

Fig. 14. *Optimum number of epochs or epoch length.* This figure shows that the most informative number of epochs for dynamic EC analysis can be inferred from the data, through a greedy search. In short, the data were generated using 3 states and 9 epochs of 80 time samples each, for 2 subjects, using the network of Fig. 10. This dataset was inverted using different combinations of window numbers (i.e., epoch lengths) and model sizes, (w, n), where w spanned [6–12] windows, n was in the [1–4] range, and the hyperparameters furnished itinerant priors ($\alpha = 4, \beta = 8$). The resulting free energies (i.e. approximate log model evidences) were evaluated for this model space, as $F(w, n)$. By pooling (summing) evidence for each model size over the range of windows numbers, the optimum number of latent states was inferred (left plot). Thereafter, a log Bayes factor was computed for each epoch length, scoring the plausibility of the orbital model with an optimal size (n_{opt}) against the stationary model (n=1). The epoch length (i.e., number of epochs) that produced the highest Bayes factor was deemed the most relevant for attributing orbital EC in the given dataset. In this case, the correct (9) number of epochs produced the highest log odds ratio (right plot).

empirical observations – enable mechanistic investigations of functional integration in the brain—and, implicitly, the role of phenomena such as metastability. In other words, using biologically-grounded models of how neuroimaging data are generated, it is possible to recover the underlying dynamics from measured signals, by solving an inverse problem. This is the basis of dynamic causal modelling. In most DCM studies, the inferences pertain to session-average responses and their underlying connectivity. Here, we present a hierarchical extension to accommodate temporal fluctuations (over successive epochs) in directed neuronal coupling. Specifically, this hierarchical model augments an established spectral DCM for resting state fMRI.

Conceptually, our hierarchical model assumes a winnerless competition among itinerant patterns of effective connectivity (i.e., states or modes of connectivity) of the brain at rest. The mathematical

image of such itinerant (i.e. metastable) behaviour is a stable heteroclinic channel (SHCh) or stable heteroclinic cycle (SHC). For three competitors (e.g., 3 states in our simulations), winnerless competition naturally results in a SHC. However, when there are more than three participants in a WLC, a SHCh is not necessarily cyclic (Afraimovich et al., 2010). SHCh can account for the reproducible itinerancy in neuronal dynamics (i.e. non-random sequences of modes) and reproducible transients induced by stimuli (Jones et al., 2007; Stopfer et al., 2003). However, a SHCh does not require the quasi-periodic recurrence of sequences. Such sequences have been referred to as quasi-periodic patterns (QPPs) in humans and rodents, with structured (and predictable) repetition rate and dwell time (Kashyap and Keilholz, 2019; Majeed et al., 2011; Thompson et al., 2014). Therefore, in virtue of the priors in our generative model, we commit to a particular class of SHCh,

which we have associated with a SHC—as an apt description of quasi-periodic connectivity patterns in the metastable brain.[9] Crucially, the transient dynamics of a SHC can be approximated by a hidden Markov model (of transitions between EC states), when the transitions possess certain itinerant characteristics; these characteristics were expressed as prior beliefs over the plausible forms of transitions in our model (Eq. (1), Fig. 3).

Having constructed the hierarchical generative model (Fig. 1), the face validity of the approach was established using simulations; i.e., synthetic resting state fMRI time series were generated (Fig. 4) and inverted (using variational Bayes, Fig. 2). We elaborated on the challenges of inverting a deep hierarchical model using sparse data (in this case, a few epochs of fMRI data per session). This problem was finessed by adopting a multi-start approach that evaluated and compared different models (cast in terms of priors) for a given data. In particular, this model space was spanned by the number of the latent states and itinerancy hyperparameters—i.e. (n, α, β). By evaluating the model evidence for different combinations of priors over these parameters, the joint probability over all combinations can be estimated. Thereafter, evidence for non-stationarity, model size and orbital state transitions can be recovered by marginalizing over subsets of this joint probability distribution (see Fig. 5).

Equipped with the priors (i.e. hyperparameter values) apt for simulated data, model inversion and comparison (using BMR) returned the expected parameter posteriors, as illustrated in Figs. 6 and 8, for the regular and irregular orbital transitions. The accompanying model evidence (i.e., free energy) was further used to perform group comparisons within and between pairs of subjects (Fig. 7), to characterize the sort of parametric changes in EC dynamics that can be detected using Bayesian model comparison.

In these simulations, the transition patterns for the two simulated groups were regular and irregular orbits, respectively. The specific choice of these transition patterns was motivated by the notion of *exit time* for perturbed SHCs (Kifer, 1981; Stone and Holmest, 1990) and practically by DFC reports of differences in *state occupancies* across populations.

A regular orbit is the simplest form of a structurally stable heteroclinic cycle, with a nearly deterministic progression through a sequence of saddles. The behaviour of such systems under random perturbations has been investigated theoretically in the literature (Kifer, 1981; Stone and Holmest, 1990). In brief, small random perturbations do not alter the global structure of the heteroclinic events, but render the duration of the events more stochastic. In our context, an event would correspond to an emergent EC mode, which the brain presumably maintains over a number of epochs. In particular, the study of (Stone and Holmest, 1990) concluded that, while weak additive noise does not essentially alter the structure of the solutions in the phase space, it induces a radical change in that it leads to "a selection of *timescales*".[10] In this sense, empirically estimated timescales of EC dynamics might be useful for informing canonical models of metastability (for the normal and pathological brain); which, in turn, can further clarify the role of noise and manifold instabilities in inducing the potentially altered heteroclinic structures and

time-scales, in particular conditions and disorders.

Simulating slower SHC dynamics (through an irregular orbit) was motivated by empirical DFC findings. For instance, a recent study (Vidaurre et al., 2017) concluded that the time spent visiting different (functional) connectivity states is not random, is heritable and predicts behavioural traits. The same research further demonstrated that *transition* patterns in brain activity and (functional) connectivity are not random; this non-random sequencing shows a hierarchical temporal structure (revealing two metastates); and that the brain cycles between distinct metastates. These findings (although phenomenological) tie in closely with the notion of SHC dynamics and the significance of their orbital periods.

Likewise, numerous other studies have established the relevance of (altered) functional state occupation and transition to demographics, consciousness, cognition, and clinical conditions: see (Preti et al., 2017) for a recent review. For instance, the authors of (Yang et al., 2014) found that the dwell time of a posterior-medial cortex seed, in a particular connectivity state, predicts inter-subject variability in mental flexibility and concept formation. Another DFC study (Cabral et al., 2017) found that closed loops exist between some FC states, and that preserved cognitive performance of healthy older adults is related to their ability to maintain certain (resting state) FC states for longer, compared to poor performers.

Furthermore, DFC researches on Schizophrenia (SZ) and post-traumatic stress disorder (PTSD) patients have reported the tendency of these subjects to linger or get trapped in certain states. These states transpired to be functionally less defined (less connected) in SZ (Damaraju et al., 2014; Du et al., 2016; Miller et al., 2016). In the case of PTSD patients, the state from which the subjects seemed unable to disengage was attributed to negative mood (Ou et al., 2015). These are all indications of pathological state dwell times (that can manifest as irregular orbits).

In addition to the inherent time-scale of the transitions, another potentially useful feature to quantify (across conditions or populations) is the state space dimension—i.e. the size of the repertoire of EC modes. This is the *model size* (n) that we inferred through BMC (Fig. 5). This feature would be particularly relevant in conditions or disorders associated with enhancement or shrinkage of the connectomic repertoire. For instance, recent studies demonstrated that psychedelics appear to enhance the dimensionality of state space and increase the traverse rate across this enlarged space (Atasoy et al., 2018; Tagliazucchi et al., 2014). Conversely, another functional connectivity study (Miller et al., 2016) suggests that SZ patients appear to be confined to a smaller state space (compared to healthy controls), both in terms of the number of realized meta-states and the distance (distinction) between the states. These observations might be usefully explored with dynamic EC, to further clarify the neuronal basis of connectivity states that are associated with psychopathology.

Having discussed some potential applications of the dynamic EC model, we now consider some pragmatic points for using and extending this model effectively. Based on our experiments so far, the limitation of this scheme is not the size of the network per se. It is rather the difficulty of estimating (fluctuating) DCM parameters from spectral features of short epochs. So as long as one intends to track a few specified connections over time (and is content with session-average values for the other connections), scaling up the network is not an issue. In fact, identification of large-scale spectral DCMs (with tens of nodes) has been achieved recently[11] (Razi et al., 2017). To apply this scheme to such networks, a few connections of interest should be specified *a priori* (based on biological hypotheses or data-driven dominant modes of FC (Seghier and Friston, 2013), to be followed over epochs.

For instance, our focus (in the simulations of this paper) was on the temporal evolution of inhibitory self-connections in the neuronal

---

[9] An interesting question could be about the location of SHCs in the brain. Cyclic (periodic) behaviour has been observed at different levels of organization of the brain, with separation of temporal scales (Cabral et al., 2017; Handwerker et al., 2012; Kashyap and Keilholz, 2019; Majeed et al., 2011; Vidaurre et al., 2017). We have assumed SHC at the level of immediate temporal dependencies of the latent states, realized over epochs. In theory, in an extended model with a deep hierarchy—encompassing different temporal scales— SHC could govern the top level of the model, and both stable heteroclinic channels and cycles may occupy lower levels. We will return to this later.

[10] By linearizing the system at the saddles, the authors were able to predict the probability distribution of the passage time (duration spent in the vicinity of a saddle) and subsequently the expected passage (exit) time. They showed that the latter depends only on the strongest unstable eigenvalue at the saddle and on the root mean square of the noise level (provided that it is small).

---

[11] Using FC modes to place prior constraints on the effective connections.

network. This choice was motivated by the following reasons: (1) The key mechanism underlying the winnerless competition in the brain is inhibition[12] (Buzsáki, 2006); (2) the explanatory power of intrinsic (i.e. self) connections has recently been emphasized in a spDCM study of seizure onset, using intracranial EEG recordings[13] (Papadopoulou et al., 2015); (3) the sensitivity of brain regions to their intrinsic inhibitory afferents speaks to the local excitation-inhibition balance—an important feature of connectivity in pathophysiology; (4) following the excitability of the regions through time allows one to estimate dwell time in different levels of the cortical hierarchy (e.g. staying longer in a sensory versus cognitive state), which is reportedly a consistent and heritable subject-specific measure that is related to cognitive traits (Vidaurre et al., 2017). However, in general, there is no operational restriction on the (intrinsic, extrinsic or combination of) connections that one may wish to track in this dynamic framework.

Another important issue is the size, shape and overlap of the windows, for partitioning the data into epochs. We have used Hanning windows (for their well-recognised trade-off between frequency and amplitude preservation), with 50% overlap. However, these choices could be optimized using BMC, for any given set of data (i.e., using the data features that provide the greatest relative evidence for any model). For example, we have shown how this procedure can identify the optimum epoch size (Fig. 14). In practice, the number of epochs is further restricted by the length of the session (in this case resting state fMRI). Related research (Park et al., 2017) used 200 time samples per epoch for the long rsfMRI sessions from the Human Connectome Project (HCP) database (Van Essen et al., 2012). Such long windows are suitable for tracking slow dynamics. On the extended network (Fig. 10A), we reduced the epoch size (to as little as 48 samples per window) to follow faster transitions. The presence of rapid itinerant dynamics was disclosed in terms of model evidence (Fig. 11). As for the posteriors, tracking very swift transitions (using 48 samples per epoch) can confound the parametric identification of EC states (Figs. 12–13), under the generative model of spectral DCM used here. However, the remaining parameter posteriors and model evidences were robust to this challenging scenario. We have also described a greedy search and BMC procedure for picking the most efficient window size, when tracking orbital dynamics of unknown rate (Fig. 14). One could also consider using adaptive epoch lengths, motivated by approaches such as change point detection (Jeong et al., 2016). Finally, we reiterate that these data feature parameters can be optimized based on the (relative) model evidence for any given data, using BMC.

A final point pertains to the depth of the hierarchy. As a proof of concept, we chose the simplest hierarchy; i.e. we adopted a fixed effects model for the effective connectivity patterns and state transitions at the group level. However, a random effects model could be considered that incorporates small between-epoch variations of the connectivity modes, per subject. It is possible, in theory, to model systematic differences in the connectivity profiles and transition patterns of different subjects as well, by entertaining additional random effects models. However, this would entail inverting deeper and deeper hierarchical models.

One can also envisage deeper hierarchical models that can entertain SHCs of SHCs/SHChs. In other words, the cyclic dynamics—that we presumed underlines EC variations—might as well be cycles of cycles (or trajectories) generated in higher (deeper) levels of a hierarchical model. That is, every saddle of a deeper (slower) SHC re-routes the trajectory between a number of lower level states, imposing a temporary form of (faster) SHC/SHCh on the subordinate level; the latter is again re-wired upon a transition to the next saddle in the higher SHC (Kiebel et al.,

2009). This conjecture is in accordance with the empirically derived (functional) metastates in (Vidaurre et al., 2017), which are located on top of a temporal hierarchy. Their study showed that metastates evince cyclic behaviour, and each metastate is associated with a number of subordinate states – that tend to switch more often into each other. This is an example of a deeper hierarchical structure (presumably a SHC of SHCs/SHChs) that might underlie the complex itinerancy of the brain's connectivity modes.

The viability of such extensions (and ensuing model inversions) for the relatively short rsfMRI sessions remains to be further examined. Nevertheless, since the proposed scheme operates within a generic Bayesian framework, it is possible to accommodate different models on the first level of the hierarchy (i.e. other than spDCM for fMRI); as a result, it may be worthwhile integrating hierarchical models (of transient dynamics) with the generative models of other neuroimaging data (such as M/EEG) or electrophysiological recordings. Although these modalities come with their own inverse problems, the higher temporal resolution and the abundance of data might resolve some of the optimisation issues inherent in the inversion of hierarchical models of fMRI timeseries.

For instance, animal models have been frequently used to investigate the neural basis of (dynamic) functional connectivity, especially because they facilitate simultaneous imaging and invasive electrophysiological recordings (Keilholz, 2014; Majeed et al., 2009; Shmuel and Leopold, 2008; Thompson et al., 2014, 2013; Zhang et al., 2019). Furthermore, image acquisition in high-field animal scanners—with shorter repetition times—furnishes data with higher temporal and spatial resolution. These studies have enhanced our understating of the contributions of different frequency bands to functional connectivity (and its fluctuations) (see (Keilholz, 2014) for review and references). Early evidence for repeated spatiotemporal patterns in spontaneous BOLD fluctuations came from anesthetized rats (Majeed et al., 2009) and subsequent work showed that the process was nearly periodic (Majeed et al., 2011).

In short, animal models may prove useful, both for empirical validation of our approach (e.g., with benchmark states generated via administration of drugs or with periodic tasks) and for exploratory research on the dynamics of neuronal networks, in different contexts (for instance in animal models of neurologic disorders). In other words, it is possible to apply our proposed scheme to animal electrophysiological, imaging, or multimodal recordings to investigate sequences of effective connectivity modes and how they might be modulated under different pharmacological or performance conditions. Of course, depending on the question (or hypothesis)—and the data modality at hand—the right type of DCM, with sufficient biophysical detail, has to be selected (Friston et al., 2017; Kiebel et al., 2008; Moran et al., 2013). But once the posteriors of the DCM parameters have been estimated, the rest of the scheme described in this technical note is quite generic.

Clearly, many aspects and details remain to be investigated in future studies. Particularly, we hope to show the predictive validity of the proposed approach using empirical data from clinical populations in future papers.

---

[12] In general, inhibitory connections mediated by interneurons are responsible for spatiotemporal transient activity, whereas excitatory cell populations and their connections ensure that the information "goes to the right place at the right time" (Afraimovich et al., 2010; Buzsáki, 2006).

[13] In this study, it was demonstrated that the intrinsic synaptic changes were sufficient to explain seizure onset.

References

Afraimovich, V., Tristan, I., Huerta, R., Rabinovich, M.I., 2008. Winnerless competition principle and prediction of the transient dynamics in a Lotka-Volterra model. Chaos, 043103. https://doi.org/10.1063/1.2991108.

Afraimovich, V.S., Muezzinoglu, M.K., Rabinovich, M.I., 2010. Metastability and transients in brain dynamics: problems and rigorous results. In: Long-Range Interactions, Stochasticity and Fractional Dynamics. Springer Berlin Heidelberg, pp. 133–175. https://doi.org/10.1007/978-3-642-12343-6_4.

Atasoy, S., Vohryzek, J., Deco, G., Carhart-Harris, R.L., Kringelbach, M.L., 2018. Common neural signatures of psychedelics: frequency-specific energy changes and repertoire expansion revealed using connectome-harmonic decomposition. Prog. Brain Res. 242, 97–120. https://doi.org/10.1016/bs.pbr.2018.08.009.

Busse, F.H., Heikes, K.E., 1980. Convection in a rotating layer: a simple case of turbulence. Science (80-.) 208, 173–175. https://doi.org/10.1126/science.208.4440.173.

Buzsáki, G., 2006. Rhythms of the Brain. Oxford University Press.

Cabral, J., Vidaurre, D., Marques, P., Magalhães, R., Silva Moreira, P., Miguel Soares, J., Deco, G., Sousa, N., Kringelbach, M.L., 2017. Cognitive performance in healthy older adults relates to spontaneous switching between states of functional connectivity during rest. Sci. Rep. 7, 1–13. https://doi.org/10.1038/s41598-017-05425-7.

Camba-Mendez, G., Kapetanios, G., 2005. Estimating the rank of the spectral density matrix. J. Time Ser. Anal. https://doi.org/10.1111/j.1467-9892.2005.00389.x.

Chang, C., Glover, G.H., 2010. Time–frequency dynamics of resting-state brain connectivity measured with fMRI. Neuroimage 50, 81–98. https://doi.org/10.1016/j.neuroimage.2009.12.011.

Chang, C., Liu, Z., Chen, M.C., Liu, X., Duyn, J.H., 2013. EEG correlates of time-varying BOLD functional connectivity. Neuroimage. https://doi.org/10.1016/j.neuroimage.2013.01.049.

Cooray, G.K., Sengupta, B., Douglas, P., Englund, M., Wickstrom, R., Friston, K., 2015. Characterising seizures in anti-NMDA-receptor encephalitis with dynamic causal modelling. Neuroimage 118, 508–519. https://doi.org/10.1016/j.neuroimage.2015.05.064.

Damaraju, E., Allen, E.A.A., Belger, A., Ford, J.M.M., McEwen, S., Mathalon, D.H.H., Mueller, B.A.A., Pearlson, G.D.D., Potkin, S.G.G., Preda, A., Turner, J.A.A., Vaidya, J.G.G., Erp, T.G. Van, Calhoun, V.D.D., Van Erp, T.G.G., Calhoun, V.D.D., 2014. Dynamic functional connectivity analysis reveals transient states of dysconnectivity in schizophrenia. NeuroImage Clin. 5, 298–308. https://doi.org/10.1016/j.nicl.2014.07.003.

Daunizeau, J., Friston, K.J., 2007. A mesostate-space model for EEG and MEG. Neuroimage 38, 67–81. https://doi.org/10.1016/j.neuroimage.2007.06.034.

Deco, G., Jirsa, V.K., 2012. Ongoing cortical activity at rest: criticality, multistability, and ghost attractors. J. Neurosci. 32, 3366–3375. https://doi.org/10.1523/JNEUROSCI.2523-11.2012.

Deco, G., Jirsa, V.K., McIntosh, A.R., 2011. Emerging concepts for the dynamical organization of resting-state activity in the brain. Nat. Rev. Neurosci. 12, 43–56. https://doi.org/10.1038/nrn2961.

Deco, G., Kringelbach, M.L., Jirsa, V.K., Ritter, P., 2017. The dynamics of resting fluctuations in the brain: metastability and its dynamical cortical core. Sci. Rep. 7, 1–14. https://doi.org/10.1038/s41598-017-03073-5.

Du, Y., Pearlson, G.D., Yu, Q., He, H., Lin, D., Sui, J., Wu, L., Calhoun, V.D., 2016. Interaction among subsystems within default mode network diminished in schizophrenia patients: a dynamic connectivity approach. Schizophr. Res. https://doi.org/10.1016/j.schres.2015.11.021.

Friston, K.J., 2011. Functional and effective connectivity: a review. Brain Connect. 1, 13–36. https://doi.org/10.1089/brain.2011.0008.

Friston, K.J., 2002. Functional integration and inference in the brain. Prog. Neurobiol. 590, 1–31. https://doi.org/10.1016/S0301-0082(02)00076-X.

Friston, K.J., Harrison, L., Penny, W., 2003. Dynamic causal modelling. Neuroimage 19, 1273–1302. https://doi.org/10.1016/S1053-8119(03)00202-7.

Friston, K.J., Kahan, J., Biswal, B., Razi, A., 2014a. A DCM for resting state fMRI. Neuroimage 94, 396–407. https://doi.org/10.1016/j.neuroimage.2013.12.009.

Friston, K.J., Kahan, J., Razi, A., Stephan, K.E., Sporns, O., 2014b. On nodes and modes in resting state fMRI. Neuroimage. https://doi.org/10.1016/j.neuroimage.2014.05.056.

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. https://doi.org/10.1016/j.neuroimage.2015.11.015.

Friston, K.J., Preller, K.H., Mathys, C., Cagnan, H., Heinzle, J., Razi, A., Zeidman, P., 2017. Dynamic causal modelling revisited. Neuroimage. https://doi.org/10.1016/j.neuroimage.2017.02.045.

Haken, H., 1983. SYNERGETICS-an introduction: nonequilibrium phase transition and self-organization in physics. Chem. Biol.

Handwerker, D.A., Roopchansingh, V., Gonzalez-Castillo, J., Bandettini, P.A., 2012. Periodic changes in fMRI connectivity. Neuroimage. https://doi.org/10.1016/j.neuroimage.2012.06.078.

Heitmann, S., Breakspear, M., 2018. Putting the "dynamic" back into dynamic functional connectivity. Netw. Neurosci. 2, 150–174.

Hindriks, R., Adhikari, M.H., Murayama, Y., Ganzetti, M., Mantini, D., Logothetis, N.K., Deco, G., 2015. Can sliding-window correlations reveal dynamic functional connectivity in resting-state fMRI? Neuroimage. https://doi.org/10.1016/j.neuroimage.2015.11.055.

Hutchison, R.M., Womelsdorf, T., Allen, E.A., Bandettini, P.A., Calhoun, V.D., Corbetta, M., Della Penna, S., Duyn, J.H., Glover, G.H., Gonzalez-Castillo, J., Handwerker, D.A., Keilholz, S., Kiviniemi, V., Leopold, D.A., de Pasquale, F., Sporns, O., Walter, M., Chang, C., 2013. Dynamic functional connectivity: promise, issues, and interpretations. Neuroimage 80, 360–378. https://doi.org/10.1016/j.neuroimage.2013.05.079.

Jeong, S.O., Pae, C., Park, H.J., 2016. Connectivity-based change point detection for large-size functional networks. Neuroimage 143, 353–363. https://doi.org/10.1016/j.neuroimage.2016.09.019.

Jirsa, V.K., Friedrich, R., Haken, H., Kelso, J.A.S., 1994. A theoretical model of phase transitions in the human brain. Biol. Cybern. https://doi.org/10.1007/BF00198909.

Jones, L.M., Fontanini, A., Sadacca, B.F., Miller, P., Katz, D.B., 2007. Natural stimuli evoke dynamic sequences of states in sensory cortical ensembles. Proc. Natl. Acad. Sci. U.S.A. https://doi.org/10.1073/pnas.0705546104.

Kashyap, A., Keilholz, S.D., 2019. Brain network constraints and recurrent neural networks reproduce unique trajectories and state transitions seen over the span of minutes in resting state fMRI. https://doi.org/10.1101/798520.

Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Am. Stat. Assoc. 90, 773–795. https://doi.org/10.1080/01621459.1995.10476572.

Keilholz, S.D. aw D., 2014. The neural basis of time-varying resting-state functional connectivity. Brain Connect. 4, 769–779. https://doi.org/10.1089/brain.2014.0250.

Keilholz, S.D., Magnuson, M.E., Pan, W. ju, Willis, M., Thompson, G.J., 2013. Dynamic properties of functional connectivity in the rodent. Brain Connect. https://doi.org/10.1089/brain.2012.0115.

Kiebel, S.J., Garrido, M.I., Moran, R.J., Friston, K.J., 2008. Dynamic causal modelling for EEG and MEG. Cogn. Neurodyn. 2, 121.

Kiebel, S.J., Von Kriegstein, K., Daunizeau, J., Friston, K.J., 2009. Recognizing sequences of sequences. PLoS Comput. Biol. 5 https://doi.org/10.1371/journal.pcbi.1000464.

Kifer, Y., 1981. The exit problem for small random perturbations of dynamical systems with a hyperbolic fixed point. Isr. J. Math. 40, 74–96.

Krupa, M., 1997. Robust heteroclinic cycles. J. Nonlinear Sci. https://doi.org/10.1007/BF02677976.

Leonardi, N., Ville, D. Van, 2015. On spurious and real fluctuations of dynamic functional connectivity during rest. Neuroimage 104, 430–436. https://doi.org/10.1016/j.neuroimage.2014.09.007.

Majeed, W., Magnuson, M., Hasenkamp, W., Schwarb, H., Schumacher, E.H., Barsalou, L., Keilholz, S.D., 2011. Spatiotemporal dynamics of low frequency BOLD fluctuations in rats and humans. Neuroimage 54, 1140–1150. https://doi.org/10.1016/j.neuroimage.2010.08.030.

Majeed, W., Magnuson, M., Keilholz, S.D., 2009. Spatiotemporal dynamics of low frequency fluctuations in BOLD fMRI of the rat. J. Magn. Reson. Imaging. https://doi.org/10.1002/jmri.21848.

Miller, R.L., Yaesoubi, M., Turner, J.A., Mathalon, D., Preda, A., Pearlson, G., Adali, T., Calhoun, V.D., 2016. Higher dimensional meta-state analysis reveals reduced resting fMRI connectivity dynamism in schizophrenia patients. PLoS One 11, e0149849. https://doi.org/10.1371/journal.pone.0149849.

Moran, R., Pinotsis, D.A., Friston, K., 2013. Neural masses and fields in dynamic causal modelling. Front. Comput. Neurosci. 7, 1–12. https://doi.org/10.3389/fncom.2013.00057.

Ou, J., Xie, L., Jin, C., Li, X., Zhu, D., Jiang, R., Chen, Y., Zhang, J., Li, L., Liu, T., 2015. Characterizing and differentiating brain state dynamics via hidden Markov models. Brain Topogr. 28, 666–679. https://doi.org/10.1007/s10548-014-0406-2.

Papadopoulou, M., Cooray, G., Rosch, R., Moran, R., Marinazzo, D., Friston, K., 2017. Dynamic causal modelling of seizure activity in a rat model. Neuroimage 146, 518–532. https://doi.org/10.1016/j.neuroimage.2016.08.062.

Papadopoulou, M., Leite, M., Mierlo, P. Van, Vonck, K., Lemieux, L., Friston, K., Marinazzo, D., 2015. Tracking slow modulations in synaptic gain using dynamic causal modelling : validation in epilepsy. Hum. Brain Mapp. J. 107, 117–126. https://doi.org/10.1016/j.neuroimage.2014.12.007.

Park, H., Friston, K.J., Pae, C., Park, B., Razi, A., 2017. Dynamic effective connectivity in resting state fMRI. Neuroimage 1–15. https://doi.org/10.1016/j.neuroimage.2017.11.033.

Preti, M.G., Bolton, T.A., Van De Ville, D., 2017. The dynamic functional connectome: state-of-the-art and perspectives. Neuroimage 160, 41–54. https://doi.org/10.1016/j.neuroimage.2016.12.061.

Rabinovich, M., Huerta, R., Laurent, G., 2008. Neuroscience: transient dynamics for neural processing. Science 80. https://doi.org/10.1126/science.1155564.

Rabinovich, M.I., Afraimovich, V.S., Bick, C., Varona, P., 2012. Information flow dynamics in the brain. Phys. Life Rev. 9 https://doi.org/10.1016/j.plrev.2011.11.002.

Rabinovich, M.I., Huerta, R., Varona, P., Afraimovich, V.S., 2008. Transient cognitive dynamics, metastability, and decision making. PLoS Comput. Biol. 4 https://doi.org/10.1371/journal.pcbi.1000072.

Rabinovich, M.I., Varona, P., 2018. Discrete sequential information coding: heteroclinic cognitive dynamics. Front. Comput. Neurosci. 12, 1–11. https://doi.org/10.3389/fncom.2018.00073.

Rabinovich, M.I., Varona, P., 2011. Robust transient dynamics and brain functions. Front. Comput. Neurosci. 5, 24. https://doi.org/10.3389/fncom.2011.00024.

Razi, A., Kahan, J., Rees, G., Friston, K.J., 2015. Construct validation of a DCM for resting state fMRI. Neuroimage 106, 1–14. https://doi.org/10.1016/j.neuroimage.2014.11.027.

Razi, A., Seghier, M.L., Zhou, Y., Mccolgan, P., 2017. Large-scale DCMs for resting state fMRI. Netw. Neurosci. 1, 222–241.

Roberts, J.A., Gollo, L.L., Abeysuriya, R.G., Roberts, G., Philip, B., Woolrich, M.W., Breakspear, M., Mitchell, P.B., Woolrich, M.W., Breakspear, M., 2019. Metastable brain waves. Nat. Commun. 10, 1056. https://doi.org/10.1101/347054.

Sakoğlu, U., Pearlson, G.D., Kiehl, K.A., Wang, Y.M., Michael, A.M., Calhoun, V.D., 2010. A method for evaluating dynamic functional network connectivity and task-modulation: application to schizophrenia. Magma 23, 351–366. https://doi.org/10.1007/s10334-010-0197-8.

Seghier, M.L., Friston, K.J., 2013. Network discovery with large DCMs. Neuroimage 68, 181–191. https://doi.org/10.1016/j.neuroimage.2012.12.005.

Shmuel, A., Leopold, D.A., 2008. Neuronal correlates of spontaneous fluctuations in fMRI signals in monkey visual cortex: implications for functional connectivity at rest. Hum. Brain Mapp. 29, 751–761. https://doi.org/10.1002/hbm.20580.

Stone, E., Holmest, P., 1990. Random perturbations of heteroclinic attractors, 50, 726–743.

Stopfer, M., Jayaraman, V., Laurent, G., 2003. Intensity versus identity coding in an olfactory system. Neuron. https://doi.org/10.1016/j.neuron.2003.08.011.

Tagliazucchi, E., Carhart-Harris, R., Leech, R., Nutt, D., Chialvo, D.R., 2014. Enhanced repertoire of brain dynamical states during the psychedelic experience. Hum. Brain Mapp. 35, 5442–5456. https://doi.org/10.1002/hbm.22562.

Tagliazucchi, E., von Wegner, F., Morzelewski, A., Brodbeck, V., Laufs, H., 2012. Dynamic BOLD functional connectivity in humans and its electrophysiological correlates. Front. Hum. Neurosci. 6, 339. https://doi.org/10.3389/fnhum.2012.00339.

Thompson, G.J., Merritt, M.D., Pan, W.J., Magnuson, M.E., Grooms, J.K., Jaeger, D., Keilholz, S.D., 2013. Neural correlates of time-varying functional connectivity in the rat. Neuroimage 83, 826–836. https://doi.org/10.1016/j.neuroimage.2013.07.036.

Thompson, G.J., Pan, W.J., Magnuson, M.E., Jaeger, D., Keilholz, S.D., 2014. Quasi-periodic patterns (QPP): large-scale dynamics in resting state fMRI that correlate with local infraslow electrical activity. Neuroimage. https://doi.org/10.1016/j.neuroimage.2013.09.029.

Tognoli, E., Kelso, J.A.S., 2014. The metastable brain. Neuron 81, 35–48. https://doi.org/10.1016/j.neuron.2013.12.022.

Van Essen, D.C., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T.E.J., Bucholz, R., Chang, A., Chen, L., Corbetta, M., Curtiss, S.W., Della Penna, S., Feinberg, D., Glasser, M.F., Harel, N., Heath, A.C., Larson-Prior, L., Marcus, D., Michalareas, G., Moeller, S., Oostenveld, R., Petersen, S.E., Prior, F., Schlaggar, B.L., Smith, S.M., Snyder, A.Z., Xu, J., Yacoub, E., 2012. The Human Connectome Project: a data acquisition perspective. Neuroimage 62, 2222–2231. https://doi.org/10.1016/j.neuroimage.2012.02.018.

Vidaurre, D., Smith, S.M., Woolrich, M.W., 2017. Brain network dynamics are hierarchically organized in time. Proc. Natl. Acad. Sci. 114, 12827–12832. https://doi.org/10.1073/pnas.1705120114.

Yang, Z., Craddock, R.C., Margulies, D.S., Yan, C.G., Milham, M.P., 2014. Common intrinsic connectivity states among posteromedial cortex subdivisions: insights from analysis of temporal dynamics. Neuroimage 93, 124–127. https://doi.org/10.1016/j.neuroimage.2014.02.014.

Zalesky, A., Breakspear, M., 2015. Towards a statistical test for functional connectivity dynamics. Neuroimage 114, 466–470. https://doi.org/10.1016/j.neuroimage.2015.03.047.

Zalesky, A., Fornito, A., Cocchi, L., Gollo, L.L., Breakspear, M., 2014. Time-resolved resting-state brain networks. Proc. Natl. Acad. Sci. U.S.A. 111, 10341–10346. https://doi.org/10.1073/pnas.1400181111.

Zhang, X., Pan, W.-J., Keilholz, S.D., 2019. The relationship between BOLD and neural activity arises from temporally sparse events. bioRxiv 644419. https://doi.org/10.1101/644419.