**Awakening: promoting transitions between different brain states in a probabilistic state space framework**

Gustavo Deco1,2,3,4, Josephine Cruzat1, Joana Cabral5,6,7, Enzo Tagliazucchi8,9, Helmut Laufs9,10, Nikos K. Logothetis 11,12 & Morten L. Kringelbach5,6,7

1. Center for Brain and Cognition, Computational Neuroscience Group, Department of Information and Communication Technologies, Universitat Pompeu Fabra, Roc Boronat 138, Barcelona, 08018, Spain

2. Institució Catalana de la Recerca i Estudis Avançats (ICREA), Passeig Lluís Companys 23, Barcelona, 08010, Spain

3. Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, 04103 Leipzig, Germany

4. School of Psychological Sciences, Monash University, Melbourne, Clayton VIC 3800, Australia

5. Department of Psychiatry, University of Oxford, Oxford, UK

6. Center for Music in the Brain, Department of Clinical Medicine, Aarhus University, DK

7. Life and Health Sciences Research Institute (ICVS), School of Medicine, University of Minho, 4710-057, Braga, Portugal

8. Institute for Medical Psychology, Christian Albrechts University, Kiel, Germany

9. Department of Neurology and Brain Imaging Center, Goethe University, Frankfurt am Main, Germany

10. Department of Neurology, Christian Albrechts University, Kiel, Germany

11. Max Planck Institute for Biological Cybernetics, 72076 Tübingen, Germany

12. Imaging Science and Biomedical Engineering, University of Manchester, Manchester M13 9PT, UK

*Corresponding authors:*

Morten L. Kringelbach, Department of Psychiatry, University of Oxford, Oxford, UK. morten.kringelbach@psych.ox.ac.uk, tel: 0044-1865613118

Gustavo Deco, Universitat Pompeu Fabra, C/ Roc Boronat, 138, 08018 Barcelona, Spain. gustavo.deco@upf.edu, Tel: 0034-935422977, Fax: 0034-935422517

*Conflict of interest: the authors declare to have no conflict of interest.*

**Running Title:** Probabilistic State Space Description of Whole-Brain Dynamics and Modeling

**Keywords:** Whole Brain Modelling, fMRI Resting State, Sleep, Perturbation of the Brain

# Abstract

A fundamental unresolved problem in neuroscience is understanding the causative mechanisms involved in transitions between different brain states such as healthy wakefulness and, say, sleep, coma or neuropsychiatric disease. In order to solve this key question, we here show the fundamental mechanisms of how to force a transition between two brain states by external stimulation. We demonstrate how to define and model a brain state quantitatively and systematically probe this model to find the best brain targets to force a transition. A brain state can be defined in terms of a quantitative description of the underlying dynamical repertoire as a probabilistic description of an ensemble of possible ‘metastable substates’. The key idea of this probabilistic state space approach is to use a probabilistic description of the underlying spatiotemporal dynamics of functional magnetic resonance imaging. This *Probabilistic Metastable Substates Space* can be precisely fitted by a whole-brain model, which can then be systematically perturbed *in silico* to predict exactly how to force a transition between brain states; for example awakening a sleeping brain. This paradigm opens up for radical translational applications; potentially honing the precision and power of stimulation technologies (e.g. DBS and TMS) to allow the re-establishment of healthy brain states from neuropsychiatric disease such that they can promote a profound reconfiguration of the dynamical landscape necessary for recovery.

NOTE: I can will update the abstract only after you read the introduction/discussion and let me know whether the changes make sense….

# Introduction

Almost thirteen decades ago the father of cognitive psychology, William James, wrote “Everybody knows what attention is” (James, 1890). Indeed, we have long been recognizing well various “states of the brain”, including sleep, wakefulness, aphasia or attention, but our understanding of the actual “brain-states” underlying such states remains a mystery. Brains have billions of neurons and trillions of synapses, nested recursive circuits at all possible spatiotemporal levels, massive connectivity, and initial condition dependent activity-evolution. In disciplines such as physics, genomics or economics, such systems are characterized as complex dynamic ones, whereby complex implies that the ultimately emerging global behavior cannot be understood by merely studying the network’s elementary nodes. The continuously evolving dynamics of such widespread networks are characterized by a condition-dependent self-organization, going through stable, “quasi-stable”, high- or low-activities and transient arrangements, termed brain-states. Given the complexity of brain-states and their probabilistic, often chaotic, state-transitions their in-depth study up to now has been minimally successful.

Attempts have been made, for instance, to define a state space by estimating the relationship of selected spatiotemporal and spectral characteristics of activity within a window shifted over time. The high dimensionality of signals in such cases can be reduced to 2 or 3 values plotted in a low dimensional state space enabling visualization and easier interpretation of the data (Gervasoni *et al.*, 2004; Gu *et al.*, 2017; Gu *et al.*, 2018; Muldoon *et al.*, 2016). Yet such a representation fails to capture the dynamics of states and their transitions.

Brain states can also be described as attractors, i.e. stable states of interacting brain regions, offering insights into repeatable and robust system configurations (Deco and Jirsa, 2012; Gu *et al.*, 2018), or by additionally using generative models that take into account context- and rank-dependent constraints at different hierarchical processing levels (Deco *et al.*, 2009; Deco *et al.*, 2015; Friston *et al.*, 2003). Interregional interactions are commonly reflected in the degree of coupling of neural activity oscillations of local (microcircuit) or remote neuromodulatory origin.

All these approaches, however, fail to capture so-called metastability, i.e. the quality of systems, including the brain, to temporarily persist in an existing equilibrium despite slight perturbations. Evidently, the dynamics of coordination between brain regions with high functional differentiation is less likely to be as stable and persistent as the coupling observed between, say, areas that are part of a single sensory or motor system. And the coordination patterns within stable states may often reflect dynamically recurring short-lived patterns that can be clustered in metastable substates for any given brain state (e.g. Laumann *et al.*, 2016; Deco and Kringelbach, 2016; Deco *et al.*, 2017c; Tognoli and Kelso, 2014).

In the current study, we first demonstrate that recurrent substates can indeed be detected and characterized in terms of probability of occurrence and alternation profiles. This novel approach dubbed as *Characterization of Probabilistic Metastable Substates Space,* fully typifies substates as stochastic subdivisions of regular and persistent brain states.

Second, we wanted to address the fundamental question in neuroscience of how the brain transitions come about between different states, e.g. from wakefulness to deep sleep and anaesthesia or to disease states such as coma and neuropsychiatric disorders (Dehaene, 2014; Stitt *et al.*, 2017; Tononi, 2012). Solving this problem requires two things: 1) a deeper understanding and quantitative definition of what constitutes a brain state, e.g. our proposed characterisation of the probabilistic Metastable Substates Space – and 2) what drives the transitions between brain states. This would allow for the possibility of forcing a transition using for example using external stimulation like deep brain stimulation (DBS) or transmagnetic stimulation (TMS) (Kringelbach *et al.*, 2011; Kringelbach *et al.*, 2007; Pascual-Leone *et al.*, 1996; Rossini *et al.*, 2015; Schiff *et al.*, 2007).

In order to do this, we need whole-brain models that can link the underlying anatomical connectivity with the functional dynamics obtained from neuroimaging (Breakspear, 2017; Cabral *et al.*, 2017a; Deco *et al.*, 2009; Deco *et al.*, 2015; Ghosh *et al.*, 2008; Honey *et al.*, 2007). We show here that a generative whole-brain model can actually accurately fit the *Probabilistic Metastable Substates Space* of the empirical data corresponding to different brain states. Moreover, we provide first evidence that significant insights into brain states and their transitions can be gained by means of *in silico* stimulation of the whole-brain model. Lastly, as proof of concept, we use this novel method to demonstrate that it can find accurate ways to promote transition from one brain state to another (as characterised by functional magnetic resonance imaging), and in particular finding ways “awaken” the brain from deep sleep to wakefulness and vice versa.

# Results

The first step of the analysis has been to identify the *Probabilistic Metastable Substates (PMS) space*, for which we use the Leading Eigenvector Dynamics Analysis (LEiDA) method (Cabral *et al.*, 2017d), described in detail in the Methods section and in **Figure 1**. The method uses BOLD Phase Coherence Connectivity to obtain a time-resolved dynamic FC (dFC) matrix that which we extracted the corresponding time-resolved leading eigenvectors, which were then used to extract the substates. The leading eigenvector captures the dominant connectivity pattern of dFC(t) at time *t*, from which one can detect a discrete number of reduced dFC patterns by applying clustering across time points and subjects. The obtained k-cluster centroids define the ‘metastable substates’, and for which we compute the probability and transition probability. We use this method on two different naturally occurring brain states: awake and deep sleep conditions obtained in healthy human participants measured with fMRI and EEG (Tagliazucchi and Laufs, 2014). The clustering as determined by the silhouettes criterion showed that three states were optimal for fitting both brain states. Nevertheless, it is entirely possible, albeit very computationally expensive, to use a higher number of states with the known networks involved in sleep (Stevner *et al.*, 2018). Here we used three states as a proof of concept.

The strategy of state definition is summarized in **Figure 2**, demonstrating a probabilistic state space that may be underlying a given brain state, e.g. wakefulness, sleep, anaesthesia or disease. In order to promote or force transitions between brain states, we propose a framework to characterize brain states (**Figure 2A**) and to then use whole-brain models to generate such brain states (**Figure 2B**) and then force the transition between two brain states (**Figure 2C**). The framework describes how in a given brain state, each point represents an instantaneous snapshot of the whole-brain at a point in time, where the totality of this probabilistic cloud of points will describe the full brain state. In **Figure 2A**, these points are projected onto a 2D space for visualisation but the original state space is of course likely to be higher dimensional. This cloud of points can be described as a probabilistic distribution in a state space. Using clustering, we can force different substates from groups of points and express the probability of each of these substates (**Figure 2B**). Given that this is not stable and changing over time, we call this approach *Probabilistic Metastable Substates (PMS) space*. In **Figure 2B**, we show how we can fit a whole-brain model to this PMS space. In **Figure 2C**, we show how the whole-brain model can be perturbed and stimulated exhaustively in order to promote and force a specific transition between two different brain states.

=================== TO BE DONE…

## Optimal spatiotemporal fit of whole-brain model to PMS space

As a second step, in order to fit the model this PMS space, we used a large-scale model, composed of local nodes representing local brain regions in a parcellation (see Methods). The different brain regions are coupled by the underlying anatomical connectivity matrix, i.e. the Structural Connectivity (SC), between those nodes. The SC matrix was obtained using diffusion MRI and tractography techniques (see Methods). The dynamics of each local brain area network are described by the normal form of a supercritical Hopf bifurcation, also called a Landau-Stuart Oscillator (Kuznetsov, 1998).

The internal parameters of the whole-brain model (coupling strength between nodes) can be optimized to fit the PMS space using the empirical measurements extracted from experimental data. Indeed, the strategy here is to use the empirical centroids of the empirical clusters defining the metastable substates, and to compute the simulated probabilistic measurements based on those empirical centres. In this way, we assure that the probabilistic measurements correspond exactly to the empirical metastable substates.

**Figure 3A** shows the performance of the whole-brain model for fitting resting state wakefulness data. In this case, we explore exhaustively the one single parameter in the model, *G*, which scales the inter-area coupling determining the dynamical working point of the system. This *G* coupling parameter scales the density of fibres expressed in the SC and can be interpreted as the synaptic conductivity of each single fibre, i.e. we are assuming, as a matter of simplicity, that these conductivities are all equal across the brain. **Figure 3A** shows two measurements of fitting as a function of *G*: 1) the phase-coherence-based functional connectivity dynamics (FCD); and 2) the probabilistic metastable substates (PMS). The first is computed by collecting the upper triangular elements of the time-resolved phase coherence connectivity dFC(t) matrices (over all participants) and then use the Kolmogorov-Smirnov (KS) distance to compare these empirical distributions with the corresponding simulated distributions from the model. The minimum value of this KS distance corresponds to the optimal fitting of the spatiotemporal characteristics and is obtained at *G*=0.283.

Nevertheless, this optimal minimum for the FCD turns out not to be the best fitting of the spatiotemporal dynamics as characterized by PMS. We extract the empirical metastable substates (**Figure 3B**) and show the transition probability matrix and graph as well as the probability state space. In order to find the optimal minimum for the PMS, we compute the symmetrized Kullback-Leibler (KL) distance between the empirical probabilities and the simulated probabilities generated by the model from the same empirical centroids. This is a stringent criterion to ensure we get the best possible fit. Indeed, the optimal PMS fitting of the empirical data is obtained at the minimum of *G*=0.245, which is a significantly better fit than the FCD fitting (**Figure 3C**, lower figure). Furthermore, the transition probabilities are also significantly better fitted as measured with Markov-Entropy distance (**Figure 3C**, upper figure). Please note that values of the diagonal of the transition matrix (i.e the probability of remaining in the same state) are much higher than the probabilities of switching states. This is important as it reflects the metastable character of the substates.

The results show that the PMS based measurements for characterizing a brain state provides the most significantly accurate working point of the whole-brain model to account for the most detailed spatiotemporal dynamical characteristics of the brain activity defining a brain state.

## Optimising whole-brain model using effective connectivity

The third step in modelling the spatiotemporal PMS space is to derive two different optimised whole-brain models for two radically different brain states. Here, as a proof a concept we used two naturally occurring significantly different states, namely wakefulness and deep sleep (Tagliazucchi and Laufs, 2014).

**Figure 4** shows how we improve on the fitting procedure showed above with a single global scaling parameter, by optimising the effectiveness of the synaptic connections between brain regions as specified by the SC (anatomical constrained effective connectivity, ANEC). More concretely, we compute the distance between the model and the empirical grand average phase coherence matrices, and adjust each structural connection separately using a gradient-descent approach (see Methods sections for details). The model is run repeatedly with the updated effective connectivity until convergence (see Methods).

**Figure 5** shows the results obtained for wakefulness (left column) and sleep (right column) with the fit of the whole-brain with ANEC and the empirical results. In each row, we show the transition probability matrices, the probability state space, effective connectivity matrix and the effective degree. As can be seen, there is an excellent fit between model and empirical data for both brain states (e.g. compare the histograms in row 2 for the probability state spaces). Also note how different the effective connectivity matrices are for the two states (bottom two rows).

## Awakening: forcing the transition

The previous results show that we can successfully create two whole-brain models with excellent fit to the empirical fMRI data from two radically different brain states. Still, the most important finding is to show that it is possible to force a transition from one whole-brain model to another via external stimulation.

**Figure 6** summarises our approach to forcing a transition between two brain states (source and target). We systematically perturb the brain regions in the whole-brain model of the source state and compare the resulting output PMS space for this model with the empirical data for the other target state (see top row of **Figure 6**). Specifically, the Hopf model allows an effective way of perturbing the model by simply changing the bifurcation parameter in a given brain region (Deco *et al.*, 2017a). The stimulation intensity, i.e. the strength of the perturbation, is directly related to the amount of shifting the local bifurcation parameter (see bottom row right, methods and (Deco *et al.*, 2017c)). We perturbed the model bilaterally in **Figure 6** (bottom row left), which shows the levels of brain state transition fitting for perturbing separately each of the 45 regions (since it is bilateral stimulation) with different stimulation intensities in source state (deep sleep). The colour scale of **Figure 6** (bottom row left) shows the level of fitting with the target state (wakefulness), ie. lower values (blue) correspond to an effective transition.

For the main transition results, we used two different protocols for external stimulation, synchronisation and noise, which shifted the local bifurcation parameter to positive and negative values, respectively. **Figure 7** (top row) shows the results for forcing a transition from source state (deep sleep) to target state (wakefulness) using a synchronisation protocol (left panel) where positive values of the local bifurcation parameter force local oscillations that promote the possibility of more synchronisation across the whole brain. The colour scale indicates the KL distance between source and target state with lower values indicating a better fit. This promotes a transition from deep sleep to wakefulness when perturbing most brain regions with sufficient stimulation intensity (a=0.08). In the right panel we show the ability of brain regions to promote transition at this stimulation intensity. It is clear from the figures that while many regions are able to promote a transition (given sufficient stimulation), other regions are less suitable for this (see grey areas). Importantly, in **Figure 8** (top row) we show that using the noise protocol to force a transition from deep sleep to wakefulness is not possible, with an increase in stimulation intensity leading to higher KL distances, i.e. poorer fit, indicated by an increase in the colours to more yellow from blue. Please also note that the colour scale is different between Figure 7 and 8, with the first column in each figure having the identical numerical KL distances (corresponding to the non-perturbation case) but appearing in different colours due to different colour scales.

**Figure 8** (bottom row) shows the results for forcing a transition from source state (wakefulness) to target state (deep sleep) using a noise protocol (left panel), where negative values of the local bifurcation parameter force local oscillations that promote the possibility of more noise and less synchronisation across the whole brain. The results show more specificity for making the wakeful brain move to deep sleep than for the inverse, with the right panel showing the ability of brain regions to promote transition at the stimulation intensity of a=-0.04 (note the increase in grey areas). Importantly, **Figure 8** (bottom row) shows that using the synchronisation protocol does not in any case result in a transition from wakefulness to deep sleep (note the increase in KL distance). This could be interpreted that is probably much easier, and thus unspecific, to promote a transition from sleep to awake than viceversa.

Finally, we explored whether stimulating multiple regions with weaker stimulation intensity would produce equal or better results. **Figure 9** shows the results of using this multi-site stimulation protocol using a greedy strategy (Deco *et al.*, 2017e) to find the best combination of multiple brain regions to force a transition between states (deep sleep to wakefulness). Using the synchronisation protocol but at the weaker stimulation intensity of a=0.02, we identify the region that best fit the target PMS space and let this region continue to be stimulated while we look for the best region among the rest in this new condition. The process was iterated over seven steps. As shown in **Figure 9A,** the combination of multisite stimulation reaches its best fit (comparable to the best fit for single-site stimulation at higher stimulation intensity, Figure 7 top row) using four bilateral stimulated regions (Front Mid, Temporal Inf, Frontal Sup and Precuneus) and then starts to get worse when more regions are added. This is shown by the black line which indicates the level of fit to the target PMS space (wakefulness) and reaches a minimum for four regions. On the other hand, the red line indicates the level of fit to the source PMS space (deep sleep) and gets monotonically worse with more stimulation sites. Finally, in order to better understand how the multisite stimulation promote transition, in **Figure 9B**, we plot the evolution of PMS 1 and 2 in the source (deep sleep) and target (wakefulness) states as a function multiregion stimulation. The black line represents the lifetime of state 1, which increases with number of regions stimulated, while red line representing lifetime of state 2 decreases. This nicely fits the transition between the lifetimes of the PMS in source and target states, shown on the right, with the optimal balance for lifetimes found with four stimulated regions.

# Methods

## Experimental Data

We based our research on BOLD fMRI data recorded in Frankfurt (Germany) where participants fell asleep during a simultaneous EEG-fMRI scanning session. The experimental results were already published and described in detail in previous publications. We describe here just a brief summary detailing technical and experimental aspects of their acquisition and pre-processing. For this study, we only considered the subset of subjects who reached deep sleep (stage N3).

*EEG-fMRI Recordings*

From a total of fifty-five subjects (thirty-six females, mean±SD age of 23.4±3.3 years) who fell asleep during a simultaneous EEG-fMRI recording previously described in Tagliazucchi and Laufs (2014), we selected the 18 subjects who reached stage N3 sleep (deep sleep). The mean duration (± standard deviation) of contiguous N3 sleep epochs for these participants was 11.67 ± 8.66 minutes. The fMRI data was recorded at 3T (Siemens Trio, Erlangen, Germany) simultaneously with EEG data using an MR-compatible EEG cap (modified BrainCapMR, Easycap, Herrsching, Germany). Sleep stages were scored manually by an expert according to the AASM criteria (AASM, 2007). fMRI data was realigned, normalized and spatially smoothed using SPM8 (www.fil.ion.ucl.ac.uk/spm). Cardiac, respiratory, and motion-induced noise were regressed out from the fMRI BOLD signals (Glover *et al.*, 2000) and data was band-pass filtered in the range 0.01-0.1 Hz (Cordes *et al.*, 2001). Please see Tagliazucchi and Laufs (2014) for full acquisition, pre-processing and sleep scoring details.

### fMRI Pre-processing

For each participant and for each brain state (i.e. wakefulness and deep sleep), we used FSL tools to extract and average the BOLD signals from all voxels within each ROI defined in the AAL atlas (considering only the 90 cortical and subcortical non-cerebellar brain regions) (Tzourio-Mazoyer *et al.*, 2002).

Pairwise Pearson correlation between all 90 regions was computed resulting in a 90x90 functional connectivity (FC) matrix for each participant and brain state. Correlation values were converted to z-values applying Fisher’s transform before averaging across participants in the same cohort, resulting in a 90x90 FC matrix for each brain state (rest and sleep).

### Structural Connectivity

In the whole-brain network model, the interactions between the 90 brain areas were scaled in proportion to their white matter structural connectivity. For the present study, we used the structural connectivity between the 90 AAL regions obtained in a previous study (Deco *et al.*, 2017b) averaged across 16 healthy young adults (5 females, mean±SD age: 24.75±2.54). Briefly, for each subject, a 90x90 structural connectivity matrix *C* was obtained by applying tractography algorithms to Diffusion Tensor Imaging (DTI) following the same methodology described in Cabral *et al.* (2012) where the connectivity *Cnp* between regions *n* and *p* is calculated as the proportion of sampled fibres in all voxels in region *n* that reach any voxel in region *p*. Since DTI does not capture fiber directionality, *Cnp* was defined as the average between *Cnp* and *Cpn*. Averaging across all 16 participants resulted in a structural connectivity matrix *C* representative of healthy young adults.

## Leading Eigenvector Dynamics Analysis (LEiDA):

First, we calculated a phase coherence matrix to capture the amount of interregional BOLD signal synchrony at each time point, for all subject and conditions (awake and N3 deep sleep). The phase coherence between each pair of nodes is given by:

(1)

where the BOLD phases at the node n, (n,t), is estimated using the Hilbert transform for each BOLD regional time course. The Hilbert transform expresses a given signal x in polar coordinates as x(t)=A(t)\*cos((t)). Using the cosine function, two areas n and p with temporarily aligned BOLD signals (i.e. with similar angles) at a given TR will have a phase coherence value dFC(n,p,t) close to 1 (since cos(0°)=1). On the other hand, time points where the BOLD signals are orthogonal (for instance, one increasing at 45° and the other decreasing at 45°) will have dFC(n,p,t) close to 0 (since cos(90°) = 0). The resulting dFC(t) for each subject in each condition is a thus three-dimensional matrix with size NxNxT, where N=90 is the number of brain areas and T is the total number of time points (which in our case is different for each subject and each condition). Note that the phase coherence matrix is undirected and, as such, for each time t, the NxN dFC(t) matrix is symmetric across the diagonal. In order, to characterize the evolution of the dFC matrix over time, we reduced the dimensionality of the problem by focusing on the evolution of the leading eigenvectors of the dFC matrices. The leading eigenvector, V1(t), is a Nx1 vector that captures the dominant connectivity pattern of the dFC(t) at each time t, given in matrix format by its outer product V1.V1T. Moreover, this approach substantially reduces the dimensionality of the data when compared to more traditional approaches considering all the values in the NxN FCt(t) connectivity matrix (Hansen *et al.*, 2015; Hutchison *et al.*, 2013; Preti *et al.*, 2016). For further details about LEiDA, the reader is invited to consult the work of Cabral and colleagues (Cabral *et al.*, 2017b; Cabral *et al.*, 2017c).

Upon computing the leading eigenvector of the phase coherence matrix dFC(t) for each TR, the next step in our analysis was to identify recurrent FC patterns in the data (‘metastable substates’). A discrete number of FC patterns was detected by clustering the leading eigenvectors V1(t) from the collapsed pre/post psilocybin fMRI data (2 conditions) including all subjects. The k-means clustering algorithm was run with k from 2 to 20. Clustering solutions output a k number of cluster centroids in the shape of Nx1 Vc vectors, which represent recurrent metastable substate FC patterns. The outer product of each cluster centroid vector VcVcT is a NxN matrix representing the dominant connectivity pattern and the elements of Vc weight the contribution of each brain area to the community structure established by the signs of the corresponding vector elements. To facilitate visualization and interpretation of FC-states, the cluster centroid vectors Vc were rendered onto a cortical surface using HCP Workbench. For our analysis, the optimal number of clusters according many criteria (Silhouette, minimal p value for significant differences between probabilities between conditions) were k=3.

Upon identifying metastable substates, we computed the probability of occurrence of each FC state in each condition. The probability of occurrence (or fractional occupancy) is simply the ratio of the number of epochs assigned to a given cluster centroid Vc divided by the total number of epochs (TRs) in each experimental condition (which is the same in all experimental conditions). The probabilities were calculated for each subject, in each experimental conditional and for the whole range of clustering solutions explored.

In addition, we computed the switching matrix, which captures the trajectories of FC dynamics in a directional manner. In more detail, it indicates the probability of, being in a given FC state (lines), transitioning to any of the other FC states (columns). Differences in probabilities of occurrence and probabilities of transition were statistically assessed between-conditions using a permutation-based paired t-test. This non-parametric test uses permutations of group labels to estimate the null distribution. The null distribution is computed independently for each experimental condition. For each of 1000 permutations a t-test is applied to compare populations and the significance threshold α = 0.05 was used.

## Whole-Brain Computational Model

We simulated the BOLD activity at the whole-brain level using a computational model– the so-called *Hopf* model - which emulates the dynamics emerging from the mutual interactions between brain areas when coupled through the anatomical structural connectivity (Deco *et al.*, 2017d; Kringelbach *et al.*, 2015). The model consists of 90 coupled dynamical units (nodes) representing the 90 cortical and subcortical brain areas from the AAL parcellation explained above. The local dynamics of each brain area (node) is described by the normal form of a supercritical Hopf bifurcation, also called a Landau-Stuart Oscillator, which is the canonical model for studying the transition from noisy to oscillatory dynamics (Kuznetsov, 1998). When coupled together using brain network architectures, the complex interactions between Hopf oscillators have been shown to successfully replicate features of brain dynamics observed in electrophysiology (Freyer *et al.*, 2011; Freyer *et al.*, 2012), MEG (Deco *et al.*, 2017b) and fMRI (Deco *et al.*, 2017d; Kringelbach *et al.*, 2015).

The dynamics of an uncoupled node *n* is given by the following set of coupled dynamical equations, which describes the normal form of a supercritical Hopf bifurcation in Cartesian coordinates:

(1)

(2)

Where is additive Gaussian noise with standard deviation *β*. This normal form has a supercritical bifurcation *an*=0, so that if *an>0* the system engages in a stable limit cycle with frequency *fn=n/2*and for *an<0* the local dynamics is in a stable fixed point representing a low activity noisy state. Within this model, the intrinsic frequency of each node is in the 0.04–0.07Hz band (*n*=1, …, 90). The intrinsic frequencies were estimated from the data, as given by the averaged peak frequency of the narrowband BOLD signals of each brain region.

To model the whole-brain dynamics we added an additive coupling term representing the input received in node *n* from every other node *p*, which is weighted by the corresponding structural connectivity *Cnp*. This input was modeled using the common difference coupling, which approximates the simplest (linear) part of a general coupling function. Thus, the whole-brain dynamics was defined by the following set of coupled equations:

(3)

(4)

Where *G* denotes the global coupling weight, scaling equally the total input received in each brain area. We fixed the noise standard deviation to *β*=0.02 and the mean structural connectivity to <C>=0.2, in order to be in the same range of parameters previously explored in Deco *et al.* (2017d).While the oscillators are weakly coupled, the periodic orbit of the uncoupled oscillators is preserved. Please note that we do not address here the case of non-linear coupling, in which the next non-vanishing higher order term following a Taylor expansion of the full coupling should be considered (Kuramoto, 1984; Pikovsky *et al.*, 2003).The variable *xn* emulates the BOLD signal of each node *n*. The global coupling parameter *G* is the control parameter with which we adjusted the model to the dynamical working region where the simulations optimally fit the empirical data (Deco *et al.*, 2017d; Jobst *et al.*, 2017).

### Empirical Fitting

*Comparing Empirical and Simulated Grand Averaged Static Functional Connectivity (FC):* The comparison was measured by computing the Pearson correlation coefficient between corresponding elements of the upper triangular part of the empirical and simulated grand averaged FC.

*Comparing Empirical and Simulated Functional Connectivity Dynamics (FCD):* We measure Kolmogorov-Smirnov distance between the upper triangular elements of the empirical and simulated FCD matrices (accumulated over all participants). For a single subject session where M time points were collected, the corresponding phase-coherence based FCD matrix is defined as a MxM symmetric matrix whose (t1, t2) entry is defined by the cosine similarity between the upper triangular parts of the two matrices dFC(t1) and dFC(t2) (previously defined, see above). For two vectors **p**1 and **p**2 the cosine similarity is given by (**p**1.**p**2)/(||**p**1||||**p**2||). Epochs of stable FC(t) conﬁgurations are reflected around the FCD diagonal in blocks of elevated inter-FC(t) correlations. The Kolmogorov–Smirnov [distance](http://en.wikipedia.org/wiki/Metric_(mathematics)) quantifies the maximal difference between the [cumulative distribution function](http://en.wikipedia.org/wiki/Cumulative_distribution_function)s of the two samples.

*Comparing Empirical and Simulated Probability Metastable Space State (PMSS) measurements:* For comparing the probabilities of the metastable states, i.e. the probabilities of the extracted empirical centers after clusterization, we used a symmetrized Kullback-Leibler distance between the simulated and empirical corresponding probabilities, i.e.:

(5)

Where *Pemp(i)* and *Psim(i)* are the empirical and simulated probabilities on the same empirical extracted metastable substates i.

*Comparing Empirical and Simulated Transition Probabilities Between Metastable States:* We calculated the entropy rate of a Markov chain, with states and transition matrix . The rate entropy is given by:

(6)

where:

(7)

The probability represents the stationary probability of state . For long realizations of the Markov chain, the probabilities of each state converge to the stationary distribution , which is solution of the following equation:

(8)

Thus, the stationary distribution is the eigenvector of the transpose of the transition matrix with associated eigenvalue equal to 1. A Markov model that makes a lot of transitions has a large rate entropy, while, a Markov model that barely transits has low entropy. For each transition matrix we obtained the stationary distribution and, then, calculated the entropy rate. The final measure comparing the two transition probability matrix it is just defined by the absolute value of the difference between both respective Markov entropy.

## Methods for Anatomically Constrained Effective Connectivity (ANEC)

Xx thus, we derived two whole-brain models, namely one for accounting the awake condition and another for accounting the N3 deep sleep condition. In both cases, we optimized the effectiveness of synaptic connections between brain areas (effective connectivity, EC). More concrete, we compute the distance between the model and empirical grand averaged phase coherence <dFC> matrices, and adjust each structural connection separately with a gradient-descent approach (see method sections for details), thereby transforming structural into effective connections. The model is run repeatedly with the updated effective connectivity until the fit converges towards a stable value.

herehow I do it....

greedy algorithm.

        for i=1:N

            for j=i+1:N

                if (C(i,j)>0 || j==N-i+1)

                    Cnew(i,j)=Cnew(i,j)+0.01\*(FCphasesemp(i,j)-FCphases(i,j));

                    if Cnew(i,j)<0

                        Cnew(i,j)=0;

                    end

                    Cnew(j,i)=Cnew(i,j);

                end

            end

        end

where FCphases is:

        for t=1:nn

            for n=1:N

                for p=1:N

                    iFC(t,n,p)=cos(Phase\_BOLD(n,t)-Phase\_BOLD(p,t));

                end

            end

        end

        FCphases=squeeze(mean(iFC));

and the same for the empirical....FCphasesemp

for nsub=subjects

    for t=1:Tmax

        for n=1:N

            for p=1:N

                iFC(t,n,p)=cos(Phase\_BOLD\_data(n,t)-Phase\_BOLD\_data(p,t));

            end

        end

    end

    FCphasesemp2(nsub,:,:)=squeeze(mean(iFC));

end

FCphasesemp=squeeze(mean(FCphasesemp2));

# Discussion

The history of physics has shown that a deep understanding of the causative underlying mechanisms of any system requires the ability to be able to actively force a transition from a given state of the system to another state (xx Feynman). During their comparatively shorter lifespans, neuroscience and more recently neuroimaging have mostly been focused on finding the best ways of recording and interpreting empirical data (Raichle and Snyder, 2007) but recently the focus has started shifted from mere phenomenology to causal understanding (Breakspear, 2017; Deco and Kringelbach, 2014; Deco *et al.*, 2017e).

### Brain states

Here, we have contributed to this emerging field of enquiry by demonstrating how to force a transition from one brain state to another. In order to this, we first show a quantitative definition of brain state from fMRI data acquired during two radically different states (wakefulness and deep sleep). This reveals the dynamical repertoire of these states as probabilistic descriptions of an ensemble of possible ‘metastable substates’. This *Probabilistic Metastable Substates (PMS) space* was then fitted to a whole-brain model, which was systematically perturbed *in silico* to predict exactly where and how to force a transition between brain states. Here, as a proof of concept, we have shown how best to awaken a sleeping brain.

Previous attempts have been made to define a brain state in a number of ways, from a point in a state space characterizing the activity of the brain at a given time (Gervasoni *et al.*, 2004; Gu *et al.*, 2017; Gu *et al.*, 2018; Muldoon *et al.*, 2016) to an attractor of interacting brain regions (Deco and Jirsa, 2012; Gu *et al.*, 2018). Other approaches have included the whole-brain connectomic measurements including dynamic functional connectivity (Allen *et al.*, 2014; Hansen *et al.*, 2015; Keilholz *et al.*, 2017; Preti *et al.*, 2017). However, these approaches do not fully capture the spatiotemporal richness of the neuroimaging data.

Instead, here we have shown that a more fundamental description of a brain state capturing the full nature of the data as an ensemble or probabilistic ‘cloud’ in a state space, which can be decomposed into ‘metastable substates’ (Baker *et al.*, 2014; Cabral *et al.*, 2017d). These are stable but variable as a function of time (Deco and Kringelbach, 2016; Deco *et al.*, 2017c; Tognoli and Kelso, 2014).

### Modelling and transition

We show that whole-brain models can be made to fit the PMS space of two radically different brain states by using an optimisation of the models with the anatomically constrained effective connectivity (ANEC), similar in spirit to Gilson and colleagues (Gilson *et al.*, 2016). In general, the advantage of having a mechanistic whole-brain model is to allow us to study the effect of perturbation off-line. Using this optimisation allows us to tailor a model to fit a specific PMS space, which is important for the studying how to optimally force a transition between brain states. This novel approach allows for an exhaustive search of the underlying parameters and locations *in silico*. This would be very difficult to carry out in animal models and impossible in humans for ethical reasons. Long-term this approach could be extremely helpful for discovering new targets for stimulation not only in existing diseases such as Parkinson’s Disease but also in neuropsychiatric disorders where animal models have been shown to be suboptimal (xx).

As proof of concept, we have shown that it is possible to force a transition in the whole-brain model adjusted to empirical fMRI data capturing two radically different brain states (wakefulness and sleep). As such the research While promising, much research needs to be carried out before this can come to have clinical utility. Nevertheless, this approach opens up for new exciting possibilities for discovering new stimulation targets for DBS and TMS in disease. This stimulation would potentially be able to better rebalance the networks and provide better alleviation of suffering for patients. But more generally, the approach described here could be used as new principled way to rebalance human brain activity in health and disease (Kringelbach *et al.*, 2011) by causally changing the ‘cloud’ of metastable substates from that found in disease to the health in order to promote a profound reconfiguration of the dynamical landscape necessary for recovery (rather than having to identify the original point of insult and repair this).

This approach significantly broadens the findings from previous whole-brain models of simultaneous neuroimaging activity and direct ON-OFF deep brain stimulation alleviating symptoms of brain disease which have revealed trigger points and underlying brain networks (Saenger *et al.*, 2017; Van Hartevelt *et al.*, 2014; van Hartevelt *et al.*, 2015). Similarly, Deco and colleagues have lesioned and perturbed brain regions and networks in whole-brain models *in silico,* demonstrating the causal contribution of brain regions to the ability of the human brain to efficiently integrate information over time (Deco *et al.*, 2017a; Deco *et al.*, 2017e).

Xx ENZO: However it is hard (at least for me) to think of an empirical intervention changing precisely that. Have you considered (perhaps for future studies, or to mention in the discussion) to change the intrinsic frequency of the oscillators (e.g. modeling a tACS intervention) or modulating the amplitude of the model (e.g. modeling a TMS intervention?) Our model here *in silico* is of course artificial, but excellent for proof-of-concept.

Xx ENZO: You mention the description of the brain in phase space as an alternative to define conscious states, and then mention your approach as a different method. I think that in essence your approach is not that difference, and that it all comes down to a question of granularity. You could represent the brain in a phase space consisting of all the parameters needed to describe each electron proton and neutron in all brain atoms. Most likely this description is at a too fine grain to be of any use to define a brain state. Decreasing the granularity you could reach a description involving cell conductances, ion and neurotransmitter concentrations, etc. And then another one encoding cell synaptic connectivity and firing. In these descriptions, the parameters move within a high dimensional space: clustering in that space essentially represents a brain state, i.e. regions of the phase space where the parameters are more likely to reside, and then jump to a different high density region afterwards. In your case the choice of granularity is based on fMRI data at the millimeter scale and large scale anatomical connections. Perhaps the discussion would be enriched with a discussion of why this choice of granularity is good for the purposes of the study. This argument could be made in terms of biological regions (i.e. the most salient changes between brain states are macroscopically measurable using EEG and fMRI), or in terms of technical reasons (it would be difficult or impossible to cluster a very high dimensionality space, as it would be very sparsely populated within the time frame of a typical experiment with humans)

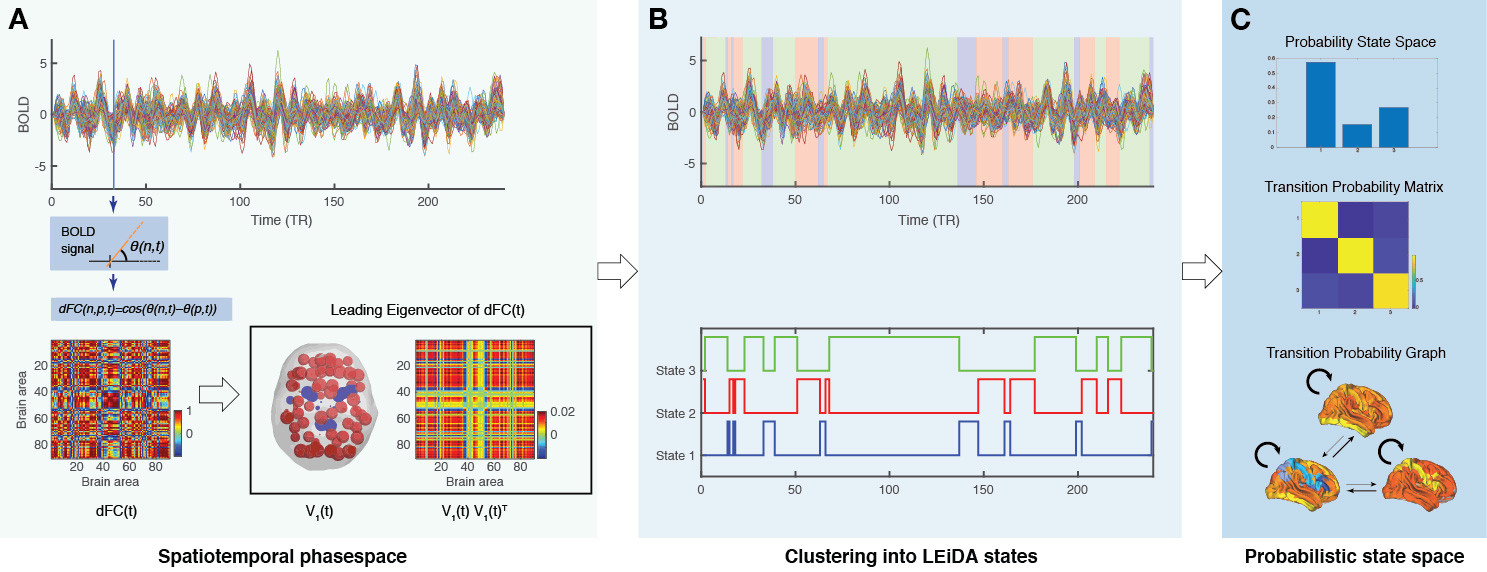
Using the proposed framework, further awakenings can be imagined, perhaps even promoting the “awakening” of locked-in patients. A whole host of stimulation technologies such as DBS and TMS could potentially be used to force such transitions between health and disease. But it might also be possible to use targeted neurotransmission to the same effect, now that whole-brain models have successfully included neurotransmission (Deco *et al.*, 2018).

Overall, the methods and results presented here may eventually allow us to build causative whole-brain models that can characterise all brain states, including levels of consciousness, disease and cognitive states. In particular this could have great clinical utility given that it could provide a principled way of discovering how to force a transition between two brain states.

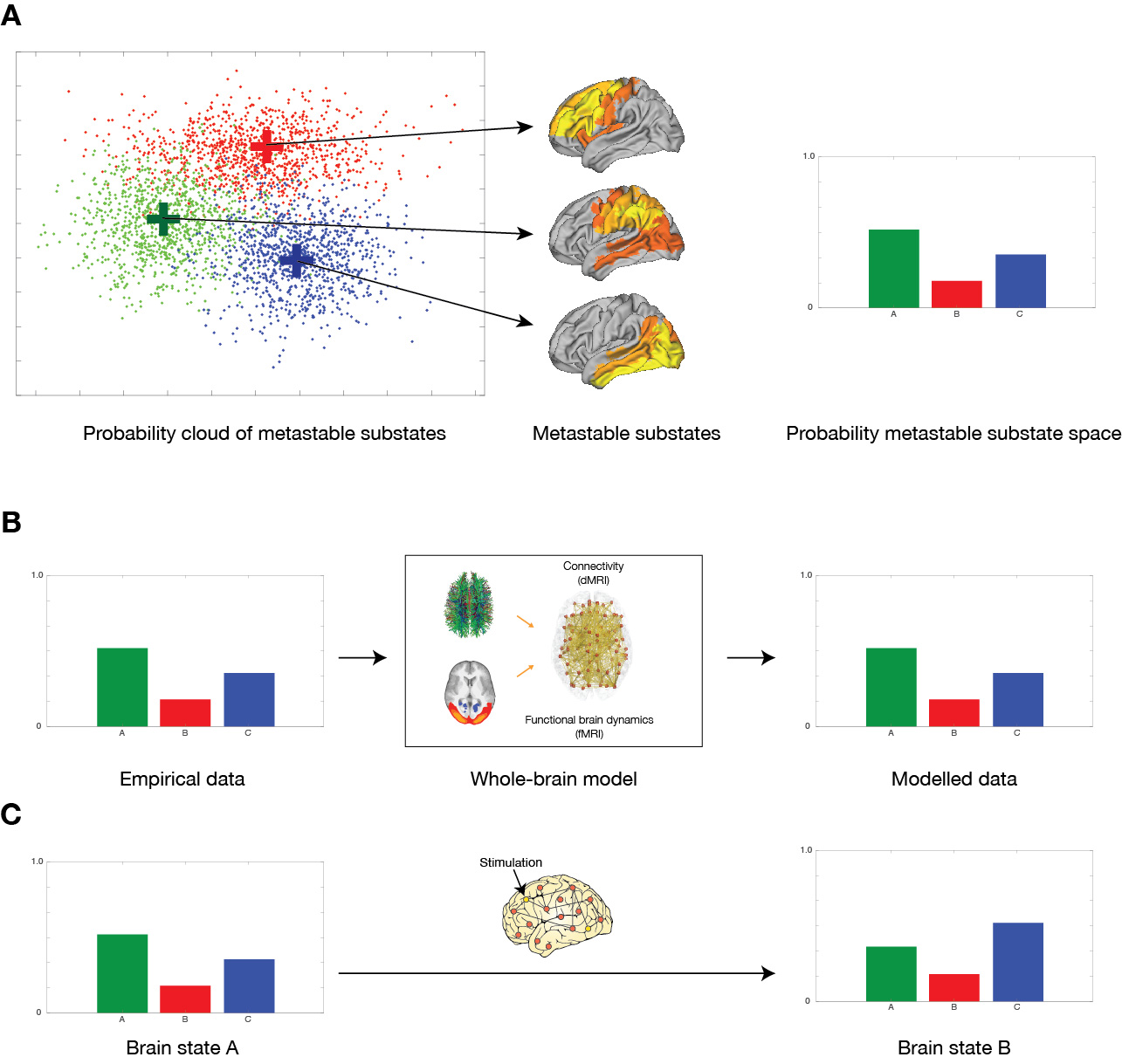
# Acknowledgements

G.D. is supported by the ERC Advanced Grant DYSTRUCTURE (n. 295129), by the Spanish Research Project PSI2016-75688-P, and the European Union’s Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreement No. 785907 (Human Brain Project SGA2). MLK is supported by the ERC Consolidator Grant: CAREGIVING (n. 615539) and Center for Music in the Brain, funded by the Danish National Research Foundation (DNRF117).

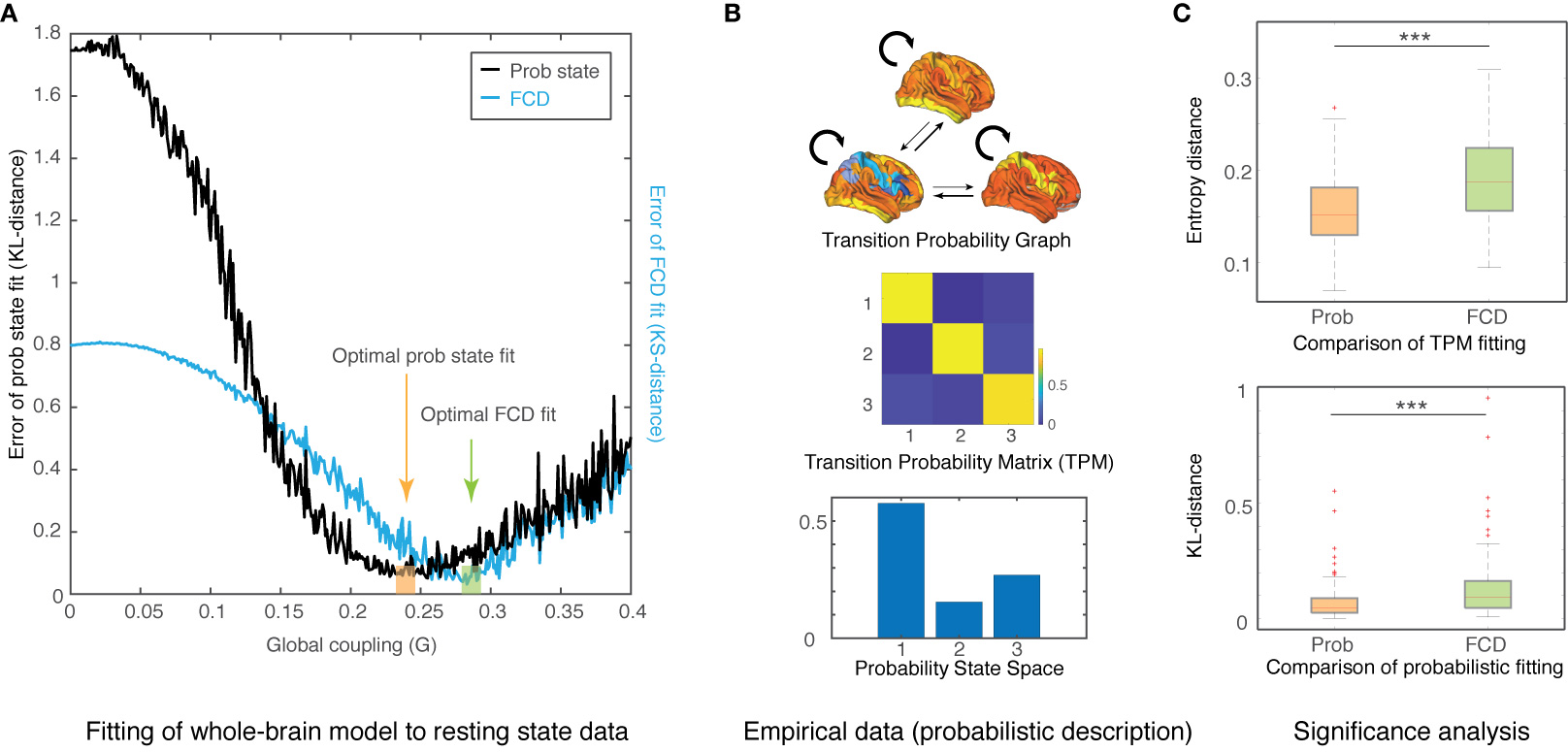
# Figures



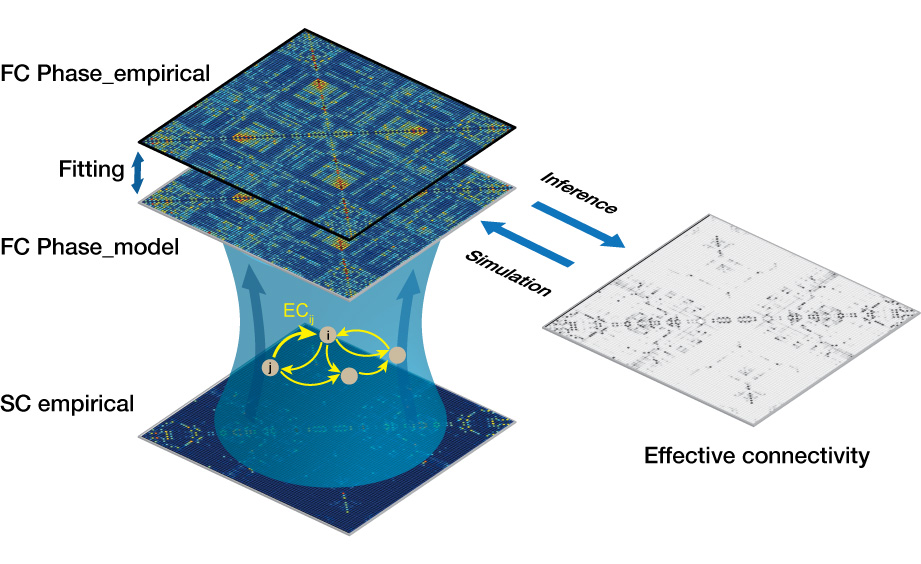
***Figure 1. Decomposing the whole-brain activity into probabilistic state space****.* ***A)*** *How to decompose time series from brain regions into a series of instantaneous snapshots of the whole-brain activity, represented by phase Coherence Connectivity matrices which can be used to obtain a time-resolved dynamic FC (dFC) matrix. From this we can extract the corresponding time-resolved leading eigenvectors used to extract substates.* ***B)*** *After clustering across time points and participants the obtained k-cluster centroids define the ‘metastable substates’.* ***C)*** *As shown, a given brain state can be represented by the probability state space, transition probability matrix and graph.*



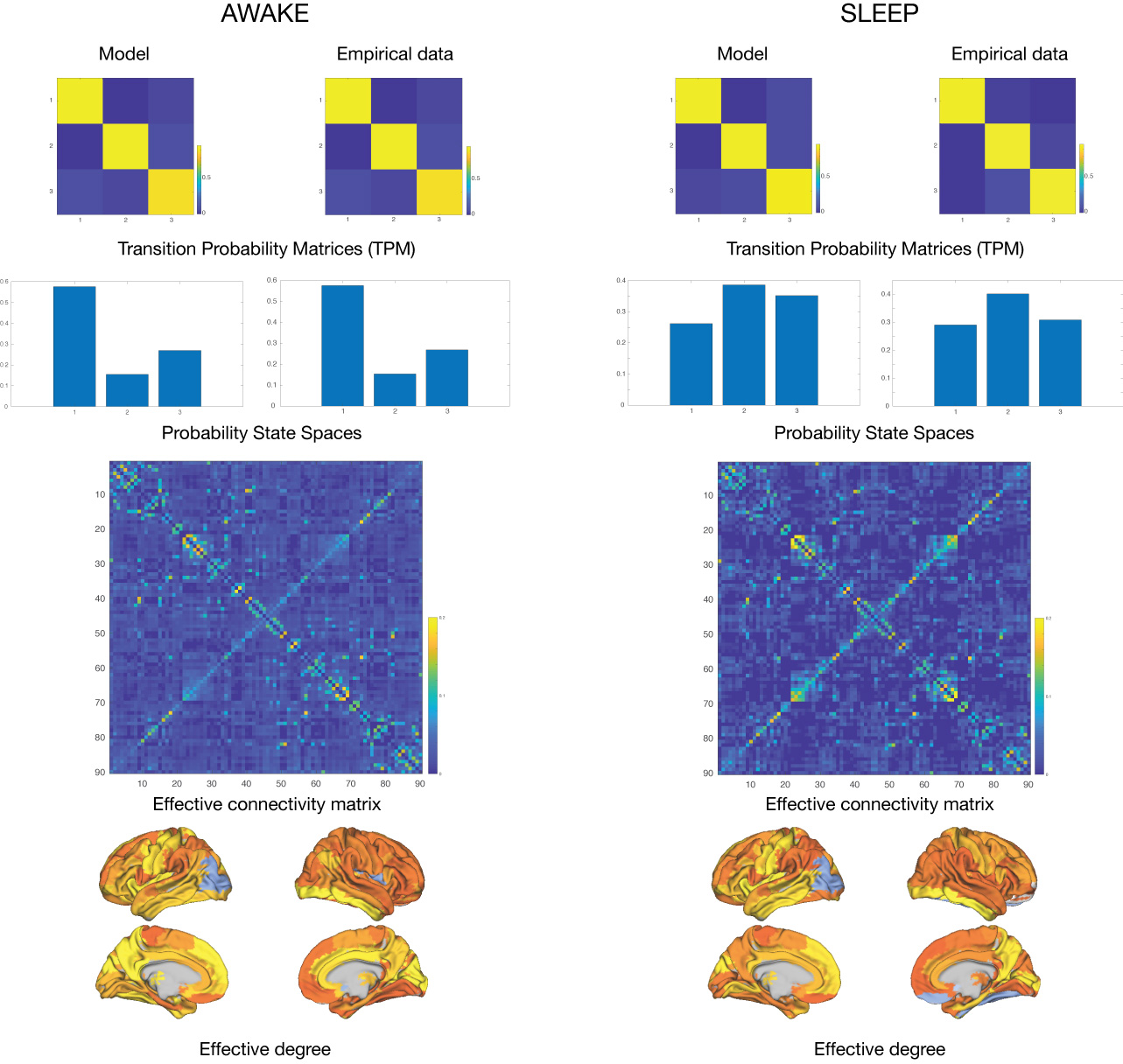
***Figure 2. Schematic of how to force a transition between different brain states.*** *This process involves three steps.* ***A)*** *The spatiotemporal description of the dynamics into Probabilistic Metastable Substates (PMS) space is shown on the two-dimensional projection of the probabilistic ‘cloud’ with each point representing an instantaneous snapshot of the whole-brain. Using clustering, we can force different substates from groups of points (green, red and blue) and express the probability of each of these substates into the Probabilistic Metastable Substates (PMS) space*. ***B)*** *A whole-brain model can be fitted to this PMS space.* ***C)*** *The whole-brain model in one source brain state can be perturbed and stimulated exhaustively in order to promote and force a specific transition to a target brain state.*



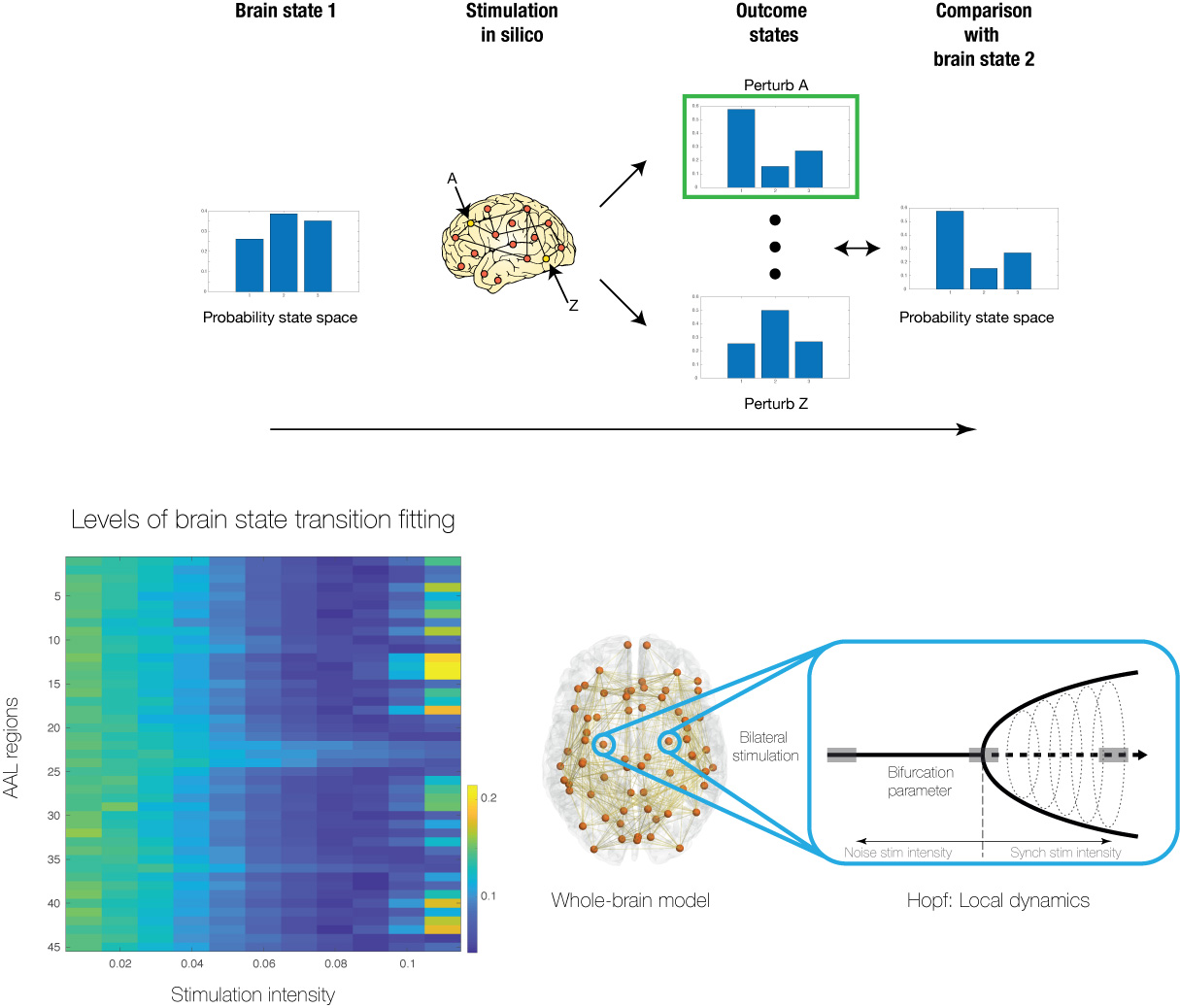
***Figure 3. Results of fitting whole-brain model to probabilistic state space of resting state wakefulness fMRI data****.* ***A)*** *The fit is determined when varying the single global coupling parameter, G, which scales the density of fibres expressed in the SC and can be interpreted as the synaptic conductivity of each single fibre. Here we show two measurements of fitting as a function of G: 1) the phase-coherence-based functional connectivity dynamics (FCD); and 2) the probabilistic metastable substates (PMS). The best fit is found when computing the symmetrized Kullback-Leibler (KL) distance between the empirical probabilities and the simulated probabilities generated by the model from the same empirical centroids with G=0.245.* ***B)*** *We show the transition probability graph, matrix and probability state space for the empirical fMRI resting state with three states.* ***C)*** *We found that the PMS fitting of the empirical data is significantly better than the FCD fitting (lower figure). The transition probabilities are also significantly better fitted as measured with Markov-Entropy distance (upper figure). Please note that values of the diagonal of the transition matrix (i.e the probability of remaining in the same state) are much higher than the probabilities of switching states, reflecting the metastable nature of substates.*



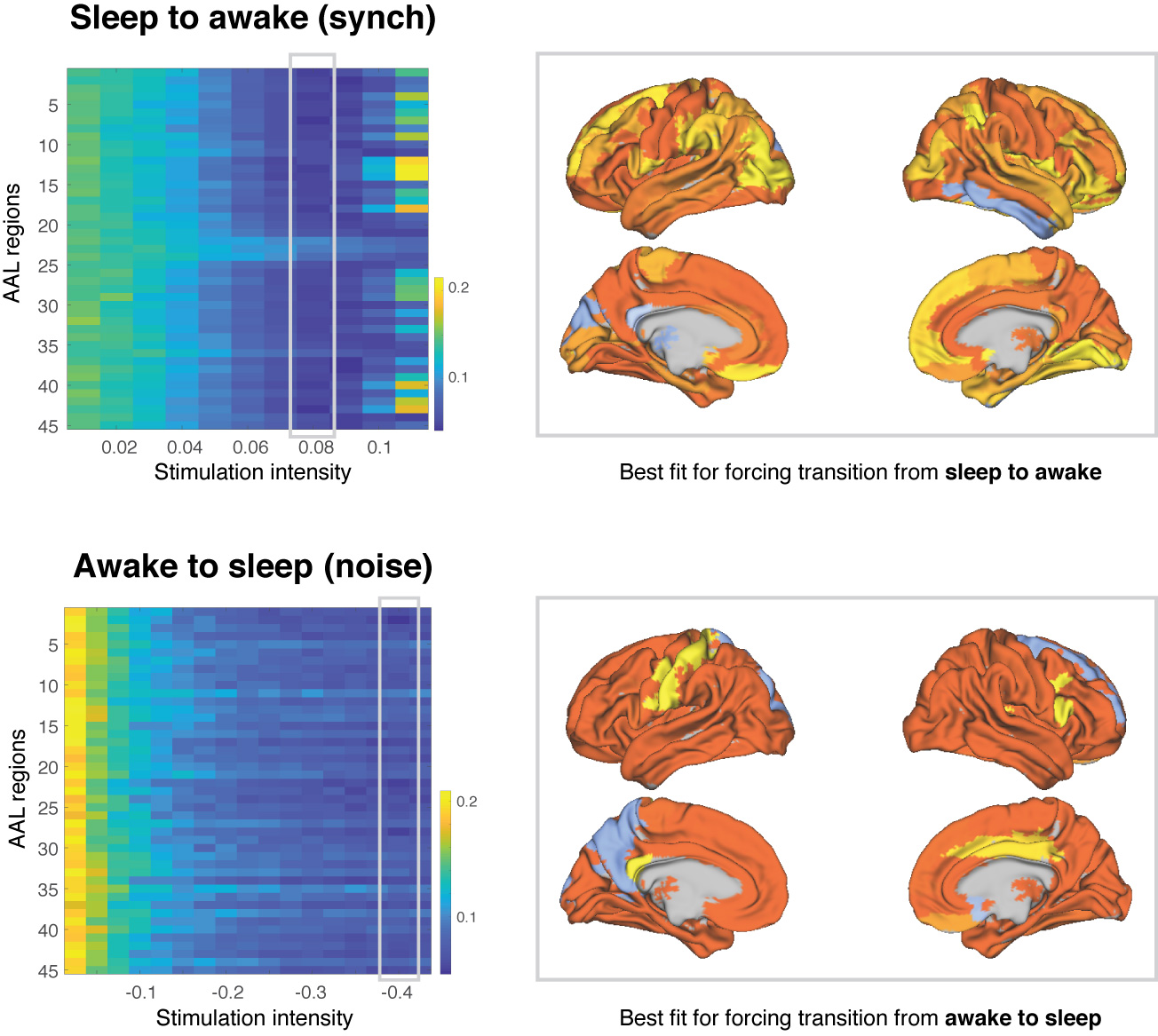
***Figure 4. Fitting whole-brain using anatomical constrained effective connectivity (ANEC).*** *We improved on the fitting procedure by optimising the effectiveness of the synaptic connections between brain regions as specified by the SC. We compute the distance between the model and the empirical grand average phase coherence matrices, and adjust each structural connection separately using a gradient-descent approach (see Methods sections for details). The model is run repeatedly with the updated effective connectivity until convergence.*



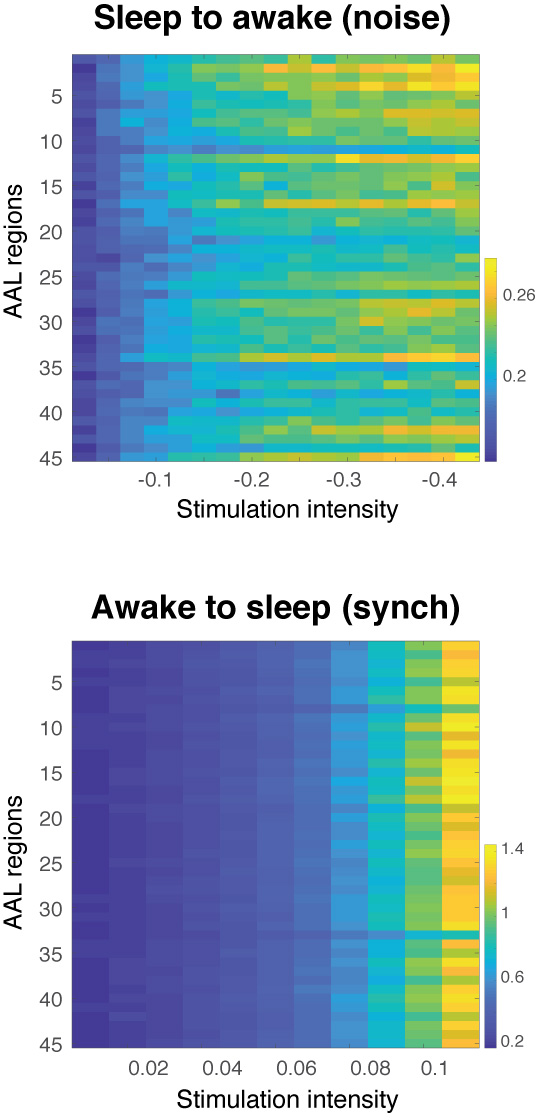
***Figure 5. Fitting the two radically different brain states (wakefulness and sleep) acquired using fMRI.*** *The figure shows the results obtained for wakefulness (left column) and sleep (right column) comparing the empirical results with the output of the fitted whole-brain ANEC model. In each row, we show the transition probability matrices, the probability state space, effective connectivity matrix and the effective degree. As can be seen, there is an excellent fit between model and empirical data for both brain states (e.g. compare the histograms in row 2 for the probability state spaces). Please also note how different the effective connectivity matrices are for the two states (bottom two rows).*



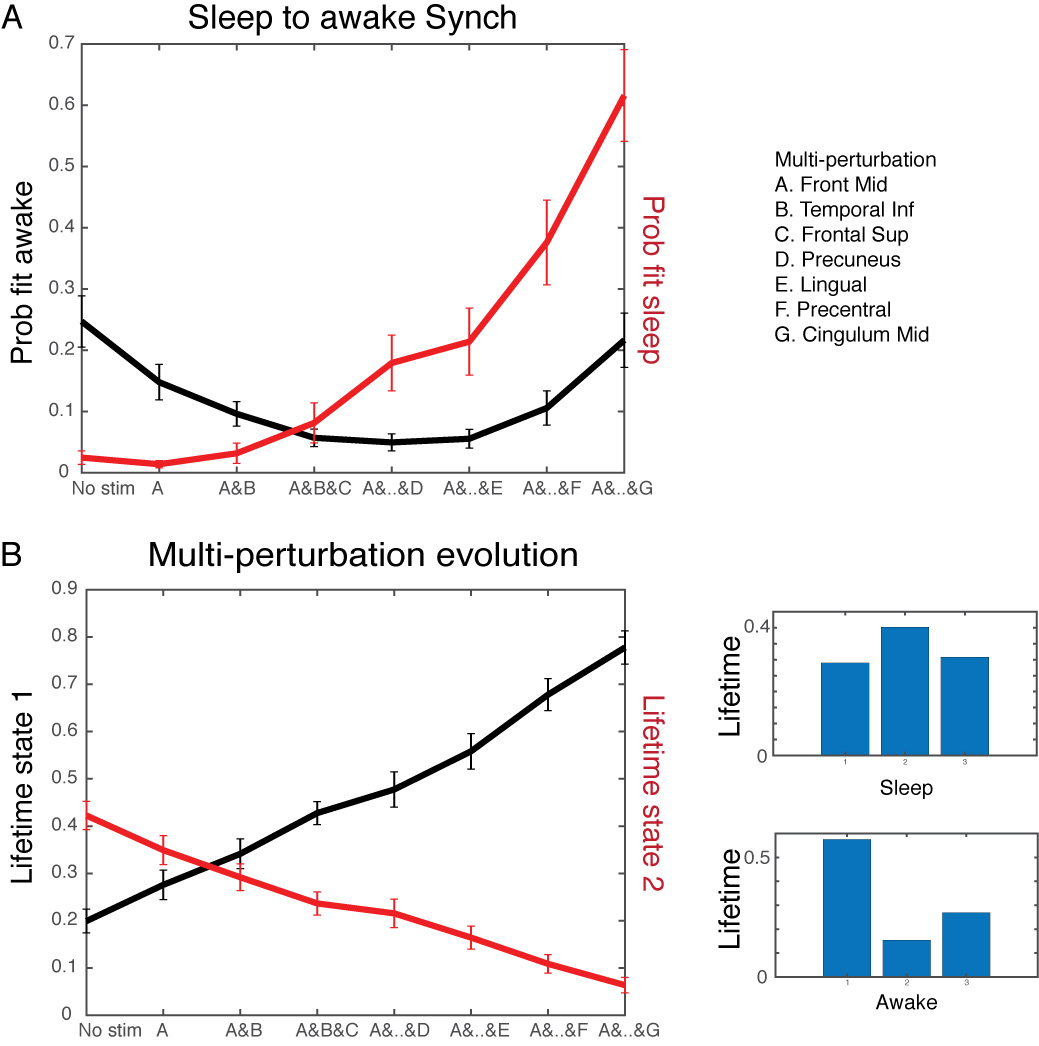
***Figure 6. Schematic of strategy for forcing transition between two source and target brain states****.* ***A)*** *The brain regions in the whole-brain model of the source state can be systematically stimulated and the results can be compared to the target state (top row). Specifically, in the local region Hopf model, it is easy to perturb the model by simply changing the bifurcation parameter* *(Deco et al., 2017a).* ***B)*** *The stimulation intensity, i.e. the strength of the perturbation, is directly related to the amount of shifting the local bifurcation parameter (bottom row, right, see Methods and* *(Deco et al., 2017c)). The results are shown bottom left of stimulating the whole-brain ANEC model bilaterally. We show the levels of brain state transition fitting for perturbing separately each of the 45 regions (for bilateral stimulation) with different stimulation intensities in source state (deep sleep). The colour scale for the results shows the level of fitting with the target state (wakefulness), ie. lower values (blue) correspond to an effective transition.*



***Figure 7. Main results of forcing a transition between brain states.*** *We used two different protocols for external stimulation, synchronisation and noise, which shifted the local bifurcation parameter to positive and negative values, respectively.* ***A)*** *The figure shows the results of forcing a transition from source state (deep sleep) to target state (wakefulness) using a synchronisation protocol (left panel) where positive values of the local bifurcation parameter force local oscillations that promote the possibility of more synchronisation across the whole brain. The colour scale indicates the KL distance between source and target state with lower values indicating a better fit. As can be seen, a transition from deep sleep to wakefulness is promoted when perturbing most brain regions with sufficient stimulation intensity (a=0.08). In the right panel we show the ability of brain regions to promote transition at this stimulation intensity. It is clear from the figures that while many regions are able to promote a transition (given sufficient stimulation), other regions are less suitable for this (see grey areas).* ***B)*** *The figure shows the results of forcing the opposite transition from source state (wakefulness) to target state (deep sleep) using a noise protocol (left panel), where negative values of the local bifurcation parameter force local oscillations that promote the possibility of more noise and less synchronisation across the whole brain. The results show more specificity for making the wakeful brain move to deep sleep than for the inverse, with the right panel showing the ability of brain regions to promote transition at the stimulation intensity of a=-0.04 (note the increase in grey areas).*



***Figure 8. Failure of forcing a transition.*** *We show that transitions are not always possible when using the opposite protocols for forcing transitions than that used in Figure 7.* ***A)*** *In particular, when using the noise protocol to force a transition from deep sleep to wakefulness increases in stimulation intensity lead to higher KL distances, i.e. poorer fit, indicated by an increase in the colours to more yellow from blue. Please also note that the colour scale is different between Figure 7 and 8, with the first column in each figure having the identical numerical KL distances (corresponding to the non-perturbation case) but appearing in different colours due to different colour scales.* ***B)*** *Similarly, it is not possible to force a transition from wakefulness to deep sleep when using the synchronisation protocol, as shown by the increase in KL distance.*



***Figure 9. Stimulation of multiple regions with weaker stimulation intensity can produce equal or better results to single-site stimulation.*** *We used a multi-site stimulation protocol with a greedy strategy to find the best combination of multiple brain regions to force a transition between states (deep sleep to wakefulness). Using the synchronisation protocol but at the weaker stimulation intensity of a=0.02, we identified the region that best fit the target PMS space and let this region continue to be stimulated while we looked for the best region among the rest in this new condition. The process was iterated over seven steps.* ***A)*** *The figure shows how the combination of multisite stimulation reaches its best fit using four bilateral stimulated regions (Front Mid, Temporal Inf, Frontal Sup and Precuneus) and then starts to get worse when more regions are added. The best fit is comparable to the best fit for single-site stimulation at higher stimulation intensity (Figure 7 top row). This is shown by the black line indicating the level of fit to the target PMS space (wakefulness) and reaches a minimum for four regions. On the other hand, the red line indicates the level of fit to the source PMS space (deep sleep) and gets monotonically worse with more stimulation sites.* ***B)*** *To better understand how the multisite stimulation promote transition, the figure plots the evolution of PMS 1 and 2 in the source (deep sleep) and target (wakefulness) states as a function multiregion stimulation. The black line represents the lifetime of state 1, which increases with number of regions stimulated, while red line representing lifetime of state 2 decreases. This nicely fits the transition between the lifetimes of the PMS in source and target states, shown on the right, with the optimal balance for lifetimes found with four stimulated regions.*

# References

AASM (2007) *The AASM manual for the scoring of sleep and associated events-rules. Terminology and technical specifications*. American Academy of Sleep Medicine: Chicago.

Allen, E. A., Damaraju, E., Plis, S. M., Erhardt, E. B., Eichele, T. and Calhoun, V. D. (2014) Tracking whole-brain connectivity dynamics in the resting state. *Cerebral cortex* **24,** 663-676.

Baker, A. P., Brookes, M. J., Rezek, I. A., Smith, S. M., Behrens, T., Probert Smith, P. J. and Woolrich, M. (2014) Fast transient networks in spontaneous human brain activity. *eLife* **3,** e01867.

Breakspear, M. (2017) Dynamic models of large-scale brain activity. *Nature neuroscience* **20,** 340-352.

Cabral, J., Kringelbach, M. L. and Deco, G. (2012) Functional graph alterations in schizophrenia: a result from a global anatomic decoupling? *Pharmacopsychiatry* **45 Suppl 1,** S57-64.

Cabral, J., Kringelbach, M. L. and Deco, G. (2017a) Functional connectivity dynamically evolves on multiple time-scales over a static structural connectome: Models and mechanisms. *Neuroimage* **160,** 84-96.

Cabral, J., Vidaurre, D., Marques, P., Magalhaes, R., Moreira, P., Soares, J., Deco, G., Sousa, N., Kringelbach, M. L. and Press, t. S. R. I. (2017b) Cognitive performance in healthy older adults relates to spontaneous switching between states of functional connectivity during rest. *Scientific reports* **In press**.

Cabral, J., Vidaurre, D., Marques, P., Magalhães, R., Moreira, P. S., Soares, J. M., Deco, G., Sousa, N. and Kringelbach, M. L. (2017c) Cognitive performance in healthy older adults relates to spontaneous switching between states of functional connectivity during rest. *Scientific Reports* **7,** 5135.

Cabral, J., Vidaurre, D., P., M., R., M., P., M., Soares, J. M., G., D., Sousa, N. and Kringelbach, M. L. (2017d) Cognitive performance in healthy older adults linked to spontaneous switching between states of functional connectivity during rest. *Scientific Reports* **7,** 5135.

Cordes, D., Haughton, V. M., Arfanakis, K., Carew, J. D., Turski, P. A., Moritz, C. H., Quigley, M. A. and Meyerand, M. E. (2001) Frequencies contributing to functional connectivity in the cerebral cortex in "resting-state" data. *AJNR. American journal of neuroradiology* **22,** 1326-1333.

Deco, G., Cabral, J., Saenger, V. M., Boly, M., Tagliazucchi, E., Laufs, H., Van Someren, E., Jobst, B., Stevner, A. and Kringelbach, M. L. (2017a) Perturbation of whole-brain dynamics in silico reveals mechanistic differences between brain states. *Neuroimage* **169,** 46-56.

Deco, G., Cabral, J., Woolrich, M. W., Stevner, A. B. A., van Hartevelt, T. J. and Kringelbach, M. L. (2017b) Single or multiple frequency generators in on-going brain activity: A mechanistic whole-brain model of empirical MEG data. *NeuroImage* **152,** 538-550.

Deco, G., Cruzat, J., Cabral, J., Knudsen, G. M., Carhart-Harris, R. L., Whybrow, P. C., Logothetis, N. K. and Kringelbach, M. L. (2018) Whole-brain multimodal neuroimaging model links human anatomy, function and neuromodulation: serotonin receptor maps explain non-linear functional effects of LSD. *Current Biology***,** in press.

Deco, G., Jirsa, V., McIntosh, A. R., Sporns, O. and Kotter, R. (2009) Key role of coupling, delay, and noise in resting brain fluctuations. *Proceedings of the National Academy of Sciences of the United States of America* **106,** 10302-10307.

Deco, G. and Jirsa, V. K. (2012) Ongoing cortical activity at rest: criticality, multistability, and ghost attractors. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **32,** 3366-3375.

Deco, G. and Kringelbach, M. L. (2014) Great Expectations: Using Whole-Brain Computational Connectomics for Understanding Neuropsychiatric Disorders. *Neuron* **84,** 892-905.

Deco, G. and Kringelbach, M. L. (2016) Metastability and Coherence: Extending the Communication through Coherence Hypothesis Using A Whole-Brain Computational Perspective. *Trends in neurosciences* **39,** 125-135.

Deco, G., Kringelbach, M. L., Jirsa, V. and Ritter, P. (2017c) The Dynamics of Resting Fluctuations in the Brain: Metastability and its Dynamical Core [bioRxiv 065284]. *Scientific Reports* **7,** 3095.

Deco, G., Kringelbach, M. L., Jirsa, V. K. and Ritter, P. (2017d) The dynamics of resting fluctuations in the brain: metastability and its dynamical cortical core. *Scientific reports* **7,** 3095.

Deco, G., Tononi, G., Boly, M. and Kringelbach, M. L. (2015) Rethinking segregation and integration: contributions of whole-brain modelling *Nat Rev Neurosci* **16,** 430-439.

Deco, G., Van Hartevelt, T., Fernandes, H. M., Stevner, A. B. A. and Kringelbach, M. L. (2017e) The most relevant human brain regions for functional connectivity: Evidence for a dynamical workspace of binding nodes from whole-brain computational modelling. *Neuroimage* **146,** 197-210.

Dehaene, S. (2014) *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts*. Viking Press: New York.

Freyer, F., Roberts, J. A., Becker, R., Robinson, P. A., Ritter, P. and Breakspear, M. (2011) Biophysical mechanisms of multistability in resting-state cortical rhythms. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **31,** 6353-6361.

Freyer, F., Roberts, J. A., Ritter, P. and Breakspear, M. (2012) A canonical model of multistability and scale-invariance in biological systems. *PLoS computational biology* **8,** e1002634.

Friston, K. J., Harrison, L. and Penny, W. (2003) Dynamic causal modelling. *NeuroImage* **19,** 1273-1302.

Gervasoni, D., Lin, S. C., Ribeiro, S., Soares, E. S., Pantoja, J. and Nicolelis, M. A. (2004) Global forebrain dynamics predict rat behavioral states and their transitions. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **24,** 11137-11147.

Ghosh, A., Rho, Y., McIntosh, A. R., Kotter, R. and Jirsa, V. K. (2008) Noise during rest enables the exploration of the brain's dynamic repertoire. *PLoS computational biology* **4,** e1000196.

Gilson, M., Moreno-Bote, R., Ponce-Alvarez, A., Ritter, P. and Deco, G. (2016) Estimation of Directed Effective Connectivity from fMRI Functional Connectivity Hints at Asymmetries of Cortical Connectome. *PLoS computational biology* **12,** e1004762.

Glover, G. H., Li, T. Q. and Ress, D. (2000) Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine* **44,** 162-167.

Gu, S., Betzel, R. F., Mattar, M. G., Cieslak, M., Delio, P. R., Grafton, S. T., Pasqualetti, F. and Bassett, D. S. (2017) Optimal trajectories of brain state transitions. *NeuroImage* **148,** 305-317.

Gu, S., Cieslak, M., Baird, B., Muldoon, S. F., Grafton, S. T., Pasqualetti, F. and Bassett, D. S. (2018) The Energy Landscape of Neurophysiological Activity Implicit in Brain Network Structure. *Scientific reports* **8,** 2507.

Hansen, E. C., Battaglia, D., Spiegler, A., Deco, G. and Jirsa, V. K. (2015) Functional connectivity dynamics: modeling the switching behavior of the resting state. *NeuroImage* **105,** 525-535.

Honey, C. J., Kotter, R., Breakspear, M. and Sporns, O. (2007) Network structure of cerebral cortex shapes functional connectivity on multiple time scales. *Proceedings of the National Academy of Sciences of the United States of America* **104,** 10240-10245.

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. and Chang, C. (2013) Dynamic functional connectivity: promise, issues, and interpretations. *NeuroImage* **80,** 360-378.

James, W. (1890) *The Principles of Psychology*. Henry Holt: New York.

Jobst, B. M., Hindriks, R., Laufs, H., Tagliazucchi, E., Hahn, G., Ponce-Alvarez, A., Stevner, A. B. A., Kringelbach, M. L. and Deco, G. (2017) Increased Stability and Breakdown of Brain Effective Connectivity During Slow-Wave Sleep: Mechanistic Insights from Whole-Brain Computational Modelling. *Scientific reports* **7,** 4634.

Keilholz, S., Caballero-Gaudes, C., Bandettini, P., Deco, G. and Calhoun, V. (2017) Time-Resolved Resting-State Functional Magnetic Resonance Imaging Analysis: Current Status, Challenges, and New Directions. *Brain connectivity* **7,** 465-481.

Kringelbach, M. L., Green, A. L. and Aziz, T. Z. (2011) Balancing the brain: resting state networks and deep brain stimulation. *Frontiers Integrative Neuroscience* **5,** 8.

Kringelbach, M. L., Jenkinson, N., Owen, S. L. F. and Aziz, T. Z. (2007) Translational principles of deep brain stimulation. *Nature Reviews Neuroscience* **8,** 623-635.

Kringelbach, M. L., McIntosh, A. R., Ritter, P., Jirsa, V. K. and Deco, G. (2015) The rediscovery of slowness: exploring the timing of cognition. *TICS* **19,** 616-628.

Kuramoto, Y. (1984) Chemical Oscillations,Waves, and Turbulence. *Springer-Verlag, Berlin*.

Kuznetsov, Y. A. (1998) *Elements of applied bifurcation theory*. Springer: New York.

Laumann, T. O., Snyder, A. Z., Mitra, A., Gordon, E. M., Gratton, C., Adeyemo, B., Gilmore, A. W., Nelson, S. M., Berg, J. J., Greene, D. J., McCarthy, J. E., Tagliazucchi, E., Laufs, H., Schlaggar, B. L., Dosenbach, N. U. and Petersen, S. E. (2016) On the Stability of BOLD fMRI Correlations. *Cerebral cortex*.

Muldoon, S. F., Pasqualetti, F., Gu, S., Cieslak, M., Grafton, S. T., Vettel, J. M. and Bassett, D. S. (2016) Stimulation-Based Control of Dynamic Brain Networks. *PLoS computational biology* **12,** e1005076.

Pascual-Leone, A., Rubio, B., Pallardo, F. and Catala, M. D. (1996) Rapid-rate transcranial magnetic stimulation of left dorsolateral prefrontal cortex in drug-resistant depression. *Lancet* **348,** 233-237.

Pikovsky, A., Rosenblum, M. and Kurths, J. (2003) *Synchronization: a universal concept in nonlinear sciences*. Cambridge University Press: Cambridge.

Preti, M. G., Bolton, T. A. and Van De Ville, D. (2017) The dynamic functional connectome: State-of-the-art and perspectives. *NeuroImage* **160,** 41-54.

Preti, M. G., Bolton, T. A. and Ville, D. V. (2016) The dynamic functional connectome: State-of-the-art and perspectives. *NeuroImage*.

Raichle, M. E. and Snyder, A. Z. (2007) A default mode of brain function: a brief history of an evolving idea. *NeuroImage* **37,** 1083-1090; discussion 1097-1089.

Rossini, P. M., Burke, D., Chen, R., Cohen, L. G., Daskalakis, Z., Di Iorio, R., Di Lazzaro, V., Ferreri, F., Fitzgerald, P. B., George, M. S., Hallett, M., Lefaucheur, J. P., Langguth, B., Matsumoto, H., Miniussi, C., Nitsche, M. A., Pascual-Leone, A., Paulus, W., Rossi, S., Rothwell, J. C., Siebner, H. R., Ugawa, Y., Walsh, V. and Ziemann, U. (2015) Non-invasive electrical and magnetic stimulation of the brain, spinal cord, roots and peripheral nerves: Basic principles and procedures for routine clinical and research application. An updated report from an I.F.C.N. Committee. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology* **126,** 1071-1107.

Saenger, V. M., Kahan, J., Foltynie, T., Friston, K., Aziz, T. Z., Green, A. L., Van Hartevelt, T. J., Stevner, A. B. A., Fernandes, H. M., Mancini, L., Thornton, J., Yousry, T., Limousin, P., Zrinzo, L., Hariz, M., Kringelbach, M. L. and Deco, G. (2017) Uncovering the underlying mechanisms and whole-brain dynamics of therapeutic deep brain stimulation for Parkinson’s disease [bioRxiv 083162]. *Scientific Reports* **7,** 9882.

Schiff, N. D., Giacino, J. T., Kalmar, K., Victor, J. D., Baker, K., Gerber, M., Fritz, B., Eisenberg, B., Biondi, T., O'Connor, J., Kobylarz, E. J., Farris, S., Machado, A., McCagg, C., Plum, F., Fins, J. J. and Rezai, A. R. (2007) Behavioural improvements with thalamic stimulation after severe traumatic brain injury. *Nature* **448,** 600-603.

Stevner, A. B. A., Vidaurre, D., Cabral, J., Rapuano, K. M., Nielsen, S. F. V., Tagliazucchi, E., Laufs, H., Vuust, P., Deco, G., Woolrich, M. W., Van Someren, E. and Kringelbach, M. L. (2018) Discovery of key whole-brain transitions during the healthy human non-REM sleep cycle. *Nature Communication***,** under review.

Stitt, I., Hollensteiner, K. J., Galindo-Leon, E., Pieper, F., Fiedler, E., Stieglitz, T., Engler, G., Nolte, G. and Engel, A. K. (2017) Dynamic reconfiguration of cortical functional connectivity across brain states. *Scientific reports* **7,** 8797.

Tagliazucchi, E. and Laufs, H. (2014) Decoding wakefulness levels from typical fMRI resting-state data reveals reliable drifts between wakefulness and sleep. *Neuron* **82,** 695-708.

Tognoli, E. and Kelso, J. A. (2014) The metastable brain. *Neuron* **81,** 35-48.

Tononi, G. (2012) *Phi: A Voyage from the Brain to the Soul*. Pantheon: New York.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B. and Joliot, M. (2002) Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage* **15,** 273-289.

Van Hartevelt, T. J., Cabral, J., Deco, G., Møller, A., Green, A. L., Aziz, T. Z. and Kringelbach, M. L. (2014) Neural plasticity in human brain connectivity: the effects of long term deep brain stimulation of the subthalamic nucleus in Parkinson's Disease. *PLoSONE* **9,** e86496.

van Hartevelt, T. J., Cabral, J., Moller, A., FitzGerald, J. J., Green, A. L., Aziz, T. Z., Deco, G. and Kringelbach, M. L. (2015) Evidence from a rare case study for Hebbian-like changes in structural connectivity induced by long-term deep brain stimulation. *Frontiers in behavioral neuroscience* **9,** 167.