Awakening: Predicting external stimulation to force transitions between different brain states

Gustavo Deco\textsuperscript{a,b,c,d,e,1}, Josephine Cruzat\textsuperscript{a,b}, Joana Cabral\textsuperscript{g,h}, Enzo Tagliazucchi\textsuperscript{i,j}, Helmut Laufs\textsuperscript{j,k}, Nikos K. Logothetis\textsuperscript{l,m,1}, and Morten L. Kringelbach\textsuperscript{a,b,1}

\textsuperscript{a}Center for Brain and Cognition, Computational Neuroscience Group, Universitat Pompeu Fabra, 08018 Barcelona, Spain; \textsuperscript{b}Department of Information and Communication Technologies, Universitat Pompeu Fabra, 08018 Barcelona, Spain; \textsuperscript{c}Instituci\`{o} Catalana de la Recerca i Estudis Avan\c{c}ats, 08010 Barcelona, Spain; \textsuperscript{d}Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, 04103 Leipzig, Germany; \textsuperscript{e}Turner Institute for Brain and Mental Health, Monash University, Melbourne, Clayton, VIC 3800, Australia; \textsuperscript{f}Department of Psychiatry, University of Oxford, Oxford OX3 7JX, United Kingdom; \textsuperscript{g}Center for Music in the Brain, Department of Clinical Medicine, Aarhus University, 8000 Aarhus, Denmark; \textsuperscript{h}Life and Health Sciences Research Institute, School of Medicine, University of Minho, 4710-075 Braga, Portugal; \textsuperscript{i}Institute for Medical Psychology, Christian Albrechts University, 24118 Kiel, Germany; \textsuperscript{j}Department of Neurology and Brain Imaging Center, Goethe University, 60528 Frankfurt am Main, Germany; \textsuperscript{k}Department of Neurology, Christian Albrechts University, 24118 Kiel, Germany; \textsuperscript{l}Department Neurophysiology of Cognitive Processes, Max Planck Institute for Biological Cybernetics, 72076 T"{u}bingen, Germany; and \textsuperscript{m}Imaging Science and Biomedical Engineering, University of Manchester, Manchester M13 9PT, United Kingdom

Contributed by Nikos K. Logothetis, June 20, 2019 (sent for review April 3, 2019; reviewed by Andreas Horn and Marcello Massimini)

A fundamental problem in systems neuroscience is how to force a transition from one brain state to another by external driving stimulation in, for example, wakefulness, sleep, coma, or neuropsychiatric diseases. This requires a quantitative and robust definition of a brain state, which has so far proven elusive. Here, we provide such a definition, which, together with whole-brain modeling, permits the systematic study in silico of how simulated brain stimulation can force transitions between different brain states in humans. Specifically, we use a unique neuroimaging dataset of human sleep to systematically investigate where to stimulate the brain to force an awakening of the human sleeping brain and vice versa. We show where this is possible using a definition of a brain state as an ensemble of “metastable substates,” each with a probabilistic stability and occurrence frequency fitted by a generative whole-brain model, fine-tuned on the basis of the effective connectivity. Given the biophysical limitations of direct electrical stimulation (DES) of microcircuits, this opens exciting possibilities for discovering stimulation targets and selecting connectivity patterns that can ensure propagation of DES-induced neural excitation, potentially making it possible to create awakenings from complex cases of brain injury.

Significance

We describe a quantitative and robust definition of a brain state as an ensemble of “metastable substates,” each with a probabilistic stability and occurrence frequency. Fitting this to a generative whole-brain model provides an innovative avenue for predicting where simulated brain stimulation can force transitions between different brain states. We provide proof-of-concept by systematically applying this model framework to neuroimaging data of the human sleep cycle and show where to stimulate to awaken the human sleeping brain and vice versa. These results suggest an avenue for using causal whole-brain models to discover in silico where to force a transition between brain states, which may potentially support recovery in disease.
recurring short-lived patterns that can be clustered in metastable substates for any given brain state (see, e.g., ref. 10).

In the current study, we provide a definition of brain states dubbed “characterization of the probabilistic metastable substates (PMS) space,” which fully typifies substates as stochastic subdivisions of regular and persistent brain states. This allows for recurrent substates to be detected and characterized in terms of probability of occurrence and alternation profiles. This allows us 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 anesthesia or to disease states such as coma and neuropsychiatric disorders (11–13). Solving this problem requires 2 things: 1) a deeper understanding and quantitative definition of what constitutes a brain state, e.g., the proposed characterization of the PMS space; and 2) what drives the transitions between brain states. This would allow for the possibility of forcing a transition using for example external stimulation like deep brain stimulation (DBS), multifocal transcranial direct current stimulation (tCDS) (14), or transmagnetic stimulation (TMS) (15–19).

In order to do this, whole-brain models are needed that can link the underlying anatomical connectivity with the functional dynamics obtained from neuroimaging (8, 9, 20–24). We show here that a generative whole-brain model can actually accurately fit this model to the simulated dynamical characteristics for explaining different brain states. Specifically, we used a unique dataset of continuous neuroimaging data from healthy participants falling asleep during simultaneous functional magnetic resonance imaging (fMRI) and electroencephalography (EEG) (25). Using this dataset, we have recently used a Markovian data-driven analysis to discover the dynamic choreography between different whole-brain networks across the wake–non-REM sleep cycle (26). However, here we provide evidence that we can externally force transitions between different brain states by means of in silico, aiming to find optimal stimulation strategies for forcing the transition from one specific brain state to another. What follows offers a synopsis of this methodology (described in detail in Materials and Methods).

The first step of our analysis has been to identify the PMS space, by using the leading eigenvector dynamics analysis (LEiDA) method (27), described in detail in Materials and Methods and in Fig. 1. Standard (FSL) tools were used to first extract, preprocess and average the blood oxygen level-dependent (BOLD) signals of 90 regions of interest (ROIs) defined in the Automated Anatomical Labeling (AAL) atlas. They included cortical and subcortical noncerebellar ROIs (nodes) that were considered for the estimation of dynamic connectivity. The average BOLD time series of each ROI were Hilbert-transformed to yield the phase evolution of the regional signals. The phase coherence for each pair of nodes at any given time can be defined as the cosine of the phase differences, the value of which varies from 1 to −1, for signals changing in exactly the same or opposite direction, respectively. This process yields a 3D matrix of N x N x T size, where N = 90 and T indicates the number of image volumes acquired in all considered sessions. Clustering the large number of N x N dynamic functional connectivity (dFC) matrices may be in principle used for estimating metastable substates. However, reducing the very large dimensionality of the coherence matrices significantly improves the signal-to-noise ratio and the reliability of any classifying process aiming at the description of states. We have therefore used the aforementioned LEiDA method that relies on the leading eigenvector of each N x N coherence matrix.

The method relies on the extraction of the first (ξ1) eigenvector, V1, of each dFC matrix, from which one can reliably detect a discrete number of reduced dFC patterns by applying clustering across time points and subjects. Due to the symmetry of the dFC matrices, each leading eigenvector may be used in turn to estimate the corresponding dFC matrix. The obtained k-cluster centroids define the “metastable substates,” for which one can compute the probability of the centroid, as well as its transition probabilities.

We use this method in the cases of 2 different naturally occurring brain states: awake and deep sleep conditions obtained in healthy human participants measured with fMRI and EEG (25). The clustering as determined by the silhouettes criterion showed that 3 metastable substates were optimal for fitting both brain states. Nevertheless, it is entirely possible, albeit computationally very expensive, to fit the whole-brain model to a higher number of metastable substates, potentially revealing more of the known networks involved in sleep shown, e.g., in our previous Markovian data-driven results (26). Here, we used 3 metastable substates as a proof-of-concept.

The strategy of brain state definition and potentially inducing a brain state transition is summarized in Fig. 1, demonstrating a probabilistic state space that may be underlying a given brain state, e.g., wakefulness, sleep, anesthesia, or disease. In order to promote or force transitions between brain states, we propose a framework to first characterize brain states (Fig. 1) and then use whole-brain models to generate such brain states (Fig. 24), and ultimately attempt to force a transition between them (Fig. 2B). Each point in the proposed framework represents an instantaneous snapshot of the whole brain at a point in time. The cloud, in turn, describes a global brain state, such as that during wakefulness or deep sleep. In Fig. 1D, the probabilistic points are projected onto a 2D space for visualization, but the original state space is of course likely to be higher dimensional.

By clustering the probabilistic points shown in Fig. 1D, we can generate substates from groups of points and express their probability (leftmost panel of Fig. 1D). Since such groups will still be changing over time, we call them PMS. In Fig. 2A, we show how we can fit a whole-brain model to this PMS space, which provides a probabilistic description of the spatiotemporal dynamics of the underlying functional fMRI time series. The whole-brain model links the structural anatomy (given by the diffusion MRI (dMRI) data) with the functional dynamics (given by the fMRI data) by adapting the free parameters (the effective conductivity of the anatomical fibers) to provide the optimal fit between the simulated and empirical PMS spaces (by using an appropriate probabilistic distance measure, namely symmetrized Kullback–Leibler (KL) distance).

In Fig. 2B, we show how the whole-brain model can be perturbed and stimulated exhaustively in order to promote and force a specific transition between 2 different brain states. Specifically, in what follows, we discuss the strategy of fitting the whole-brain model to the PMS space, and inducing a transition from one specific brain state to another.

**Optimal Spatiotemporal Fit of Whole-Brain Model to PMS Space.** The whole-brain, large-scale model consisted of local nodes representing local brain regions in a parcellation (Materials and Methods). The connectivity of different brain regions was constrained 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 (Materials and Methods), while the dynamics of each local brain area network were described by the normal form of a supercritical Hopf bifurcation, also known as the Landau–Stuart oscillator (28). The internal parameters of the whole-brain model (e.g., the coupling strength between nodes) can be optimized to fit the PMS space.
Computing the probabilistic metastable substate (PMS) space for whole-brain activity. Briefly, we use the leading eigenvector dynamics analysis (LEiDA) method, (A) where for every time point, t, in every brain region of each participant we extract the BOLD signal, compute the phase of the BOLD signal, and (B) compute the BOLD phase coherence matrix, dFC(t), between brain regions, and (C) extract the leading eigenvector V1(t) of this matrix. (D) To obtain a probabilistic metastable substate (PMS) space, we then take all of the leading eigenvectors for all time points in all participants and use an algorithm to cluster them (here, we show 3 clusters). More specifically, in A, we show for each of the 90 AAL parcels (shown in different colors), we extract the BOLD signal for a given region (here in orange). For each brain area, the original BOLD signal (orange) is first band-pass filtered between 0.02 and 0.1 Hz (blue) and then transformed into an analytic signal, which can be represented by its time-varying amplitude A and its phase θ (with real and imaginary components) using the Hilbert transform. The Hilbert phase, θ, can be represented over time in the complex plane by e^{iθ}, where the real part is given by cos(θ), while the imaginary part by sin(θ) (black dotted lines). The red arrows represent the BOLD phase at each TR. As can be seen, much of the original BOLD signal is captured by the BOLD phase, cos(θ) (dotted line). (B) For a given time point (gray box shown at t = 50 in A), the Left panel shows the BOLD phases in all 90 AAL regions represented in the complex plane (i.e., the unit circle with real and imaginary axis, where all phases are centered at the same origin). The Right panel shows the phase coherence matrix at a given time t with the BOLD phase coherence (PC) between each pair of brain regions. (C) The leading (i.e., largest magnitude) eigenvector of this phase coherence matrix at time t, V1(t), is the vector that best captures the main orientation of all BOLD phases, where each element in V1(t) corresponds to the projection of the BOLD phase in each region into V1(t). The elements of V1(t) are colored according to their sign (red, positive; blue, negative; same color scheme in all phase representations). (D) Finally, to determine the PMS space, we gather all leading eigenvectors V1(t) for all participants across time as a low-resolution representation of the BOLD phase coherence patterns. We apply a clustering algorithm (k-means) to divide the sample into a predefined number of clusters k (here, k = 3). Each cluster is represented by a central vector (blue, orange, and red), which we take to represent a recurrent pattern of phase coherence, or substate, which occurs with a given probability. Any brain state can thus be represented by this PMS space.

Indeed, the strategy here is to use the centroids of the empirical clusters defining the metastable substates, and to compute the simulated probabilistic measurements based on those empirical centers. In this way, we assure that the probabilistic measurements correspond exactly to the empirical metastable substates.

Fig. 3A shows the performance of the whole-brain model for fitting data of wakefulness state. In this case, we exhaustively explore the parameter G, which scales the interarea coupling, thereby determining the dynamical working point of the system. The G coupling parameter scales the density of fibers expressed in the SC and can be interpreted as an indicator of axonal conductivity, under the simplistic assumption that conductivities of axons related to long-range connectivity are equal across the brain, assuming similar myelination density.

Fig. 3A shows 2 measurements of fitting as a function of G: 1) the phase-coherence-based functional connectivity dynamics (FCD); and 2) the 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 $G_\text{opt} = 0.283$. In this case, where the whole-brain model was fitted to the FCD, the optimal minimum turned out not to be the best fitting of the spatiotemporal dynamics as characterized by PMS. Instead, in the second case, we obtained more accurate results by extracting the empirical PMS, shown in Fig. 3B with the PMS space as well as the transition probability matrix (TPM). To find the optimal minimum for the PMS, we computed the symmetrized KL distance between the empirical probabilities and the simulated probabilities, generated by the model from the same empirical centroids. This was a stringent criterion to ensure that we would get the best possible fit. Indeed, the optimal PMS fitting of the empirical data were obtained at the minimum of $G_\text{opt} = 0.245$, which clearly was a significantly better fit than that of the FCD (Fig. 3C, Upper). Furthermore, the transition probabilities are also significantly better fitted as measured with the Markov entropy distance (Fig. 3C, Lower). 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.

Overall, these results showed that the PMS-based measurements for characterizing a brain state provide 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.
Fit whole-brain model to Probabilistic Metastable Substates (PMS) for a brain state

Connectivity (fMRI)

Functional brain dynamics (fMRI)

Empirical data

Fit whole-brain model

Whole-brain model fit

Brain state X

Perturbation of model

Brain state Y

A

0.5

B

0.5

0

0

A

B

C

A

B

C

Fig. 2. Schematic of how to model a brain state and force a transition between different brain states. This process involves 3 steps. (A) First, we empirically characterize a given brain state in terms of its probabilistic metastable substrates (PMS) space (using the method shown in Fig. 1). This gives rise to 3 different substrates (blue, orange, and red), which have different probabilities associated with each substrate. Second, we fit a whole-brain model directly to this PMS space (Materials and Methods). (B) Third, the whole-brain model in source brain state X can be perturbed and stimulated exhaustively in order to promote and force a specific transition to a target brain state Y.

Optimizing the Whole-Brain Model by Using Effective Connectivity.

The third step in modeling the spatiotemporal PMS space is to derive 2 different optimized whole-brain models for 2 radically different brain states. Here, as a proof-of-concept, we used 2 naturally occurring significantly different states, namely, wakefulness and deep sleep (25).

Fig. 4 demonstrates the improvements in the aforementioned fitting procedure, by optimizing the effectiveness of the synaptic connections between brain regions [effective connectivity (EC)]. Specifically, we computed 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 Materials and Methods for details). The model was run repeatedly with recursive updates of EC until convergence was reached (Materials and Methods).

Fig. 5 shows the results obtained during wakefulness (Fig. 5A) and sleep (Fig. 5B), with the first row showing the PMS space and the second row, the TPM, and with left column showing the empirical results and the right column, the fit of the whole-brain model. Below this, we show EC matrix and the effective degree. For the whole-brain model fit of PMS space, we found a KL distance of 0.0169 for wakefulness, and 0.0045 for sleep. Similarly, for whole-brain fit of the TPM, we found a Markov entropy distance of 0.098 for wakefulness and 0.109 for sleep. Also note how different the EC matrices are for the 2 states (bottom 2 rows).

Awakening: Forcing a Brain State Transition.

The previous results show that we can successfully create 2 whole-brain models with excellent fit to the empirical fMRI data from 2 radically different brain states. However, the most important finding of this study is the demonstration of the possibility to force a transition from one whole-brain model to another via external stimulation.

Fig. 6 summarizes the process of forcing a transition between 2 brain states (source and target). We systematically perturbed the brain regions in the whole-brain model of the source state and compared the resulting PMS space for this model with the empirical data for the other target state (Fig. 6A). Specifically, the Hopf model allowed an effective way of perturbing the model by simply changing the bifurcation parameter in a given brain region (29). The stimulation intensity, i.e., the strength of the perturbation, is directly related to the amount of shifting the local bifurcation parameter (Fig. 6B, Right, Materials and Methods, and ref. 30). We perturbed the model bilaterally in Fig. 6B, 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 color scale of Fig. 6B, Left, shows the level of fitting with the target state (wakefulness), i.e., lower values (blue) correspond to an effective transition.

For the main transition results, we used 2 different protocols for external stimulation, synchronization and noise, which shifted the local bifurcation parameter to positive and negative values, respectively. Fig. 7A shows the results for forcing a transition from source state (deep sleep) to target state (wakefulness) using a synchronization protocol where positive values of the local bifurcation parameter force local oscillations that promote the possibility of more synchronization across the whole brain. The color 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 = 0.05). In 7A, Right, we show a rendering 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 burgundy areas). Importantly, in Fig. 7C, 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 colors to more yellow from blue. Also note that the color scale is different between Fig. 7A and B compared to Fig. 7C and D, with the first column in each figure having the identical numerical KL distances (corresponding to the nonperturbation case) but appearing in different colors due to different color scales.

Fig. 7B shows the results for forcing a transition from source state (wakefulness) to target state (deep sleep) using a noise protocol, where negative values of the local bifurcation parameter force local oscillations that promote the possibility of more synchronization and less synchronization 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 = 0.4 (note the increase in burgundy areas). Importantly, Fig. 7D shows that using the synchronization 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 it is probably much easier, and thus unspecific, to promote a transition from sleep to wake than vice versa.

Finally, we explored whether stimulating multiple regions with weaker stimulation intensity would produce equal or better results. Fig. 8 shows the results of using this multisite stimulation protocol using a greedy strategy (31) to find the best combination of multiple brain regions for forcing a transition between states (deep sleep to wakefulness). Using the synchronization 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 7 steps. As shown in Fig. 8, the combination of multisite stimulation reaches its best fit (comparable to the best fit for single-site stimulation at higher stimulation intensity; Fig. 7A) using 4 bilateral stimulated regions (frontal middle gyrus, temporal inferior gyrus, frontal superior gyrus, and precuneus, shown rendered in the right upper panel) 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 4 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 Fig. 8B, we plot the evolution of the 3 substrates in the source (deep sleep) and target (wakefulness) states as a function multiregional stimulation. The blue line represents the probability of substrate A,
Results of fitting whole-brain model to awake data

A Fitting of whole-brain model to awake data

- PMS
- FCD

B Empirical data

C Significance analysis

**Fig. 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 fibers expressed in the SC and can be interpreted as the synaptic conductivity of each single fiber. Here, we show 2 measurements of fitting as a function of $G$: 1) the phase-coherence–based functional connectivity dynamics (FCD), and 2) the probabilistic metastable substrates (PMS) space. The best fit of the FCD can be found by computing the Kolmogorov–Smirnov (KS) distance (black line), while the fit to PMS space can be computed using the symmetrized Kullback–Leibler (KL) distance between the empirical probabilities and the simulated probabilities generated by the model using the same empirical centroids (blue line). The optimal fit of for PMS is with $G = 0.245$. (B) We show the PMS space and the TPM when using 3 substrates. (C) We found that the PMS fitting of the empirical data are significantly better than the FCD fitting (upper figure). The transition probabilities are also significantly better fitted as measured with Markov entropy distance (lower figure). Please note how the values of the diagonal of the transition matrix which increases with number of regions stimulated, while orange and red lines representing probability of substrates B and C decrease. 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 4 stimulated regions.

**Discussion**

We successfully used a whole-brain model to fit the PMS space of the 2 radically different brain states of sleep and wakefulness. The states were extracted from a unique continuous neuroimaging data of healthy participants falling asleep during simultaneous fMRI and EEG (25). We provided the evidence that in silico stimulation of this whole-brain model can be used to show where to force transitions between different brain states and thus “awaken” the brain from deep sleep to wakefulness and vice versa.

We provided evidence primarily for role of single regions to promote this transition but also provided proof-of-principle that multisite stimulation can achieve similar results with less stimulation in each site. As expected, we found an optimum number of regions that will promote a transition between states, while adding more regions will significantly worsen the transition probability, reflecting the metastable nature of the brain states. Delivering multisite stimulation to the brain is by its very nature more difficult with techniques such as DBS but perhaps possible with a multifocal tDCS array (14). Still, at this point it is too early to speculate whether single or multisite stimulation would be more optimal. In ongoing work, we have also demonstrated that we can achieve awakening with more biophysical realistic models such as the dynamic mean field model, which speaks to the potential for clinical benefit.

Interestingly, the results show that it is easier to awaken the model, i.e., promote a transition from sleep to wakefulness, rather than promote a transition in the opposite direction. It requires less stimulation in more regions. Evolutionarily, this makes sense given that in order to promote survival we have to be able to easily wake from slumber, while falling asleep easily could be seen as less critical.

More broadly, these results have to be seen in the context of the well-known Latin statement “post hoc, ergo propter hoc” (“after this, therefore because of this”), which is a classical logical fallacy, confusing correlation and temporal sequences with causality. There have long been many spectacular examples of cases, in which the effect may even precede its cause, as in the case of backward causation (e.g., the retrocausality in the fields of quantum mechanics). Richard Feynman (F)—with his F-clocks and F-gates—was one of the historical physicists who attempted to understand so-called plateaus of complexity (POCs), i.e., the physical states of complex systems, by enforcing state transitions and describing the POCs associated with them (32). One thing became clear over time: state transitions in any discipline can only be correctly interpreted, predicted, or even successfully enforced, if the states themselves are well understood at a necessary complexity level, and the self-organization and dynamics of the systems can be modeled, potentially leading to theories predicting empirical data.

Still, it goes without saying that understanding states and their dynamics in brain-like systems makes physics to appear as an “easy” matter (32). The organization of the brain daily challenges most of our efforts to study it, as its understanding requires concurrent studies of local and global networks at different spatio-temporal scales, as well as of their dynamical evolution over time. At the macroscopic level, the developments of neuroimaging over the last few decades has become a valuable source of empirical data (e.g., ref. 33), but quantitative description of such data, modeling of
systems, and even perturbing them in silico for understanding causal relationship between neural and cognitive states is a field in its early childhood, albeit with promising perspectives (23, 24, 31, 34, 35).

State definition in systems neuroscience justifiably has drawn and continues to draw a great deal of attention. Attempts were made to define a brain state in many different ways, often by using heuristic definitions of signals or signal combinations, and points in an empirically defined state space characterizing the activity of the brain at a given time (2–5), by defining attractors of interacting brain regions (5, 6). Other approaches utilized the whole-brain connectomic measurements including dFC (36–39). However, most of these approaches fail to capture and exploit the intriguing role of various band-limited–power neurophysiological signals and the spatiotemporal richness of fMRI data.

The PMS space strategy we describe here captures an unusually large proportion of data as an ensemble or probabilistic “cloud” in a state space, which can be decomposed into so-called metastable substates (27, 40), which are temporarily stable but variable as a function of time (30, 41, 42). Fitting was optimized by fine-tuning the model on the basis of the EC.

A clear advantage of using such data-constrained whole-brain model is its potential use for studying stimulation-induced state transitions, as it enables an exhaustive search and optimization of all underlying parameters and locations in silico, and it may ultimately offer insights into the self-organization of widespread networks. Such insights may be impossible to gain even with direct electrical stimulation (DES) of brain sites in experimental animals, or occasionally in human patients, as DES has been shown to violate the most basic principles of cortical microcircuits, disrupting cortico-cortical signal propagation by silencing the output of any neocortical area whose afferents are electrically stimulated (43–45). Stimulation of lateral geniculate nucleus (LGN), for example, will yield fMRI maps with positive BOLD responses, in the striate cortex, colliculus, and pulvinar, and negative bold responses in all extrastriate areas, disynaptically connected to LGN. Intracranial recordings, concurrently carried out with fMRI, consistently confirm that a short excitatory response occurring immediately after a stimulation pulse was followed by a long-lasting spiking inhibition. This inhibition is synaptic, rather than due to excitability changes, as injection of GABA antagonists, such as bicuculline, restores the stimulation-induced excitatory responses. It has long been known that, in the cortical canonical microcircuits, excitation and inhibition are inseparable. Sensory stimulation will first excite glutamatergic and, ~2 ms later, GABAergic neurons (46). Both have a strong recurrent self-amplification, and the output of the microcircuit depends on so-called excitation–inhibition (E–I) balance at any given time. Electrical stimulation of the cortical afferents, nulls the excitation differences between E- and I-neurons, activating both groups simultaneously, and shunting the output of the microcircuits through the aforementioned powerful synaptic inhibition and before the initiation of the recurrent self-amplification (44, 47).

Not surprisingly thus, the stimulation of the LGN in monkeys in DES-fMRI experiments demarcated all monosynaptic targets of the stimulated brain site, or alternatively a selection of polysynaptic nuclei reached through the antidromic stimulation of collaterals of cortical infragranular projection neurons. It follows that DES in animals or humans propagates through cortico-subcortico-cortical pathways, rather than cortico-cortical connectivity. The often unpredictable active connectivity combined with the very fact that we can still not precisely define the element activated by DES (19, 48). This explains the fact that a large amount of Parkinson patients suffer from adverse stimulation effects (49). Understanding states and optimizing in silico stimulations can only be instructive for future attempts to deal with the DES technology and interpretation of its results.

The approach presented here may open new, exciting possibilities for discovering new stimulation targets for DBS and TMS in disease. A stimulation strategy that minimizes the unavoidable

---

**Fig. 4.** Fitting whole-brain using effective connectivity (EC). We improved on the fitting procedure by optimizing 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 *Materials and Methods* for details). The model is run repeatedly with the updated EC until convergence.

---

**Fig. 5.** Fitting a whole-brain model to 2 radically different brain states (wakefulness and sleep) measured using fMRI. The figure shows the results obtained for wakefulness (left column; A) and sleep (right column; B) comparing the empirical results with the output of the fitted whole-brain EC model. In each row, we show the PMS space, transition probability matrices, EC 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 bar plots for the PMS spaces). Please also note how different the EC matrices are for the 2 states (bottom 2 rows).
Overall, the methods and results presented here may eventually allow us to build causative whole-brain models that can characterize 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 2 brain states.

Materials and 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 described in detail in previous publications (25). All subjects gave written informed consent with approval by the local ethics committee. For the present study, we only considered the subset of subjects who reached deep sleep (stage N3). A summary of the technical and experimental aspects of data acquisition and preprocessing can be found in SI Appendix.

LEIDA. First, we calculated a phase coherence matrix at each time point to capture the amount of interregional BOLD signal synchrony at any given time point, for all subjects and conditions (awake and N3 deep sleep). Specifically, the discrete time series of each ROI, termed as the discrete input is 2 a discrete time series of each ROI, is termed as the discrete input parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function. The parameter function.
the probability of, being in a given substate (rows), transitioning to any of the other substates (columns). Differences in probabilities of occurrence and probabilities of transition were statistically assessed between conditions using a permutation-based paired t test. This nonparametric test uses permutations of group labels to estimate the null distribution. The null distribution is computed independently for each experimental condition. For each of 1,000 permutations, a t test is applied to compare populations and the significance threshold $\alpha = 0.05$ was used.

Whole-Brain Computational Model. We simulated the BOLD activity at the whole-brain level by using so-called Hopf computational model, which emulates the dynamics emerging from the mutual interactions between brain areas, considered to be interconnected on the basis of the established graphs of anatomical SC (30, 57). The model consists of 90 coupled dynamical units (ROIs or nodes) representing the 90 cortical and subcortical brain areas from the aforementioned AAL parcellation. The local dynamics of each brain area (node) is described by the normal form of a supercritical Hopf bifurcation in Cartesian coordinates:

$$\frac{dx_n}{dt} = [a_n - x_n^2 - y_n^2]x_n - \omega_n y_n + \beta_n \eta_n(t),$$

and

$$\frac{dy_n}{dt} = [a_n - x_n^2 - y_n^2]y_n + \omega_n x_n + \beta_n \eta_n(t),$$

where $\eta_n(t)$ is additive Gaussian noise with SD $\beta$. This normal form has a supercritical bifurcation $a_n = 0$, so that if $a_n > 0$ the system engages in a stable limit cycle with frequency $f_n = \omega_n/2\pi$ and when $a_n < 0$ the local dynamics are in a stable fixed point representing a low-activity noisy state. Within this model, the intrinsic frequency $\omega_n$ of each node is in the 0.04- to 0.07-Hz band ($n = 1, \ldots, 90$). 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 SC $C_{np}$. 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:
adjusted the model to the dynamical working region where the simulations optimally fit the empirical data (30, 63).

**Empirical Fitting.**

*Comparing empirical and simulated grand-averaged static functional connectivity.* 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 functional connectivity (FC).

*Comparing empirical and simulated FCD.* We measure KS 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 \((t_1, t_2)\) entry is defined by the cosine similarity between the upper triangular parts of the 2 matrices \(dFC(t_1)\) and \(dFC(t_2)\) (previously defined; see above). For 2 vectors \(p_1\) and \(p_2\), the cosine similarity is given by \(\langle p_1, p_2 \rangle / (|p_1||p_2|)\). Epochs of stable FC(t) configurations are reflected around the FCD diagonal in blocks of elevated inter-FC(t) correlations. The KS distance quantifies the maximal difference between the cumulative distribution functions of the 2 samples.

*Comparing empirical and simulated probability metastable space state measurements.* For comparing the probabilities of the metastable states, i.e., the probabilities of the extracted empirical centers after clusterization, we used a symmetrized KL distance between the simulated and empirical corresponding probabilities, i.e.:

\[
KL(P_{emp}, P_{sim}) = 0.5 \left( \sum_i P_{emp}(i) \log \left( \frac{P_{emp}(i)}{P_{sim}(i)} \right) + \sum_i P_{sim}(i) \log \left( \frac{P_{sim}(i)}{P_{emp}(i)} \right) \right),
\]

where \(P_{emp}(i)\) and \(P_{sim}(i)\) are the empirical and simulated probabilities on the same empirical extracted metastable substates \(i\).

**Comparing empirical and simulated transition probabilities between metastable substates.** We calculated the entropy rate \(S\) of a Markov chain, with \(N\) states and transition matrix \(P\). The rate entropy \(S\) is given by the following:

\[
S = S_1 + S_2 + \ldots + S_N,
\]

where

\[
S_i = -p(i) \sum_{j=1}^{N} p(i,j) \log(p(i,j)).
\]

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

\[
p^T p = p.
\]

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 high entropy, while a Markov model that barely transmits has low entropy. For each transition matrix, we obtained the stationary distribution, and then calculated the entropy rate. The final measure comparing the 2 TPMs is just defined by the absolute value of the difference between both respective Markov entropy values.

**Methods for updating EC.** We derived 2 whole-brain models, namely, 1 for accounting the awake condition and another for accounting the N3 deep sleep condition. In both cases, we optimized the EC between brain regions. Specifically, we compute the distance between the model FC_{phase-mod} and empirical FC_{phase-emp} grand-averaged phase coherence matrices, and adjust each structural connection separately with a gradient-descent approach, thereby transforming structural into effective connections. The model is run repeatedly with the updated EC until the fit converges toward a stable value.

We start with the anatomical connectivity obtained with probabilistic tractography from dMRI and use the following procedure to update:

\[
C_{ij} = C_{ij} + \epsilon (C_{ij} - C_{ij}^{phase-mod} + C_{ij}^{phase-emp}).
\]

We update all connections in order to access the potential missing components of the anatomical connectivity obtained with dMRI, which is known to miss connections in the opposite hemisphere, given that most known tracts...
are bilateral. We use $e = 0.01$ and the grand average phase coherence matrix is defined as follows:

$$F_{OC}^j(t) = \cos \psi_j(t) - \phi_j(t),$$

where $\psi_j(t)$ is the phase of the BOLD signal in brain region $j$ at time $t$ extracted with the Hilbert transform, with the bracket denoting the average across time.

ACKNOWLEDGMENTS. G.D. is supported by the Spanish Research Project PSI2016-75688-P (Agencia Estatal de Investigación/Fondo Europeo de Desarrollo Regional, European Union); by the European Union’s Horizon 2020 Research and Innovation Programme under Grant Agreements 720270 (Human Brain Project [HBP] GA1), 785907 (HBP SGA2); and by the Catalan Agency for Management of University and Research Grants Programme 2017 SGR 1545. J. Cabral is supported by Portuguese Foundation for Science and Technology (CEECIND/03325/2017, Portugal. M.L.K. is supported by the European Research Council Consolidator Grant: CAREGIVING (615539) and Center for Music in the Brain, funded by the Danish National Research Foundation (DNRF117).

1. W. James, The Principles of Psychology (Henry Holt, New York, 1890).
2. S. Gu et al., Optimal trajectories of brain state transitions. Neuroimage 148, 305–317 (2017).
3. D. Gervasoni et al., Global forebrain dynamics predict rat behavioral states and their transitions. J. Neurosci. 24, 11137–11147 (2004).
4. F. S. Muldoon et al., Stimulation-controlled dynamics of dynamic brain networks. PLoS Comput. Biol. 12, e1005076 (2016).
5. S. Gu et al., The energy landscape of neurophysiological activity implicit in brain network structure. Sci. Rep. 8, 2507 (2018).
6. G. Deco, V. K. Jirsa, Ongoing cortical activity at rest: Criticality, multistability, and ghost attractors. J. Neurosci. 32, 3366–3375 (2012).
7. J. Friston, L. Harrison, W. Penny, Dynamic causal modelling. Neuroimage 19, 1233–1302 (2003).
8. G. Deco, V. Jirsa, A. R. McIntosh, O. Sporns, R. Kötter, Key role of coupling, delay, and noise in resting brain fluctuations. Proc. Natl. Acad. Sci. U.S.A. 106, 10302–10307 (2009).
9. G. Deco et al., Topology, speed, and scale of brain connectivity: Contributions of whole-brain modelling. Nat. Rev. Neurosci. 16, 430–439 (2015).
10. T. O. Laumann et al., On the stability of BOLD fMRI correlations. Cereb. Cortex 27, 4719–4732 (2017).
11. G. Tonioni, A. Schindler, T. Varela, S. Massimini, F. Schnakers, M. L. Dauwels, A. K. Roldan, R. Kötter, S. Wibral, J. M. Bressler, D. Deco, T. J. van Hartevelt, H. M. Fernandes, A. Stevner, M. L. Kringelbach, 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 (2017).