Eider Pérez-Ordoyo Bellido

Ph.D. Student at the Computational Neuroscience Group, Center for Brain and Cognition, Department of Information and Communication Technologies

Email: eider.perezordoyo@upf.edu

Forced Brain State Transitions via External Stimulations: A Computational Approach

Research proposal

1. Introduction and State of the Art

Consciousness is considered by some a buzzword, a concept around which one has to tread lightly. Looking at the amount of debate around it, be it in philosophical discussions or scientific theories that try to find its neuronal correlates, it is easy to understand why it is a controversial concept. Even in the midst of all the confusion, one thing seems to be clear: there appear to be different states of consciousness. Some of them are accessed physiologically, others are accessed through training or drugs, and others are caused by injuries or neuropsychological conditions. Over the last few years, these different brain states have been studied with the help of whole-brain models. Even more recent is the concept of *in silico* stimulation: using these models to simulate brain stimulation. Together, they have opened the imagination to ways of finding targets for brain stimulation that allow the brain to transition between different states furthering the work towards the personalised treatment of patients.

1.1. On brain states

The brain is a complex and dynamic organ that exhibits a wide range of states and activities, from wakefulness to sleep, from focused attention to mind wandering, and from stress to relaxation. These different *brain states* are associated with distinct patterns of neural activity and functional connectivity that underlie various cognitive, emotional, and behavioural processes. Understanding the nature and mechanisms of brain states and their transitions is a fundamental challenge in neuroscience, as it could shed light on how the brain processes information, adapts to changing environments and produces conscious experience.

The definition of “brain state”, though, has been lacking the view of the brain as a dynamic system. If we see it as such, we can see that the emergence of brain states comes from its complexity. As a complex system, we can see that it shows different properties expected from similar systems. For example, self-organisation without centralised control [1] and other emergent properties.

A definition of brain state that takes that into account is the one given by Kringelbach and Deco [2] in which they refer to them as “the continuously evolving dynamics of widespread networks that are characterised by condition-dependent self-organisation, going through stable, quasi-stable, high or low activities, and transient arrangements”. This definition also alludes to *metastability*: the property of states that fall outside of the natural equilibrium of the system but stay for extended periods of time [3]. This concept of metastability has been around in the field of neuroscience for some time now [4], and has been used for a diverse range of studies: ERP studies during development [5], decision-making [6], resting-state networks studies [7], the effect of lesion and damage on this property [8] and many more. Computational studies even show that the resting brain is “maximally metastable”, the optimal compromise between stability and flexibility, allowing for rapid and context-dependent transitions between different states to adapt to the current needs of the system [9].

1.2. Consciousness

Another behaviour that we could attribute to emergence (although it is still a very debated topic philosophically) is consciousness, i.e. the subjective experience of existing. A lot of work has been done on scientific theories of consciousness, trying to find the neural correlates that make this subjective experience viable, that make us able to lose it physiologically during sleep and that makes some people lose it under the influence of what we call *disorders of consciousness* (DoC). A recent review on the explanatory models of consciousness is that of Signorelli, Szczotka and Prentner (2021) [10]. They use a three-dimensional classification of models of consciousness regarding their mode, mechanism and target of explanation. There, the most common theories are commented on, such as Thalamo-Cortical loops and Sensorimotor Couplings by Llinás and colleagues [11], Information Integration Theory, by the group of Giulio Tononi [12] and the Global Neuronal Workspace theory by Dehaene and colleagues [13]. From all these theories and their respective supporting data and studies, one thing comes clear: for consciousness to be present there needs to be an interplay between different brain areas and nuclei. No single area or neural circuit is entirely responsible (or sufficient) for producing such emergence.

Identifying the presence of subjective experience of other beings is not clear-cut and the clinical tools to determine the level of consciousness are normally simplified versions that serve the purpose they were designed for (being easy to use and consistent). For example, one of the most widely used tools for determining patients' unconsciousness level is the Glasgow Coma Scale (GCS) [14] which assesses a patient's level of consciousness based on their responses to stimuli. The scale itself studies patients in three dimensions with a very coarse parameterization and ends up giving a unidimensional value which is an oversimplification when talking about conscious experience. That being said, it is pointed out that for diagnostic purposes other variables have to be taken into account. Also, individual assessments have to be done using the three dimensions as separate, only groups having assigned a single value assessment.

The use of responses to stimuli when assessing consciousness is widely criticised, as response to stimuli has more to do with arousal and perceptual gating. Several papers discuss the disparities between various cognitive processes that are frequently misidentified as consciousness and what we can call “true” consciousness. One such paper is the one by

Koch and Tsuchiya (2007) about the difference between attention and consciousness [15]. Some authors, like Tassi and Muzet (2001) conclude that “an inevitable consequence of our inability to identify clear-cut non-occurrence of subjective awareness is the logical impossibility of identifying the biological basis of such awareness” [16]. These authors also talk about the different concepts that are mistaken for consciousness, for example, distinguishing between awareness and consciousness when we are sleeping and dreaming (conscious but not aware of our surroundings). They also mention that there are evoked states of consciousness, such as those obtained by self-training through meditation or trances, those obtained by hypnosis and the ones obtained by pharmacological means. Some work has been done in the field of generative computational whole-brain models regarding these evoked states, such as the work done by De Filippi [17] and Escrichs [18] on meditation states, and the effects of pharmacologically induced states of consciousness are also being studied with these models [19, 20].

A big part of altered states of consciousness is the so-called disorders of consciousness, characterised by alterations in awareness or arousal or awareness. Common causes of these disorders are cardiac arrests, traumatic brain injuries (TBI), intracerebral haemorrhages and ischaemic strokes [21, 22]. To study these conditions in a controlled manner, animal models are frequently used. In particular, rodent models of TBI, primate models of global ischemia and models of hypoxia and anoxia in other species. These animal lesional models have their drawbacks, and are, of course, irreversible. For that reason, anaesthetics are used as a way to induce loss of consciousness (LoC) [23] in animals to be treated as models.

1.3. Stimulation and Whole Brain Models

Some of the proposed treatments for DoC are stimulation of certain brain areas [21], be it Direct Electrical Stimulation (DES), Transcranial Magnetic Stimulation (TMS), Deep-Brain Stimulation (DBS) or any other. For example, a study by Tasserie and colleagues [24] sheds some light on using DBS to “wake up” an animal model from LoC. But this venue for the therapeutic use of brain stimulation is not only present for DoC, as it is also proposed for other neurological and neuropsychiatric disorders [25]. All of those “deviations from normalcy” share that their dynamics during resting-state are also different from that of an “ideal healthy brain”, and that they can be studied using a whole-brain modelling framework [26].

Stemming from that idea, we have the approach of Deco and colleagues (2019) [27]. Starting from awake resting-state and sleep fMRI data they are able to create a framework to differentiate between the different brain states and, using a whole-brain model to fit the data. They simulate perturbations to the model showing how, theoretically, you could force the brain to transition from one brain state to another. These perturbations would represent brain stimulation. This approach is very powerful, but to my knowledge, it is yet to be tested in either a predictive or descriptive manner.

Continuing with the idea of stimulating (or perturbing) whole-brain models, there is present work on what these *in silico* perturbations lead us to. For example, Sanz Perl and collaborators (2021) [28] show how perturbing the model gives us insights into the different consciousness levels, like the stability and level of consciousness. To help with the connection between theory and practice, Dynamic Sensitivity Analysis [29] is a procedure

that could be followed to better understand which kinds of stimulation would have a greater impact on the restoration of optimal brain dynamics. Most recently, a group of clinicians and neuroscientists from the international Curing Coma Campaign put together a framework (currently a preprint) that functions as a compendium for a lot of information on modelling DoC and tries to, as they say, “close the gap” between modelling and the goal of restoring consciousness [30].

Whole-brain models put together the anatomical connectivity of the brain, usually measured with tools like Diffusion Tensor Imaging (DTI), sometimes together with neuronal tracer studies; with the local dynamics of the brain regions to fit the empirical functional data [19, 25, 27]. There are different approaches to whole-brain modelling, focusing mainly on different ways to represent local dynamics. Examples of these approaches include spiking models, Dynamic Mean Field (DMN) models, neural mass models, or nonlinear dynamical models that rely on deterministic chaos or noise-driven multistability [19, 31, 32]. The so-named Hopf model, which represents the local dynamics by the means of a Stuart-Landau oscillator, has been shown to provide the best fit as commented by Kringelbach and colleagues (2023) [33], and it has been used in the past to tackle different brain functions such as neurodegenerative diseases [34] and stroke [35], to name some examples; as well as our sought after consciousness [28]. Another great property of these computational models is that they can be used to leverage already available datasets that are already available to find new insights.

1.4. Using a low-dimensional approach to consciousness

As previously mentioned, consciousness does not have a clear-cut definition or measure, and as such the tools used to determine the “level of consciousness” of a patient are not designed for an actual definition of the consciousness state, but more likely for practical purposes of clinical practice [14].

It is true that handling a multifaceted concept like consciousness as a way to deal with patients is challenging and a lower-dimensional approach is needed for easier analysis and interpretation. For that reason, Sanz Perl et al. (2020) [37] implemented a Variational Autoencoder (VAE) to reduce the dimensionality of the data and differentiate between brain states. An autoencoder is a model comprised of two artificial neural networks: an encoder and a decoder. The encoder compresses the data into a low-dimensional space (called latent space), and the decoder tries to reconstruct the original data starting from the low-dimensional representation. The training of the neural networks aims to minimise the reconstruction error. In the VAE, apart from minimising the reconstruction error, another condition is created in order to have the distributions of data in the latent space follow a probability distribution (normally a Gaussian distribution), a renormalisation. This allows the VAE to become a generative model and the points in the latent space not covered by the initial data obtain some meaning [38]. This approach has been used in other studies with the same purpose[34, 36].

This approach faces one issue regarding the usage of fMRI data: the training of neural networks usually requires great amounts of data and fMRI data is costly and the frequency of acquisition leads to long acquisition times. To alleviate this problem, the concept of data augmentation based on dynamical systems can be used [39]. This consists of the creation of surrogate data by a whole-brain model fitted to the empirical data, allowing

the creation of virtually any amount of data needed for training your model. This concept is something that facilitates the creation of personalised treatments since it allows us to obtain insights into small amounts of data with tools that would require more.

Seeing that consciousness has many faces and that whole-brain models can be used as a tool to simulate brain states and their dynamics, this thesis will focus on the study of transitions between brain states, especially those related to consciousness. The tools described in this introduction will be used to the ends of simulating brain stimulation and the forcing of brain state transitions.

2. Research objectives

The main objective of this thesis is to close the gap between using whole-brain modelling as a way to simulate brain stimulation, and the experimental data obtained by brain stimulation.

To that purpose, I plan on using datasets that obtain recordings from subjects before and after (and during if possible) brain stimulation. Modelling the brain state before and after the stimulation and finding stimulations that best reproduce the ones done in the experimental setting. If possible, I would like to be able to find general forms of these stimulations that will allow more accessible modelling of brain stimulation and a “cleaner” theoretical framework.

Another great objective that depends on outside teams would be the prediction of stimulation, by using a whole-brain model to predict the outcome of stimulation and collaborate with a team that could perform the experiment and see the match between the predictions from the model and the data.

In the development of the thesis, I would like to make my tools as accessible and streamlined as possible, allowing people from different backgrounds to use them easily. I would also like to find ways of optimising the pipeline so that it can be used close to real-time (which would be the expectation for a clinical setting).

3. Research plan and methodology

The methodology that will be used is similar to the one used in past literature [19, 27, 37, 39]. It consists of the creation of whole-brain models to fit fMRI data of different states of consciousness, create a representation of the states in a low-dimensional space using a Variational Autoencoder, and stimulate the models to try and find targets for stimulation that will move the system from a low-consciousness state to a higher consciousness state.

The dataset we use is the one from the research by Tasserie et al. (2022) [24], in which monkeys have implanted a DBS lead in the thalamus and anaesthetised in order to demonstrate that stimulation to the central thalamus improves their measures of awareness and alertness.

Through the next subsections, I will review the methods used so far and that I will continue using throughout my project. Figure 3a presents an overview of the methodology used.

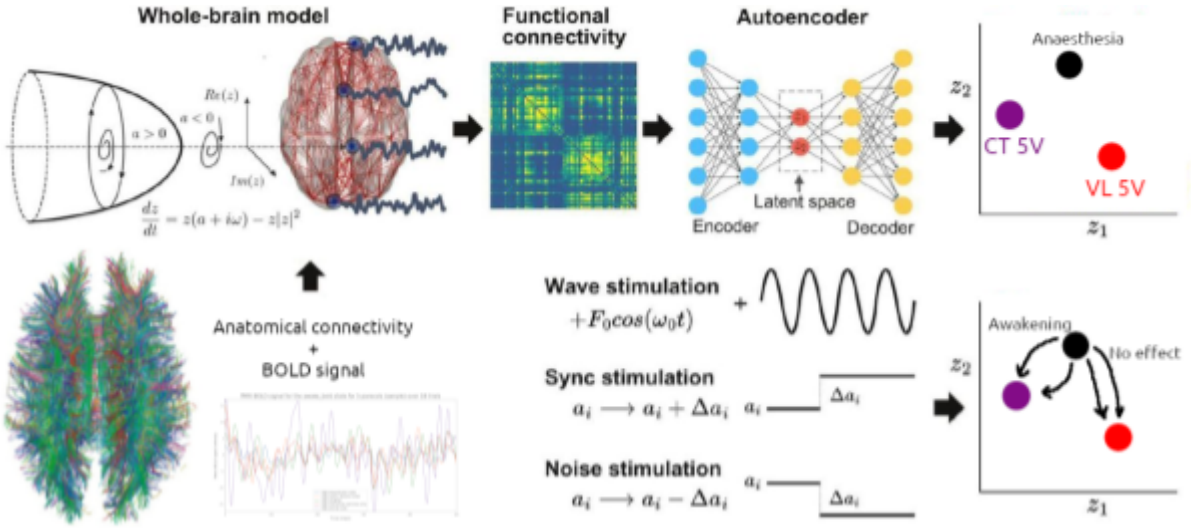

Figure 3a. An overview of the methodology covered. Starting from the BOLD signal and the anatomical connectivity (obtained due to DTI, tractography, tracer studies or others and the parcellation), a whole-brain model is created. In this case, the whole-brain model is a Hopf model. The model is fitted to the data with parameter exploration and genetic algorithms and is then used to generate data to train a Variational Autoencoder (VAE). The latent layer of the autoencoder is then used as a "latent space" where we can differentiate the different brain states. Through the application of different perturbations to the model, transitions between the different brain states can be observed in the latent space. *(Figure adapted from [34] and [39])*

3.1. Data acquisition

The fMRI data were obtained from [24]. In particular, the data used for analysis is that of monkey T. For visualisation purposes, the data of monkey B is also displayed in some figures. As a summary:

Functional MRI data of rhesus macaques (*Macaca mulatta*) was recorded during awake resting-state (for monkey B) and anaesthesia, after the implantation of a DBS electrode (Medtronic, Minneapolis, MN, USA, lead model 3389) (monkey T). The stimulation via DBS was done in the Central Thalamus (CT) as the target area, in order to recover awareness and arousal from the monkey, and in the Ventrolateral nuclei of the Thalamus (VL) as the control condition. Stimulation to both areas was performed at 3V (low condition) and 5V (high condition), resulting in 5 different recording conditions for monkey T (anaesthesia without stimulation, CT stimulation 3V, CT stimulation 5V, VL stimulation 3V, VL stimulation 5V). The anaesthetic used was propofol, induced by a mixture of ketamine and dexmedetomidine. A muscle blocker (cisatracurium) was used to avoid artefacts. The DBS leads were later reconstructed by both computational reconstruction and histological studies. The fMRI data were parcellated under the CoCoMac parcellation (82 cortical areas). The BOLD time series was band-passed filtered between 0.0025 and 0.05 Hz in order to remove physiological artefacts and noise, being this band one that shows to contain information relevant to the function of the brain.

3.2. Anatomical connectivity

The anatomical connectivity used is a fully weighted whole-cortex macaque structural connectivity matrix (connectome) derived by combining the information from fibre-tracing and tractography [40]. The connectome is publicly available:

https://zenodo.org/record/1471588#.X44C6dAzY2x. The tractography algorithm was optimised to best reproduce the weighted but partial-cortex tracer connectome from Markov et al. [41], before estimating the whole-cortex connectome weights. For details see Shen et al. [40].

3.3. Whole-Brain Models

The whole-brain model used in this study is the denominated Hopf model [3]. It consists of a network of nonlinear oscillators (Stuart-Landau oscillators) coupled together by the anatomical connectivity. To implement the model, two main assumptions are made: (i) the dynamics of macroscopic neural masses can range from fully synchronous to a stable asynchronous state driven by random fluctuations; (ii) fMRI can capture the dynamics from both regimes with sufficient accuracy and precision to be modelled by the equations. Each oscillator represents the dynamics of one of the 82 regions of the parcellation used to obtain the anatomical connectivity. The Hopf bifurcation changes the qualitative nature of the solutions from a stable fixed point to a limit cycle, allowing the model to represent the emergence of self-sustained oscillations. The dynamics of brain region *j* were modelled by the following equations:

$$\frac{dx_j}{dt} = (a_j + x_j^2 + y_j^2)x_j - \omega_j y_j + G\sum_{i\neq j} C_{ij}(x_i - x_j) + \beta\eta_j$$

$$\frac{dy_j}{dt} = (a_j + x_j^2 + y_j^2)y_j + \omega_j x_j + G\sum_{i\neq j} C_{ij}(y_i - y_j) + \beta\eta_j$$

In the equations, ω_j is the intrinsic oscillation frequency of node *j*. This intrinsic oscillation frequency is the highest-powered frequency from the band-passed fMRI signal of each area. The parameter a_j is known as the bifurcation parameter, and the dynamics of the system change from a stable fixed point to a limit cycle depending on the value of this parameter. The coupling comes into play in the third term of the equations. Nodes *i* and *j* are coupled by C_{ij} (the *i,j* entry of the SC matrix). β_j is fixed at 0.04 and represents the scaling factor of additive Gaussian noise (η_j) for each node. The parameter *G* is the global coupling factor and scales the SC equally for all nodes to ensure oscillatory dynamics for *a>0*. The empirical SC matrix was globally scaled to a maximum of 0.2 (weak coupling assumption). The equations were integrated using the Euler-Maruyama algorithm with a time-step of 0.01 seconds to simulate the empirical fMRI signals. In this model, when *a* is close to the bifurcation, the additive Gaussian noise gives rise to complex dynamics as the system continuously switches between both sides of the bifurcation.

3.3.1. Fitting the models to the empirical data

The empirical observable to be fitted by the model was the trial-averaged single-subject functional connectivity (FC), also called the zero-lag covariance matrix. A model was created for each experimental condition for monkey T (anaesthesia, low CT stimulation, high CT stimulation, low VL stimulation and high VL stimulation), and an extra model was fitted to the awake condition of monkey B for visualisation purposes.

The above described whole-brain model was used to simulate regional fMRI signals. The first step towards fitting the model is exhaustively exploring the parameter space for the coupling parameter (G), using a homogenous condition for the model with the bifurcation parameters close to the bifurcation ($a \approx 0$, $a = -0.02$), which according to previous literature leads to the best fit in a homogenous model [3]. The metric used to fit the data was the structural similarity index (SSIM) [42] as implemented by the scikit-image package under the SciPy package [43] in the Python programming language. This metric had been used previously for the purpose of data augmentation [39]. From the exhaustive exploration, the *G* value that maximises the SSIM is selected for each model.

Afterwards, a fitting of the *a* parameters is done by the application of a genetic algorithm following previous work [44]. In this case, instead of lowering the dimensionality of the parameter space for the bifurcation parameters by using resting-state networks or other networks, the dimensionality was reduced from the 82 different areas to 41 parameters by the assumption that homotopic areas in both hemispheres would have a similar enough cytoarchitecture to have similar local dynamics. In a work by Shamir and colleagues, they report that the cytoarchitecture of homotopic areas is highly symmetric after their clustering [45]. The drawback of this assumption is that it cannot represent asymmetries (which have been reported by some literature, for example, by Kringelbach and colleagues (2023) regarding interhemispheric connections [33]). Then, the individuals of each generation would have a genome of 41 parameters, which are used to run the Hopf model and obtain the SSIM between the model and the empirical data. Each generation, comprised of 20 individuals, would then be ranked by the GoF and the two highest models go on intact to the next generation (elitist selection). The rest of the 18 individuals are made up of children of the individuals from the previous generation. Parents are selected with different probabilities depending on their GoF, and their genomes are mixed to create a new individual. The genomes of these individuals have the chance to mutate, changing the value of a random parameter following a normal probability distribution centred on the respective value. This process goes on for 200 generations or until the highest GoF has not varied more than a given value for the last 10 generations (i.e. if it has converged). This process is run 100 times and the parameters of the model that outperforms the rest are chosen as the models to run the data augmentation.

3.3.2. Data augmentation

The fitted models were used to generate surrogate data for later training of the VAE. 5000 FC matrices were generated for each condition for monkey T for training. Additionally, another 500 FC matrices were generated for visualisation, in order for them to be not from the training or validation sets, i.e. to be previously unseen by the VAE model. The awake condition for monkey B was also used for the generation of 500 matrices for visualisation.

3.3.3. *In silico* stimulation of the brain

Stimulations to the models were done following the procedures in Sanz Perl et al. (2022) [34]. Two main procedures were used to stimulate an area: applying an extra oscillatory term (wave stimulation), and shifting the value of the bifurcation parameter of said area (noise stimulation if decreasing the value, sync stimulation if increasing the value). The wave stimulation is given by an additive term given by $F_0 \cos(\omega_0 t)$. F_0 is the forcing

amplitude, and it varies from 0 to 2 in order to parametrize the perturbation as a function of the strength. ω_0 is the natural frequency of the node that is being stimulated.

3.3.4. Variational Autoencoder (VAE)

The VAE implementation used follows that of Sanz Perl and colleagues (2022) [36]. The data encoded into the latent space are the FC matrices, which for training purposes are then decoded and the reconstruction error is minimised. Since this is a VAE, it maps inputs to probability distributions in the latent space, which are regularised during this training process. This allows the latent space to become meaningful, meaning that coordinates not used by empirical data, can still be decoded and produce interpretable results.

The architecture of the VAE consists of three parts: the encoder networks, a middle variational layer and the decoder network. The architecture follows that of the mentioned paper, to which one can refer to for further information [36].

The training is done with the generated matrices from monkey T under the conditions of anaesthesia, high CT stimulation and high VL stimulation. The training procedure consisted of batches of 256 samples and 5 epochs. Later, the extra FCs from states of low CT, low VL and awake (from monkey B) are encoded for visualisation in the latent space.

4. Summary of previous work

During these last 18 months I have been working on the described pipeline and dataset. Right now I have concluded the first round of exhaustive stimulations. In the next subsections, I will describe the results I have obtained step by step and will comment on ideas that I might want to pursue further.

4.1. Individualised monkey models

Following the pipeline described for fitting the model to the data, an exhaustive search for a *G* value that maximises the SSIM for an homogeneous model returns different values of *G* for each condition. As an example, see figure 4.1a. It shows different metrics for comparison between the empirical FC and the simulated FC for the different *G* values for the anaesthesia condition. In this case, as we chose SSIM to be our metric, the *G* value chosen for the anaesthesia condition is 0.45. This is done for all conditions even when only the plots for the anaesthesia condition are shown here.

Once an appropriate *G* value has been obtained, the *a* parameters are fit as described. An evolution of the GoF can be seen in figure 4.1b for the anaesthesia condition. In the same figure the empirical FC matrix for the same condition is shown, alongside a simulated FC matrix produced with the best fitted model. A GoF close to 0.75 is achieved for this condition.

On figure 4.1c violin plots for the different fitted conditions are shown, revealing the level of fitting for the different conditions. They show the results from all the 100 runs from the genetic algorithm. Only the parameters from the best fit were used for the following processes.

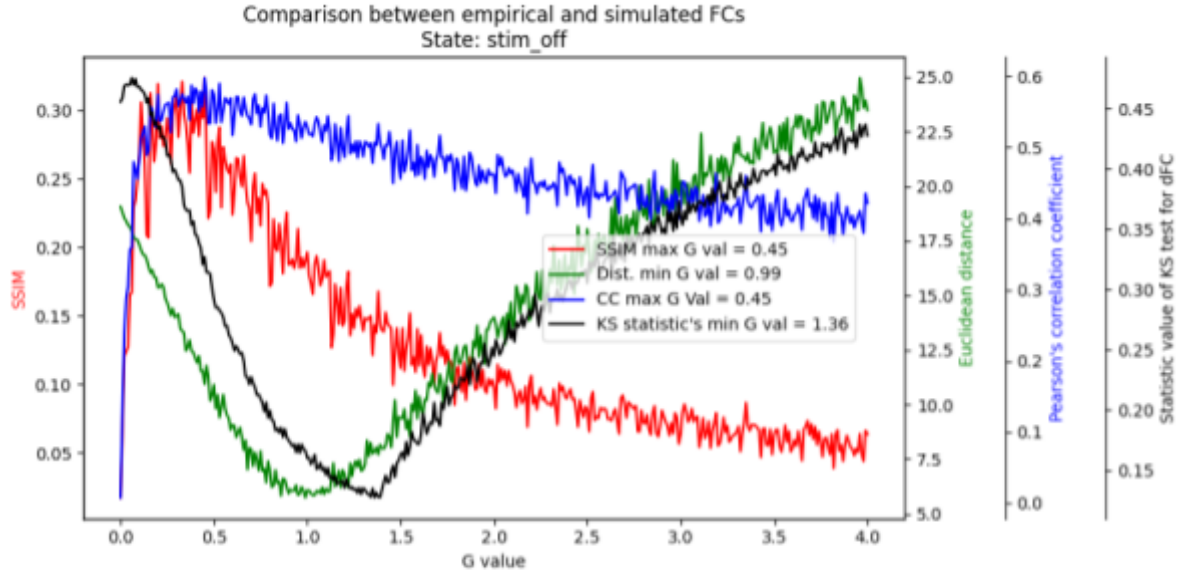

Figure 4.1a. Different metrics as GoF comparing the empirical data with the simulated data. Both the correlation coefficient (CC, blue) and the structural similarity index (SSIM, red) compare the grand average FC, while the euclidean distance (dist, green) and the kolmogorov-smirnov distance (KS statistic, black) compare the dynamical functional connectivity

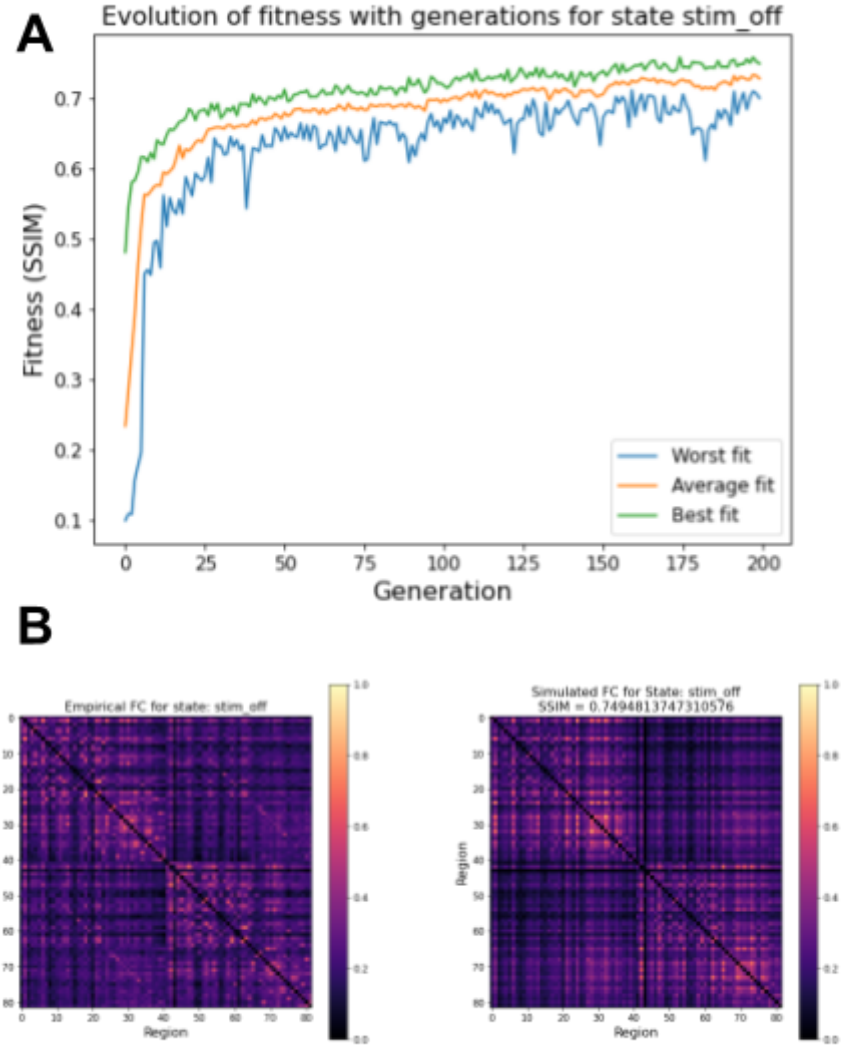

Figure 4.1b. Genetic algorithm fitting for the anaesthesia condition. (A) The evolution of the GoF over all the generations from the genetic algorithm. (B) On the left the empirical FC for the anaesthesia condition is shown. On the right, an FC generated with the fitted model is shown. The SSIM between these two matrices is approximately 0.75.

For each condition the distribution of the bifurcation parameters obtained across all genetic algorithm runs was obtained, and statistical analyses were performed between conditions. That being said, due to the dependence on *G* that the bifurcation parameters have, we believe that making inferences from those analyses is not entirely solid. For that reason, those analyses have been left out.

Regarding this part, I believe that I would like to try different methods of optimization and parameter tuning in order to make it less computationally expensive and be able to add asymmetry in the model. Also, I believe that using metrics that take into account the dynamics of the system might improve not only the fitting but also will end up being more informative than using grand time averages. Also, the implementation of weighted connections (GCAT framework [33]) I believe would improve the model quite a bit.

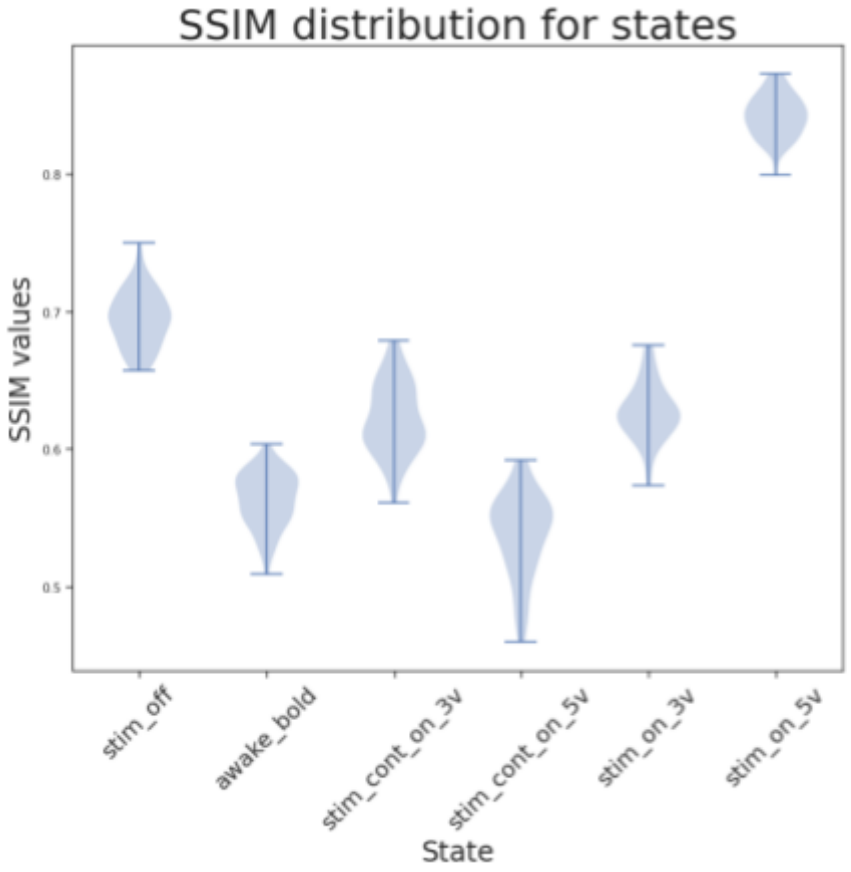

Figure 4.1c. Distributions of SSIM values for the best fit in each run of the genetic algorithm. As can be seen, the distributions are differentiated meaning that the assumptions from the model work better for some conditions. Only the best result out of each distribution was used later on.

Description of states: stim_off is the anaesthesia state, awake_bold is the awake state (from monkey B), stim_cont_on_3v and stim_cont_on_5v are the VL stimulations, low and high voltage (respectively), and both stim_on_3v and stim_on_5v are CT stimulation (low and high voltage respectively).

4.2. Representation in the latent space

Once the models have been fit, FC matrices are generated for the training of the VAE. The VAE is trained on three conditions as previously mentioned (anaesthesia, CT high stimulation, VL high stimulation) and then the rest of the states of monkey T (along with the awake state from monkey B) are encoded for visualisation purposes. On figure 4.2a the latent space with the validation data and with the new data (not previously seen by the model) are shown. The relative orientations and distances of the states are kept through

different training sessions, showing that the states have underlying properties that tie them together.

Further work in this area regards the implementation of the VAE and the study of variations of the system. One interesting variation that is currently being discussed with some collaborators is the ß-VAE, which weights the two training conditions (reconstruction and latent space normalisation) and might lead to a better organised latent space. Another route I would like to explore is the study of the topology/geometry of these latent spaces, since they seem to be robust through training sessions, which I feel is very interesting and might allow us to extract more insights from the different brain states and their relations.

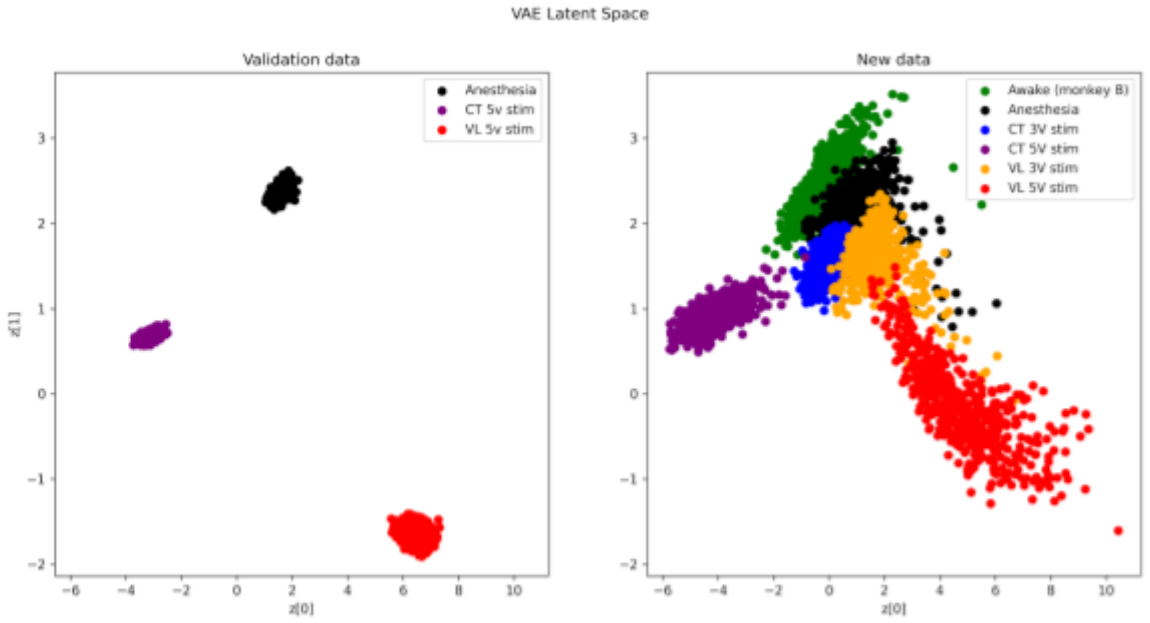

Figure 4.2a. The left panel shows the validation data as encoded in the latent space by the encoder from the VAE. The right panel shows the encoding of previously unseen data by the VAE. Each point is an encoded FC matrix.

4.3. Model stimulation and awakening

After having the Whole-brain models fitted and the VAE trained, the anaesthesia state was exhaustively stimulated on homotopical regions with different methods as previously mentioned. Figures 4.3a and 4.3b show some of the most interesting results, with some transitions to VL and CT stimulated states, as well as the area where the awake condition is expected to be. It can be seen that different stimulations drive the system to the stimulations that were performed experimentally, but also to where we would expect the awake condition to lay (where the awake condition from the other monkey is). Even more impressive is that stimulating V2 (areas 18 and 59), the brain area that was significantly activated during the low VL stimulation, the system is moved towards the place in latent space where the FC matrices from the model fit to that type of experimental stimulation are. If we try to stimulate only the right V2 (as reported by the paper [24]), we do not observe the result. This could be to the left V2 activating sub-significantly or because in the structural connectivity the long-range connections are usually misrepresented and stimulating the other hemisphere in the same way would make up for this effect. Previous studies have

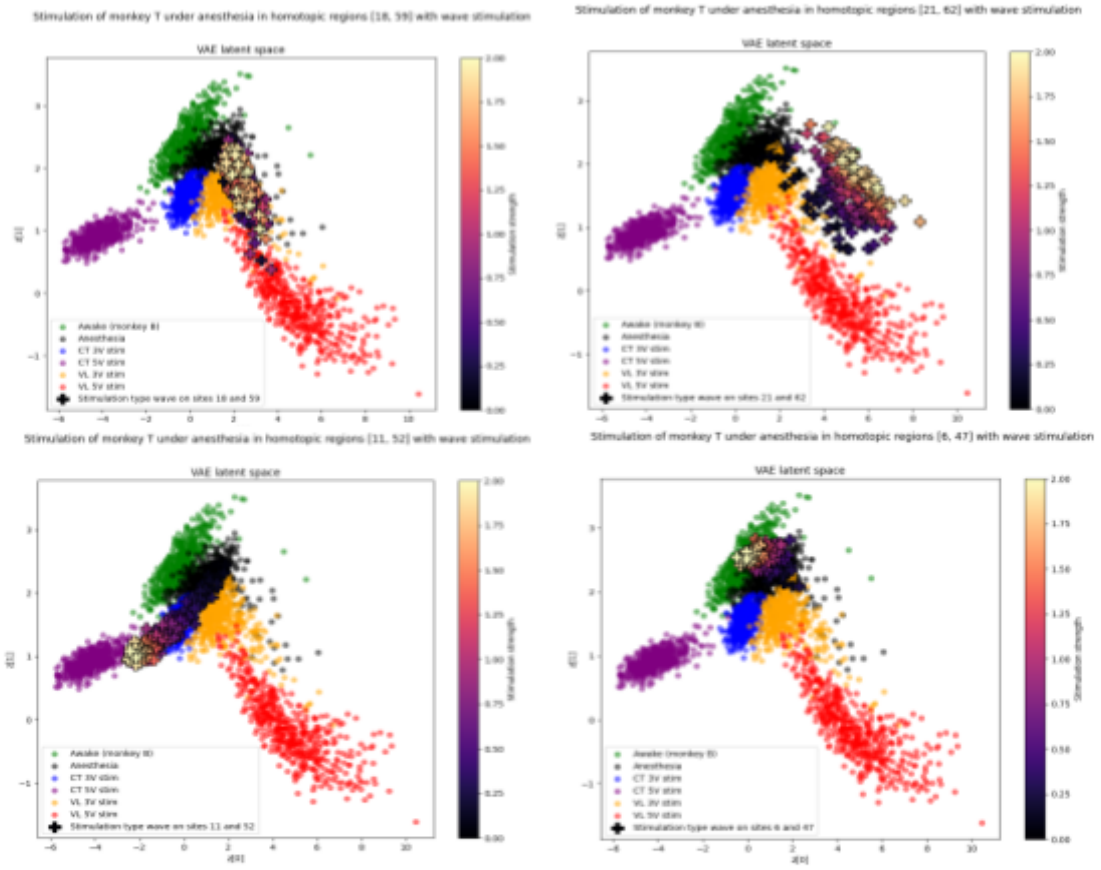

Figure 4.3a. Wave stimulation on different brain areas of our model. Top left shows the stimulation of area 18 (V2) and the resulting states being close to the VL low stimulation, as reported. Top right shows an stimulation that takes the brain to a region (the dorsal part of the anterior visual area, VACd) that is not described by the previously done stimulations, but opposes both the target stimulation (CT) and the awake state of the other monkey. Bottom left shows a stimulation that moves the system through the path defined by both CT stimulations (low and high). Here we see the phenomenon of saturation. The area stimulated is the ventrolateral premotor cortex (PMCvl). Bottom right shows a stimulation that moves the system towards where we would expect the awake state to be (where the awake state of monkey B is). The area stimulated is the central temporal cortex (TCc).

been done on the functional importance of these long range connections. For example, the study conducted by Deco and colleagues (2021) [46] using a similar whole-brain approach shows the information processing importance of some long range connections.

Some other types of stimulation are being studied, such as different waveforms (different spectral properties), different frequencies, adding more noise, or changing the bifurcation parameters in other ways. Also, following the study by Kringelbach and colleagues (2023) [33] in which they develop a way to weight individual connections to fit the model, it would be interesting to be able to define stimulations based on modifications to these weights.

Apart from that, and for the purpose of modelling the experiment, a tractography of the connections between the CT and VL regions of the thalamus and the cortex is being sought in order to be able to stimulate those areas and see whether the system behaves as it did in the experiment and what type for stimulation would best fit the DBS data.

Current work on this is being done by decoding all the coordinates of the latent space to see how a FC matrix would look at any point, as well as by obtaining metrics for the states and the stimulations (following [34] and [36]).

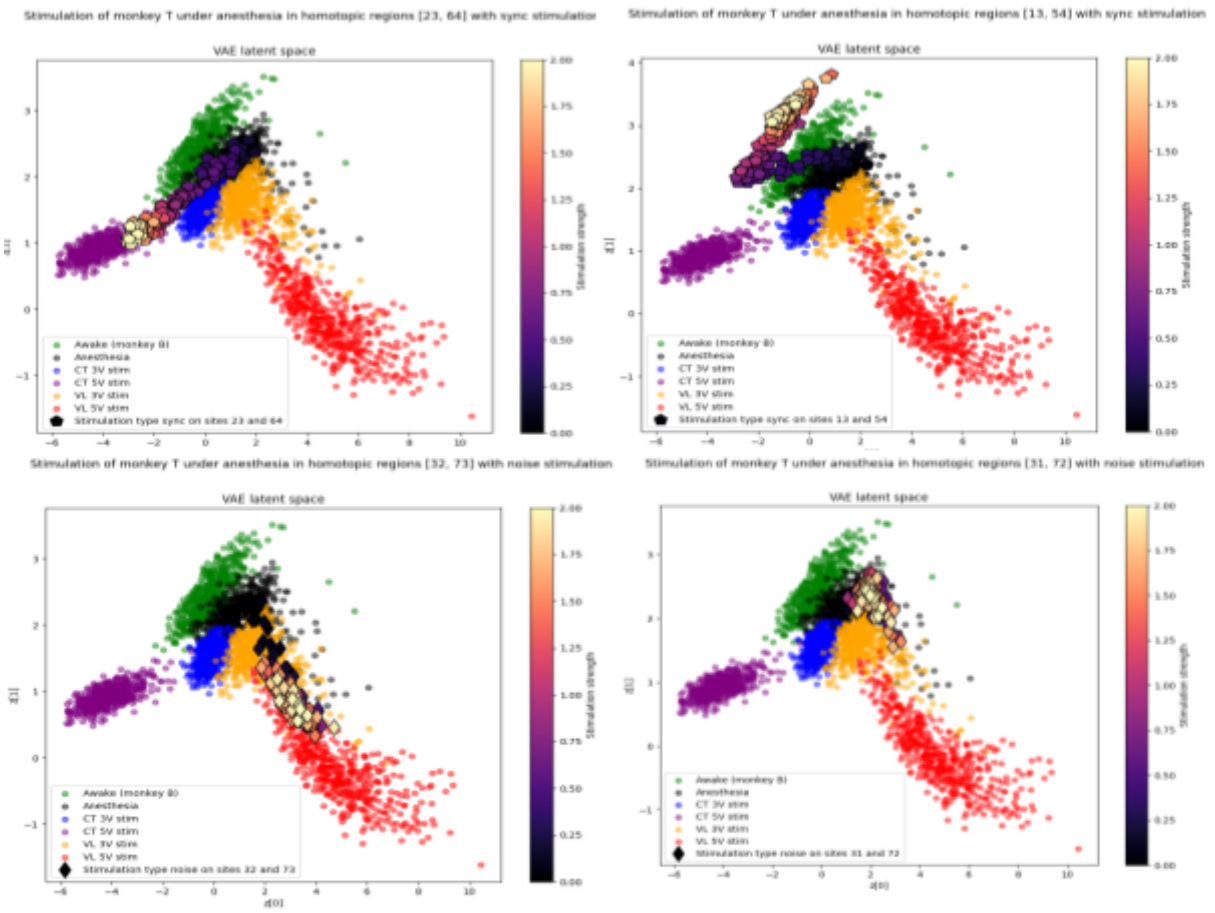

Figure 4.3b. Stimulations that modify the local dynamics (bifurcation parameter). Top left: a sync stimulation to the centrolateral prefrontal cortex (PFCcl) that moves the state towards the place where the high CT stimulation model lays. Top right: a sync stimulation to the posterior insula (Ip) that moves the model towards where we would expect the awake condition to be. This stimulation moves it past the awake condition for monkey B. Bottom left: Few noise conditions move the state towards where the VL stimulations lay, like thi one to the inferior parietal cortex (PCi). Bottom right: Most noise stimulations display a similar pattern, not moving the system in any direction, like in this case of stimulation to the primary motor cortex (M1).

5. Significance

With this thesis, we expect to close the gap between whole-brain modelling and the usage of perturbations to model real-life brain stimulation, and the actual experimental stimulations that take place.

Another great aspect of this framework is that we do not force the system to display one state or the other. By purely trying to create a model that behaves like the data, we end up being able to separate the consciousness states by an unsupervised neural network, showing that there are underlying dynamics that make those states different and that stimulation of states moves the system between them.

In the future, this could be a step towards the path of new personalised treatments for people with disorders of consciousness, by showing that individual whole-brain models not only can be developed but can also be used to find targets for brain stimulation.

References

1. Kelso, J.A.S. (1995) **Dynamic patterns: the self-organization of brain and behavior**. *Complexity 2, 45-46*
2. Kringelbach, M. L., & Deco, G. (2020). **Brain States and Transitions: Insights from Computational Neuroscience**. *Cell Reports, 32(10), 108128.* https://doi.org/10.1016/j.celrep.2020.108128
3. Deco, G., Kringelbach, M. L., Jirsa, V. K., & Ritter, P. (2017). **The dynamics of resting fluctuations in the brain: metastability and its dynamical cortical core**. *Scientific Reports, 7(1).* https://doi.org/10.1038/s41598-017-03073-5
4. Tognoli, E., & Kelso, J. (2014). **The Metastable Brain**. *Neuron, 81(1), 35–48.* https://doi.org/10.1016/j.neuron.2013.12.022
5. McIntosh, A. R., Kovacevic, N., & Itier, R. J. (2008). **Increased Brain Signal Variability Accompanies Lower Behavioral Variability in Development**. *PLoS Computational Biology, 4(7), e1000106.* https://doi.org/10.1371/journal.pcbi.1000106
6. Rabinovich, M. I., Huerta, R., Varona, P., & Afraimovich, V. S. (2008). **Transient Cognitive Dynamics, Metastability, and Decision Making**. *PLoS Computational Biology, 4(5), e1000072.* https://doi.org/10.1371/journal.pcbi.1000072
7. Cabral, J., Hugues, E., Sporns, O., & Deco, G. (2011). **Role of local network oscillations in resting-state functional connectivity**. *NeuroImage, 57(1), 130–139.* https://doi.org/10.1016/j.neuroimage.2011.04.010
8. Hellyer, P. J., Scott, G., Shanahan, M., Sharp, D.J., Leech, R. (2015). **Cognitive Flexibility through Metastable Neural Dynamics Is Disrupted by Damage to the Structural Connectome**. *Journal of Neuroscience, 35 (24) 9050-9063.* https://doi.org/10.1523/JNEUROSCI.4648-14.2015
9. Kringelbach, M.L., McIntosh, A.R., Ritter, P., Jirsa, V.K., Deco, G. (2015). **The Rediscovery of Slowness: Exploring the Timing of Cognition**. *Trends in Cognitive Sciences, 19(10), 616-628.* https://doi.org/10.1016/j.tics.2015.07.011
10. Signorelli, C. M., Szczotka, J., & Prentner, R. (2021). **Explanatory profiles of models of consciousness - towards a systematic classification**. *Neuroscience of Consciousness, 2021(2).* https://doi.org/10.1093/nc/niab021
11. Llinás, R., Ribary, U., Contreras, D., & Pedroarena, C. (1998). **The neuronal basis for consciousness**. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 353(1377), 1841–1849.* https://doi.org/10.1098/rstb.1998.0336
12. Oizumi, M., Albantakis, L., Tononi, G. (2014). **From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0**. *PLOS*

Computational Biology 10(5)
https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003588
13. Dehaene, S., Naccache, L. (2001) **Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework,** *Cognition, 79(1-2), 1-37*
https://doi.org/10.1016/S0010-0277(00)00123-2
14. Teasdale, G., Maas, A., Lecky, F., Manley, G., Stocchetti, N., & Murray, G. (2014). **The Glasgow Coma Scale at 40 years: standing the test of time**. *The Lancet Neurology, 13(8), 844–854.*
https://doi.org/10.1016/s1474-4422(14)70120-6
15. Koch, C., & Tsuchiya, N. (2007). **Attention and consciousness: two distinct brain processes**. *Trends in Cognitive Sciences, 11(1), 16–22.*
https://doi.org/10.1016/j.tics.2006.10.012
16. Tassi, P., & Muzet, A. (2001). **Defining the states of consciousness.** *Neuroscience & Biobehavioral Reviews, 25(2), 175–191.*
https://doi.org/10.1016/s0149-7634(01)00006-9
17. De Filippi, E., Escrichs, A., Càmara, E., Garrido, C., Marins, T., Sánchez-Fibla, M., Gilson, M., & Deco, G. (2022). **Meditation-induced effects on whole-brain structural and effective connectivity**. *Brain Structure and Function, 227(6), 2087–2102.*
https://doi.org/10.1007/s00429-022-02496-9
18. Escrichs, A., Perl, Y. S., Uribe, C., Camara, E., Türker, B., Pyatigorskaya, N., López-González, A., Pallavicini, C., Panda, R., Annen, J., Gosseries, O., Laureys, S., Naccache, L., Sitt, J. D., Laufs, H., Tagliazucchi, E., Kringelbach, M. L., & Deco, G. (2022). **Unifying turbulent dynamics framework distinguishes different brain states**. *Communications Biology, 5(1).*
https://doi.org/10.1038/s42003-022-03576-6
19. Cofré, R., Herzog, R., Mediano, P. A., Piccinini, J., Rosas, F. E., Sanz Perl, Y., & Tagliazucchi, E. (2020). **Whole-Brain Models to Explore Altered States of Consciousness from the Bottom Up**. *Brain Sciences, 10(9), 626.*
https://doi.org/10.3390/brainsci10090626
20. Vohryzek, J., Cabral, J., Vuust, P., Deco, G., & Kringelbach, M. L. (2022). **Understanding brain states across spacetime informed by whole-brain modelling**. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 380(2227).*
https://doi.org/10.1098/rsta.2021.0247
21. Edlow, B. L., Claassen, J., Schiff, N. D., & Greer, D. M. (2020). **Recovery from disorders of consciousness: mechanisms, prognosis and emerging therapies**. *Nature Reviews Neurology, 17(3), 135–156.*
https://doi.org/10.1038/s41582-020-00428-x
22. Goldfine, A. M., & Schiff, N. D. (2011). **Consciousness: Its Neurobiology and the Major Classes of Impairment**. *Neurologic Clinics, 29(4), 723–737.*
https://doi.org/10.1016/j.ncl.2011.08.001
23. Alkire, M. T., Hudetz, A. G., & Tononi, G. (2008). **Consciousness and Anesthesia**. *Science, 322(5903), 876–880.*
https://doi.org/10.1126/science.1149213

24. Kringelbach, M. L., Green, A. L., & Aziz, T. Z. (2011). **Balancing the Brain: Resting State Networks and Deep Brain Stimulation**. *Frontiers in Integrative Neuroscience, 5.* https://doi.org/10.3389/fnint.2011.00008
25. Tasserie, J., Uhrig, L., Sitt, J. D., Manasova, D., Dupont, M., Dehaene, S., & Jarraya, B. (2022). **Deep brain stimulation of the thalamus restores signatures of consciousness in a nonhuman primate model.** *Science Advances, 8(11).* https://doi.org/10.1126/sciadv.abl5547
26. Deco, G., & Kringelbach, M. (2014). **Great Expectations: Using Whole-Brain Computational Connectomics for Understanding Neuropsychiatric Disorders**. *Neuron, 84(5), 892–905.* https://doi.org/10.1016/j.neuron.2014.08.034
27. Deco, G., Cruzat, J., Cabral, J., Tagliazucchi, E., Laufs, H., Logothetis, N. K., & Kringelbach, M. L. (2019). **Awakening: Predicting external stimulation to force transitions between different brain states**. *Proceedings of the National Academy of Sciences, 116(36), 18088–18097* https://doi.org/10.1073/pnas.1905534116
28. Sanz Perl, Y., Pallavicini, C., Pérez Ipiña, I., Demertzi, A., Bonhomme, V., Martial, C., Panda, R., Annen, J., Ibañez, A., Kringelbach, M., Deco, G., Laufs, H., Sitt, J., Laureys, S., & Tagliazucchi, E. (2021b). **Perturbations in dynamical models of whole-brain activity dissociate between the level and stability of consciousness**. *PLOS Computational Biology, 17(7), e1009139.* https://doi.org/10.1371/journal.pcbi.1009139
29. Vohryzek, J., Cabral, J., Castaldo, F., Sanz-Perl, Y., Lord, L. D., Fernandes, H. M., Litvak, V., Kringelbach, M. L., & Deco, G. (2023). **Dynamic sensitivity analysis: Defining personalised strategies to drive brain state transitions via whole brain modelling.** *Computational and Structural Biotechnology Journal, 21, 335–345.* https://doi.org/10.1016/j.csbj.2022.11.060
30. Luppi, A. I., Cabral, J., Cofre, R., Mediano, P. a. M., Rosas, F. E., Qureshi, A., Kuceyeski, A., Tagliazucchi, E., Raimondo, F., Deco, G., Shine, J., Kringelbach, M. L., Orio, P., Ching, S., Perl, Y. S., Diringer, M. N., Stevens, R. D., & Sitt, J. (2023). **Computational modelling in disorders of consciousness: closing the gap towards personalised models for restoring consciousness**. *PsyRxiv.* https://doi.org/10.31234/osf.io/5dp3j
31. Deco, G., Jirsa, V. K., Robinson, P. A., Breakspear, M., & Friston, K. (2008). **The Dynamic Brain: From Spiking Neurons to Neural Masses and Cortical Fields**. *PLoS Computational Biology, 4(8), e1000092.* https://doi.org/10.1371/journal.pcbi.1000092
32. Piccinini, J., Ipiñna, I. P., Laufs, H., Kringelbach, M., Deco, G., Sanz Perl, Y., & Tagliazucchi, E. (2021). **Noise-driven multistability vs deterministic chaos in phenomenological semi-empirical models of whole-brain activity**. *Chaos: An Interdisciplinary Journal of Nonlinear Science, 31(2), 023127.* https://doi.org/10.1063/5.0025543
33. Kringelbach, M. L., Perl, Y. S., Tagliazucchi, E., & Deco, G. (2023). **Toward naturalistic neuroscience: Mechanisms underlying the flattening of brain**

hierarchy in movie-watching compared to rest and task. *Science Advances, 9(2).* https://doi.org/10.1126/sciadv.ade6049

34. Perl, Y. S., Fittipaldi, S., González Campo, C., Moguilner, S., Cruzat, J., Herzog, R., Kringelbach, M., Deco, G., Prado, P., Ibañez, A., & Tagliazucchi, E. (2022). **Model-based whole-brain perturbational landscape of neurodegenerative diseases.** *BioRxiv.* https://doi.org/10.1101/2022.09.26.509612
35. Idesis, S., Favaretto, C., Metcalf, N. V., Griffis, J. C., Shulman, G. L., Corbetta, M., & Deco, G. (2022). **Inferring the dynamical effects of stroke lesions through whole-brain modeling**. *NeuroImage: Clinical, 36, 103233.* https://doi.org/10.1016/j.nicl.2022.103233
36. Perl, Y. S., Pallavicini, C., Piccinini, J., Demertzi, A., Bonhomme, V., Martial, C., Panda, R., Alnagger, N., Annen, J., Gosseries, O., Ibañez, A., Laufs, H., Sitt, J., Jirsa, V., Kringelbach, M., Laureys, S., Deco, G., & Tagliazucchi, E. (2022). **Low-dimensional organization of global brain states of reduced consciousness**. *BioRxiv*. https://doi.org/10.1101/2022.09.28.509817
37. Perl, Y. S., Bocaccio, H., Pérez-Ipiña, I., Zamberlán, F., Piccinini, J., Laufs, H., Kringelbach, M., Deco, G., & Tagliazucchi, E. (2020). **Generative Embeddings of Brain Collective Dynamics Using Variational Autoencoders.** *Physical Review Letters, 125(23).* https://doi.org/10.1103/physrevlett.125.238101
38. Kingma, D. P., & Welling, M. (2013). **Auto-Encoding Variational Bayes.** ArXiv, https://doi.org/10.48550/arXiv.1312.6114
39. Perl, Y. S., Pallavicini, C., Ipiña, I. P., Kringelbach, M., Deco, G., Laufs, H., & Tagliazucchi, E. (2020). **Data augmentation based on dynamical systems for the classification of brain states**. *Chaos, Solitons & Fractals, 139, 110069.* https://doi.org/10.1016/j.chaos.2020.110069
40. Shen, K., Bezgin, G., Schirner, M., Ritter, P., Everling, S., & McIntosh, A. R. (2019). **A macaque connectome for large-scale network simulations in TheVirtualBrain.** *Scientific Data, 6(1).* https://doi.org/10.1038/s41597-019-0129-z
41. Markov, N. T., Ercsey-Ravasz, M. M., Ribeiro Gomes, A. R., Lamy, C., Magrou, L., Vezoli, J., Misery, P., Falchier, A., Quilodran, R., Gariel, M. A., Sallet, J., Gamanut, R., Huissoud, C., Clavagnier, S., Giroud, P., Sappey-Marinier, D., Barone, P., Dehay, C., Toroczkai, Z., . . . Kennedy, H. (2012). **A Weighted and Directed Interareal Connectivity Matrix for Macaque Cerebral Cortex.** *Cerebral Cortex, 24(1), 17–36.* https://doi.org/10.1093/cercor/bhs270
42. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P. (2004) **Image quality assessment: from error visibility to structural similarity**, *IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600-612* https://doi.org/10.1109/TIP.2003.819861
43. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., . . . SciPy 1.0 Contributors. (2020) **SciPy 1.0: Fundamental Algorithms for**

Scientific Computing in Python. *Nature Methods, 17(3), 261-272.* https://doi.org/10.1038/s41592-019-0686-2

44. Ipiña, I.P., Kehoe, P.D., Kringelbach, M., Laufs, H., Ibañez, A., Deco, G., Perl, Y.S., Tagliazucchi, E. (2019) **Modeling the relationship between regional activation and functional connectivity during wakefulness and sleep.** *arXiv* https://doi.org/10.48550/arXiv.1907.04412
45. Shamir, I., Assaf, Y., & Shamir, R. (2022). **Clustering laminar cytoarchitecture: in vivo parcellation based on cortical granularity**. *biorXiv.* https://doi.org/10.1101/2022.10.29.514347
46. Deco, G., Sanz Perl, Y., Vuust, P., Tagliazucchi, E., Kennedy, H., & Kringelbach, M. L. (2021). **Rare long-range cortical connections enhance human information processing**. *Current Biology, 31(20), 4436-4448.e5.* https://doi.org/10.1016/j.cub.2021.07.064