Brain dynamics predictive of response to psilocybin for treatment-resistant depression

Jakub Vohryzek\textsuperscript{1,2,3}, Joana Cabral\textsuperscript{1,2,4,5}, Louis-David Lord\textsuperscript{1,2}, Henrique M Fernandes\textsuperscript{1,2}, Leor Roseman\textsuperscript{6}, David J Nutt\textsuperscript{6}, Robin L Carhart-Harris\textsuperscript{6,7}, Gustavo Deco\textsuperscript{3,8,9,10}, Morten L Kringelbach\textsuperscript{1,2,4}

1. Department of Psychiatry, University of Oxford, Oxford, United Kingdom
2. Center for Music in the Brain, Aarhus University, Aarhus, Denmark
3. Center for Brain and Cognition, Computational Neuroscience Group, Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, Spain.
4. Life and Health Sciences Research Institute (ICVS), School of Medicine, University of Minho, Braga, Portugal
5. ICVS/3B’s - PT Government Associate Laboratory, Braga/Guimarães, Portugal
6. Centre for Psychedelic Research, Department of Brain Sciences, Imperial College London, London, United Kingdom
7. Psychedelics Division, Neuroscape, Department of Neurology, University of California San Francisco, California, US
8. Institució Catalana de la Recerca i Estudis Avançats (ICREA), Barcelona, Spain
9. Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany
10. School of Psychological Sciences, Monash University, Melbourne, Australia

Corresponding author:
Jakub Vohryzek, Center for Brain and Cognition, Computational Neuroscience Group, Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, Spain
jakub.vohryzek@upf.edu

Short Title: Brain dynamics in psilocybin for depression
Abstract

Psilocybin therapy for depression has started to show promise, yet the underlying causal mechanisms are not currently known. Here we leveraged the differential outcome in responders and non-responders to psilocybin (10mg and 25mg, 7 days apart) therapy for depression - to gain new insights into regions and networks implicated in the restoration of healthy brain dynamics. We used large-scale brain modelling to fit the spatiotemporal brain dynamics at rest in both responders and non-responders before treatment. Dynamic sensitivity analysis of systematic perturbation of these models enabled us to identify specific brain regions implicated in a transition from a depressive brain state to a healthy one. Binarizing the sample into treatment responders (>50% reduction in depressive symptoms) versus non-responders enabled us to identify a subset of regions implicated in this change. Interestingly, these regions correlate with in vivo density maps of serotonin receptors 5-Hydroxytryptamine 2a and 5-Hydroxytryptamine 1a, which psilocin, the active metabolite of psilocybin, has an appreciable affinity for, and where it acts as a full-to-partial agonist. Serotonergic transmission has long been associated with depression and our findings provide causal mechanistic evidence for the role of brain regions in the recovery from depression via psilocybin.

Keywords: Large-scale Brain Modelling, Psilocybin Treatment, Depression

Introduction
Behavioral differences between healthy and depressed individuals can sometimes be conspicuous but identifying causal contributions from brain dynamics is more challenging. Discrete global brain states, such as those that pertain to sleep, healthy waking consciousness and the psychedelic state, have their own characteristic spatio-temporal dynamics, involving large-scale spatial communities temporally evolving in transient arrangements. With recent advancements in non-invasive neuroimaging techniques, it has become possible to describe complex spatio-temporal dynamics in terms of their spatial and temporal information. Still, one of the challenges for systems neuroscience is to understand what the most appropriate description of such dynamics is and how transition between one state to another is made possible.

A common method for characterizing global brain function, involves assessing how activity is temporally correlated across spatially separate brain areas over an entire recording period, defining static and state-specific ‘functional connectomes’. However, the last decade has brought clear evidence that finer-grained, more dynamic analysis of brain states, can deepen our understanding of their properties and relationship to behavioural states. There is a growing taxonomy of approaches to characterize the dynamics of functional interactions, from data-driven heuristic clustering methods across time, dynamical systems informed phase-locking approaches, Hidden Markov Models to spatio-temporal networks.

From a clinical perspective major depressive disorder (MDD) lends itself ideal for the dynamic analysis as its symptomatology is often defined by ruminative states with self-critical and inflexible periods of thinking. Such behaviour has been linked to the abnormal spatio-temporal organisation of resting state activity. Converging findings have found the overactivation of default-mode network and hypoconnectivity in the control executive network in MDD.
This is consistent with the triple-network model whereby depressive states promote the inward and self-referential experience at the expense of the interactions and attention towards the external environment.\textsuperscript{24,26} In addition, the MDD brain state has been found to reflect greater synchronisation and temporal stability of overall brain activity compared to healthy participants.\textsuperscript{27}

Efforts and methods are advancing for understanding response to neuropharmacological interventions for depression. Understanding the therapeutic actions of interventions promise - not only to shed light onto the mechanistic relationship between various brain states implicated in health and pathology - but also to provide inspiration for the development of new, improved interventions. However, there are considerable practical and ethical challenges for answering mechanistic questions in humans, elevating the use of animal models (with sometimes questionable translational validity) or small clinically relevant populations.\textsuperscript{28,29} One potential advance in this direction, is the use of large-scale brain modeling - as a tool for understanding pathological changes in neuropsychiatric disorders, and, potentially, for clinical diagnosis and prediction.\textsuperscript{4} We are mindful, however, that the predictive power of any model depends on how well it can describe and predict experimental data to which it is fitted.\textsuperscript{30–33}

The present paper focuses on large-scale brain network models where region specific stimulation or excitation can be tested in silico, and used to describe and predict empirical-informed target states\textsuperscript{1} – such the global brain state found in people with intractable depression. These models link regional dynamics with the neuroanatomical structure of the brain to describe the spatio-temporal activity of functional data.\textsuperscript{34} This approach bypasses the ethical constrains of human or non-human animal experimental settings, enabling many types of stimulation to be tested, in order to evaluate the role of regions and their excitation on transit
between states – with relevance to empirical phenomena of interest. The validity of this strategy has previously been demonstrated in the context of sleep and awake states.\textsuperscript{1}

Here, we build on this notion of dynamic sensitivity analysis to gain insight into response to psilocybin therapy for treatment resistant depression. We define brain states in terms of spatial subdivisions and their probability of occurrence across time, characterised as Probabilistic Metastable Substates (PMS). These recurrent metastable substates can be characterized by their probability of occurrence. Beyond the quantitative description of brain states, we wish to understand which brain regions play a prominent role in the recovery from depression after treatment with psilocybin.\textsuperscript{35}

Using data from a trial of psilocybin-therapy for treatment-resistant depression, the sample was binarized into ‘responders’ and ‘non-responders’ to psilocybin therapy. FMRI data was collected before and one-day after the second of two psilocybin-therapy dosing sessions. Using parameters from the empirical data, modeled brain states - and stimulation parameters therein, could then be used to predict treatment response, defined as a >50% reduction in symptom severity from baseline - determined at a key 5-week post-treatment endpoint.\textsuperscript{36}

Psychedelic medicine has shown a promising avenue for treating depression.\textsuperscript{37} For depression treatment, one current hypothesis is that: via a psychedelic drug x psychological intervention combination, there is an increase in global brain flexibility, translating into a window of opportunity for breaking free of negative cognitive biases and associated ruminations.\textsuperscript{38} Indeed, the current research on the acute effects of psychedelic drugs suggests an increase in the repertoire of brain activity substates.\textsuperscript{39–41} From a neuropharmacological perspective, psilocybin – an active compound in magic mushrooms – binds with high affinity to the serotonergic 5-
Hydrotryptamine 2a (5HT\(_{2a}\)) receptors but other serotonergic receptors are also implicated.\(^{42,43}\) Psilocybin acts as an agonist resulting in higher neuronal excitability, modulating the excitatory-inhibitory balance (in favour of excitation) in the cortical brain regions with more 5HT\(_{2a}\) receptors.\(^{44}\) Recently, a large-scale brain computational study focusing on the human brain action of lysergic acid diethylamide (LSD) – which has a similar pharmacology to psilocybin/psilocin – demonstrated, for the first time, the causal impact of 5HT\(_{2a}\) agonism-induced excitation on global brain dynamics.\(^{45}\)

Here, in empirical fMRI data, we identified recurrent brain substates in terms of the PMS space across all the subjects in the pre- and post-treatment conditions. Furthermore, we use a computational large-scale brain model – where each brain area is represented by a Hopf-bifurcation model\(^ {46}\) - to simulate the brain network dynamics in patients before the treatment. Through dynamic sensitivity analysis, we were able to identify brain regions responsible for treatment response at a key 5-week endpoint.\(^ {1,47}\) A priori, we hypothesised that regions permitting transition to a healthy brain state (as predicted by the 5-week endpoint) would relate to the distribution of the 5HT\(_{2a}\) and 5-Hydrotryptamine 1a (5HT\(_{1a}\)) receptors in the human brain, as determined by prior in vivo positron emission tomography (PET) mapping.\(^ {48}\)

**Material and methods**

**Experimental Data**

**Functional MRI**

We carried out the analysis on previously published dataset of patients with treatment-resistant depression undergoing treatment with psilocybin at Imperial College London.\(^ {36}\) In brief, we investigated 15 patients (without excessive movement and other artefacts from the original 19
patients) who were diagnosed with treatment resistant major depression. The MRI scanning sessions were completed pre-treatment with psilocybin and one-day post-treatment with the treatment consisting of two oral doses of psilocybin (10mg and 25mg, 7 days apart). The patients were split into responders and non-responders to the treatment based on the Quick Inventory of Depressive Symptomatology (QIDS) at 5-weeks post-treatment with 6 out of the 15 patients meeting criteria for response.  

**Structural Connectivity**

In this study, white-matter (structural) connectivity of 90 Automated Anatomical Atlas (AAL) brain areas from a previously obtained dataset was used for the large-scale brain network model. In brief, the group consisted of 16 healthy young adults (5 females, mean age: 24.7 ± 2.54). Diffusion Tensor Imaging (DTI) was applied following the methodology described in. Undirected structural connectivity $C_{np}$ was obtained were $n$ and $p$ are brain areas and the connectivity weights are defined as the proportion of sampled fibers in all voxels in region $n$ that reach any voxel in region $p$. Finally, the individually structural connectomes were averaged across the 16 subjects to obtain a group-based template.

**Probabilistic Metastable Substates**

Firstly, we calculated the instantaneous phased relationship between individual brain regions by expressing the demeaned regional fMRI signal $x(t)$ as an analytical signal i.e. in terms of its time-varying phase $\theta(t)$ and amplitude $A(t)$ as $x(t) = A(t) \times \cos (\theta(t))$. We excluded the first and last three timepoints to account for the boundary artefacts introduced by the Hilbert transform. Hence for every time point $t$ and pair of brain regions $n$ and $m$, we obtain the phase coherence matrix dPC as follows:
\[ dPC(n,m,t) = \cos(\theta(n,t) - \theta(m,t)) \] (1)

By decomposing the signal in this way, we can look at when the brain regions \( n \) and \( m \) are aligned with similar angles, \( \cos(0) = 1 \), orthogonal to each other \( \cos(\pi/2) = 1 \) and anti-aligned \( \cos(\pi) = -1 \). As the phase coherence is a measure of undirected connectivity, the phase coherence matrix \( dPC \) is symmetric and all the meaningful information is captured in the upper-triangular matrix.

For further analysis, we used only the \( 1 \times N \) leading eigenvector \( V_1(t) \) of the dPC matrix as described in the Leading Eigenvector Dynamics Analysis.

In detail, at every timepoint \( t \) of the \( dPC(t) \), we performed the eigendecomposition taking the first (most dominant) eigenvector to describe the \( dPC(t) \) pattern. The \( dPC(t) \) is decomposed as \( dPC(t) = V(t)D(t)V^{-1}(t) \) where \( D \) is the diagonal matrix carrying the real-valued eigenvalues and \( V_1(t) \) and \( V^{-1}_1(t) \) are the left and right corresponding orthogonal eigenvectors respectively. The dominant connectivity pattern can be simply reconstructed by the following matrix multiplication \( V(t)V^{-1}(t) \).

The next step was introduced to find recurrent phase-locking substates in the dominant connectivity patterns varying in time. To that end, we clustered all the leading eigenvectors obtained from all the fMRI scans obtained from both responders and non-responders thus achieving a common phase-locking substates for both groups. For the clustering, we used the unsupervised k-means algorithm, with varying cluster number \( k \) from 2 to 10 clusters. The algorithm was run in an n-dimensional space where \( n = 90 \) (number of AAL brain regions) with condition x subject x timepoints number of observations. For each run we randomly initialised the algorithm 20 times to ensure stability in the clustering. Again, by computing the
matrix multiplication of the $1\times N$ cluster centroids $V_{ca}$ as $V_{ca}(t)V_{ca}^T(t)$ we obtain the dominant connectivity pattern of each cluster. In the current analysis, we considered the cluster solution $k = 3$ as an optimal choice between the quality measures - Dunns, Davies-Bouldin and Silhouette Score, Davies (Supplementary Figure 2), and the maximising of the statistical significance between patient groups (p-values).

After calculating the phase-locking states, we defined the probability of occurrence of the individual substates by simply dividing their occurrence in each recording session by the total number of time points recorded (same for all recordings).

**Large-scale Brain Computational Model**

In order to simulate the ultra-slow fluctuations in fMRI signal detected during rest, we used the Landau-Stuart oscillator canonical model, describing the transition from a noisy to an oscillatory dynamics. The so-called supercritical Hopf-bifurcation model was used locally at every brain region (node) to emulate the local dynamics. To achieve a large-scale brain level description, the individual Hopf models were coupled in a structural connectivity (SC) network, describing the large-scale white-matter map of the human brain. The emerging and complex interactions in the large-scale brain network of coupled Hopf models have been shown to describe many aspects known from experimental recordings in MEG and fMRI.

Formally, the normal form of the supercritical Hopf-bifurcation model for a single uncoupled region of interest ($n$) in Cartesian coordinates is described by the following set of coupled equations:
\[ \frac{dx_n}{dt} = (a_n - x_n^2 - y_n^2)x_n - \omega_n y_n + \beta \eta_n(t) \quad (2) \]

\[ \frac{dy_n}{dt} = (a_n - x_n^2 - y_n^2)y_n + \omega_n x_n + \beta \eta_n(t) \quad (3) \]

with \( \beta \eta_n(t) \) being the Gaussian noise with standard deviation of \( \beta = 0.02 \). The bifurcation parameter \( a \) positions the system at the supercritical bifurcation point when \( a = 0 \), noise activity governed by \( \beta \eta_n(t) \) in regime when \( a < 0 \), and stable limit cycle with oscillatory behaviour of frequency defined by \( f_n = \omega_n/2\pi \) when \( a > 0 \). The values of the intrinsic frequency \( \omega \) were calculated from the experimental fMRI signals in the 0.04 – 0.07 Hz band by taking the peak frequency of the Gaussian-smoothed power spectrum of each brain area.

To describe the coupled large-scale brain computational model, we introduced the coupling term (modelled as the common difference coupling i.e., describing the linear term of a general coupling function) between the individual nodes weighted by the corresponding values of the SC matrix. To be noted, we do not consider the next nonlinear coupling term following Taylor expansion of the full coupling, in case the linear coupling is non-existent.\(^{57,58}\) The equations 2 and 3 can be hence expanded as follows:

\[ \frac{dx_n}{dt} = (a_n - x_n^2 - y_n^2)x_n - \omega_n y_n + G \sum_{p=1}^{N} C_{np} (x_p - x_n) + \beta \eta_n(t) \quad (4) \]

\[ \frac{dy_n}{dt} = (a_n - x_n^2 - y_n^2)y_n + \omega_n x_n + G \sum_{p=1}^{N} C_{np} (y_p - y_n) + \beta \eta_n(t) \quad (5) \]
where $C_{np}$ is the SC weight between node $n$ and $p$, and $G$ is the global coupling weight with equal contribution between all the nodal pairs. The SC matrix was rescaled to have the mean value $<C> = 0.2$ in order to be consistent with previous literature’s range of parameters.\textsuperscript{1,46} The simulated signal is described by the $x_n$ equation for every node $n$. The variables $G$ and $a$ are the control parameters used for the model fitting to the experimental data and the stimulation protocol respectively.\textsuperscript{1,46}

**Objective Function**

In order to validate the simulated signal different realisations of the experimental data can be used.\textsuperscript{30} The most standard approach is comparison of the simulated data with grand-averaged static functional connectivity as computed by the Pearson correlation \textsuperscript{34,59} or metastability defined as the standard deviation of the Kuramoto Order Parameter (Supplementary Figure 4 - Metastability). To account for the temporally varying nature of the Blood Oxygen Level Dependent (BOLD) signal, recent literature has focused on the comparison between the simulated and empirical Functional Connectivity Dynamics (FCD) spectrums (quantified by Kolmogorov-Smirnov distance) i.e. the distributions of the cosine distance between the consecutive timepoints as described by the leading eigenvector\textsuperscript{45,46} (Supplementary Figure 4 – Functional Connectivity Dynamics). As alluded to in the previous section, the fMRI signals organise into spatially meaningful phase-locking states. Here, we compare the simulated data to the probabilities of occurrence of the phase-locking states found in the experimental recordings.\textsuperscript{1} We used the symmetrised Kullback-Leibler Divergence (KL divergence) of the simulated and empirical probabilities of occurrence as follows:

$$KL(P_{emp}, P_{sim}) = 0.5 \left( \sum P_{emp}(i) \ln \frac{P_{emp}(i)}{P_{sim}(i)} + \sum P_{sim}(i) \ln \frac{P_{sim}(i)}{P_{emp}(i)} \right)$$ (6)
with $P_{emp}$ and $P_{sim}$ being the empirical and simulated probabilities of occurrence of the same phase-locking states respectively.

**Statistical Analysis**

We use non-parametric permutation tests for the experimental analysis with 1000 permutations. For the receptor analysis we used standard Pearson’s correlation. Both tests were implemented in MATLAB.

**Results**

In summary, a quantitative characterization of the spatio-temporal dynamics recorded with fMRI was obtained using leading eigenvector dynamics analysis (LEiDA), resulting in the definition of Probabilistic Metastable Substates (PMS) (Figure 1A), whose probability of occurrence was compared across conditions (i.e., within-subjects design – therefore, before versus after treatment). We then constructed two large-scale brain models representative of the pre-treatment brains to psilocybin therapy. This was done by fitting their PMS descriptions to those obtained from the experimental data (Figure 1B). Finally, a dynamic sensitivity analysis was implemented to both responder and non-responder pre-treatment models to identify the brain regions that permit a transition to the healthy PMS (described by responders’ (as predicted by the 5-week endpoint) one-day post treatment brains) (Figure 1C and 1D).

**Placeholder for Figure 1**
As described in the methods section, we computed the PMS pre- and post-treatment with psilocybin (where ‘post’ = 1 day post psilocybin dosing session two), for both responders and non-responders (determined 5 weeks hence). Here, we focused on a three-substate solution – the lowest k-level with statistically significant differences between the two groups as well as optimal quality measures across clustering solutions (Supplementary Figure 2). When contrasting responders versus non-responders, the occurrence of substate 3 was significantly different pre- versus post-treatment (p = 0.0258, signed rank-sum test), as well as in the post-treatment data alone (p = 0.0141, rank-sum test; Figure 2A). Furthermore, we also computed the Global Brain Connectivity (GBC), metastability and FCD measures (Supplementary Figure 1). These results clearly indicated the necessity of considering both spatial and temporal dimensions to differentiate between conditions as GBC, synchrony and metastability show non-significant results. Conversely, the FCD measure showed significant differences in the temporal similarities of spatial patterns between pre- and post-treatment responders (p = 0.0163, signed rank-sum permutation test), and pre- and post-treatment non-responders with post-treatment responders respectively (p = 0.0183 and p = 0.0273, rank-sum permutation test), further supporting the use of spatio-temporal measures to capture the alterations in large-scale brain dynamics across conditions.

Placeholder for Figure 2

To obtain large-scale brain computational models representative of the two groups of patients (responders and non-responders before treatment), we first defined a generalized brain network model, where each of 90 cortical and subcortical brain regions (defined using automated anatomical labelling) was described by a Stuart-Landau oscillator (see methods), and regions were coupled according to realistic structural connectivity obtained from diffusion MRI.
To adjust the model to each group of patients, first the intrinsic frequency of each brain region was set to the peak frequency in fMRI signals averaged across patients in the same group (Supplementary Figure 3). Subsequently, the global coupling parameter, $G$, was tuned to optimize each model to its appropriate working point. This was achieved by minimizing the divergence between the experimental and simulated PMS spaces - see Figure 2B. In Supplementary Figure 4, we report optimisation curves for other observables such as the static FC, metastability and FCD. For the responders and non-responders before treatment, we found $G = 0.185$ (KL divergence $= 0.0064$) and $G = 0.165$ (KL divergence $= 0.0187$) respectively to minimise the difference. Figure 2C shows on the left, the experimental results for both groups before treatment; in the middle, the optimal simulated fits for both groups, and on the right the experimental results after treatment (with the results of responders after treatment serving as the target PMS for rebalancing). These findings clearly demonstrate the large-scale model's ability to fit the spatio-temporal dynamics, as described by the PMS space, of the studied groups of responders and non-responders before the treatment.

Subsequently, we considered dynamic sensitivity analysis to determine the optimal perturbation strategies to rebalance the PMS distribution to the healthy state (as defined by the PMS space of responders after one-day after treatment). Figure 3 illustrates the dynamic sensitivity analysis, whereby the bifurcation parameter $a$ is used to change the nodal dynamics in terms of its response to added noise, ranging from a more noise-driven regime (the more $a$ is negative) to an oscillatory regime (with larger amplitude the more $a$ is positive). We focused on homological nodal perturbation of the large-scale brain model, meaning that bilateral regions were perturbed equally, resulting in 45 pairs of regions perturbed at gradually varying...
values of $a$. Figure 3A, shows the dynamic sensitivity analysis of driving a transition to the healthy state for models of both responders and non-responders before treatment. Again, an average of the KL divergence between either the perturbed pre-treatment responders or non-responder models and the healthy PMS space was shown. In the noise-driven regime ($a < 0$), a deterioration of the fit was observed for both groups, while in the oscillatory regime ($a > 0$), an initial improvement across all 45 runs was depicted, before subsequent deterioration away from the optimal fit for both groups. Conversely, when replacing the target healthy state by the depressive state (i.e., by comparing with the average PMS in non-responders after treatment), we found that the KL divergence was minimal without perturbation (i.e., keeping $a=0$), showing a worse fit for both groups when brain areas became more oscillatory and no effect of the noisy perturbation (Figure 3B). This analysis demonstrates the optimal response for the amelioration of brain dynamics towards the optimal brain dynamics, as described by the PMS space of post-treatment responders, to be in the oscillation-driven regime ($a>0$) for both responders and non-responders.

To evaluate which regions permitted transition to a healthy state, we first defined the optimal perturbation strength as the minimum of the averaged KL divergence (across the 45 runs) of the responder group to the treatment. This stimulation intensity was found at $a = 0.07$. Then, we inspected the difference between the responders and non-responders at that given value of $a$ to assess what nodal perturbations were permitting the transition to the healthy state in responders but not in non-responders (Figure 4).
Figure 4A shows the rank ordered regional differences in KL divergence between perturbations of the responder and non-responder models before treatment at a stimulation intensity of $a = 0.07$. We highlighted the regions with the largest KL divergence working in responders but not non-responders to promote a transition to the healthy state. These regions are the Temporal Superior pole, Rolandic operculum, Fusiform gyrus, Supplementary Motor Area, Parietal Inferior gyrus, Angular gyrus, Supramarginal gyrus, Frontal Inferior gyrus (opercular), Frontal Middle gyrus (orbital) and the Parahippocampal gyrus. Figure 4B, shows the cortical rendering of these differences. In summary, these regions are implicated in transition away from ‘depressed brain’ pathology and towards the ‘healthy brain’ configurations of treatment responders.

Correlation with Serotonin Receptor Maps

Given the unique and known neuropharmacology of psychedelics whereby psilocybin – the active component of magic mushrooms – binds with high affinity to the serotonergic receptors (mainly the $5-HT_{2a}$) and in effect increases neural activity, we assessed whether the regions working in responders but not non-responders overlapped with the 5-HT density maps derived from PET imaging data previously obtained by an independent research group. Figure 5A and 5B shows the correlation between the $5-HT_{2a}$ and $5-HT_{1a}$ receptor density maps and regional differences in KL divergence between perturbations of the responder and non-responder models before treatment at a stimulation intensity of $a = 0.07$ (same as in Figure 4). (Spearman $\rho = 0.227$, $p = 0.032$ and Spearman $\rho = 0.284$, $p = 0.007$ respectively). Figure 5C, shows non-significant correlations to other 5-HT components – namely the $5-HT_{2b}$ (Spearman $\rho = 0.064$, $p = 0.055$) and $5-HT_{4}$ receptors (Spearman $\rho = 0.055$, $p = 0.607$), and the 5-HT transporter ($5-HTT$) (Spearman $\rho = -0.172$, $p = 0.106$). This analysis shows that the ability to
force transition towards the optimal spatio-temporal dynamics of a given region correlates with density of $5HT_{2a}$ and $5HT_{1a}$ neuroreceptors of that region.

**Discussion**

In this work, we employed a large-scale brain modelling approach to evaluate potential brain-change causes of response to psilocybin therapy for treatment-resistant depression. Using a novel combination of empirical data and in silico modelling, systematic perturbations to brain regions modelled in silico, revealed a subset of regions implicated in transition away from ‘depressed brain’ pathology and towards the ‘healthy brain’ configurations of treatment responders. Notably, these regions matched those with the highest density of $5HT_{2a}$ and $5HT_{1a}$ neuroreceptors. This relationship is plausible given that psilocin (psilocybin’s active metabolite) is known to have an appreciable-to-high affinity for the 5-HT$_{1a}$ and 5-HT$_{2a}$ receptors, respectively, where it acts as an agonist; in the case of the 5-HT$_{2a}$, potentially stimulating plasticity-related signaling cascades relevant to an antidepressant action.$^{61,62}$

Another complementary perspective of the findings, which show the correlation between 5-HT$_{2a}$ and 5-HT$_{1a}$ serotonergic receptors and regions with difference in the intervention between the groups, is that of the overactivation of default-mode network in depression both in terms of its connectivity profile and temporal characterization.$^{22-24,63}$ It has been shown that serotonergic raphe nuclei have direct anatomical projections to the regions of Default Mode Network (DMN) and 5HT$_2$ receptors are expressed in DMN regions,$^{64}$ and it is thus possible
that the psychedelic action enables modulation of the DMN via serotonergic pathways in responders but not non-responders.

A summary of complex spatio-temporal dynamics, in terms of brain substates and their transitions, has drawn a lot of attention in systems neuroscience due to its utility to evaluate the impact of pharmacological and electromagnetic interventions for treating brain and behavioural disorders. Brain substates have been characterised in different ways; by minimal energy\textsuperscript{65} as attractor landscapes,\textsuperscript{16,34} and more heuristically, through sliding-window analysis and unsupervised clustering.\textsuperscript{8,9} However, it has been challenging to find a model that is sufficiently simple and yet accurate to account for temporally and spatially complex and non-stationary datasets. Here, PMS are built on a description of the data in terms of a probabilistic "cloud" in substate space and as such can be extended to different neuroimaging modalities with higher temporal resolution, such as EEG and MEG, or potentially to more fine-grained spatial resolutions\textsuperscript{1,11}

Cutting-edge non-invasive brain stimulation techniques such as Transcranial Magnetic Stimulation (TMS) and Direct Electrical Stimulation (DES), and new neuropsychopharmacological drugs for treatment of psychiatric disorders have heralded a new era of localized and system-wide brain perturbations as medical interventions. For example, TMS has been considered for treatment of many psychiatric disorders such as depression, schizophrenia and addiction,\textsuperscript{66} and classic psychedelic (drug) therapy, which in part, targets a specific neuroreceptor (i.e., principally the 5-HT\textsubscript{1a} receptor) is showing efficacy in the treatment of a broad range of conditions such as depressive, anxiety and addiction disorders.\textsuperscript{38} However, it seems highly likely that the mechanistic action of these interventions lies potentially well downstream of their initial action, and this action may not be straightforward.\textsuperscript{67}
For example, how Direct Electrical Stimulation (DES) induced signal propagates within neuronal microcircuits remains unclear and often paradoxical,\textsuperscript{68} and motivates theoretical neuroscience studies and \textit{in-silico} perturbation protocols. In this perspective, our study attempted to understand how the psychedelic-induced perturbation impacts regional activity that alter the brain-wide dynamics thus potentially bridging the initial (as done experimentally) and downstream actions (as done in-silico) of such intervention.

Beyond \textit{in-silico} perturbations, the exhaustive stimulation protocol can also be used as a dynamic sensitivity analysis tool from the complex systems perspective. Traditionally, statistical differences in measures summarising spatio-temporal dynamics are obtained using signal detection theory. Such approaches can be enhanced by considering large-scale brain models and their structural differences between conditions, for example as described by the global coupling ($G$) parameter. Moreover, rather than describing and assessing expressions of spatio-temporal dynamics, an exhaustive protocol allows a shift of focus onto transitions to a target state and this can be used to identify differences between groups, such as treatment responders versus non-responders, as we have done here.

Forcing transitions in large-scale brain networks has also been investigated through the prism of control network theory. In such scenarios, control strategies are deployed to navigate complex systems from a source (initial) state to a target (final) state.\textsuperscript{69} This approach has obtained a lot of attention due to its wide-ranging engineering applicability in technological, social and cyberphysical systems across various experimental scenarios.\textsuperscript{70,71} However, the conceptual understanding of controlling neuronal signals from source to target might be problematic as the brain operates in self-sustained and non-equilibrium state, and the notion of well-defined pathway between them might be ill-posed.\textsuperscript{72} On the contrary, the approach
considered in this work describes spatio-temporal dynamics in terms of Probabilistic Metastable Substates and, through systematic perturbation, rebalances the spatio-temporal dynamics between two PMS spaces. Through this approach the brain is rebalanced to its healthy working point, without specific instructions of what the relevant pathway might be.

To obtain a PMS approximation of the brain substate of interest, several methodological choices are made which inevitably introduce several caveats. Firstly, a regional parcellation must be chosen, which might introduce artificial spatial boundaries especially when dealing with dynamics. Secondly, the choice of clustering algorithm defines the type of substates that can be obtained. Here, we use the unsupervised learning algorithm \textit{k-means} clustering which has been shown to adequately represent functionally meaningful brain substates.\(^{16}\) However, alternative algorithms could be used for this purpose (e.g. k-medoids). Related to the experimental data, the design is an uncontrolled open-label feasibility pilot study, and as such has no placebo group and suffers from small sample size. In this context, it is relevant to consider the current study as exploratory. Hence, future replication studies are warranted to ensure robustness of the findings. Moreover, the healthy state is defined here in terms of the 1-day post-treatment scan but the responders/non-responders’ assessment is done 5 weeks after. Lastly, the large-scale brain models constructed are based on group approximations of the functional brain information and structural connectivity group template. For clinical relevance, further research will be needed to create individual-based large-scale brain models that might allow for future in-silico assisted personalised psychiatry.\(^{73}\)

\textbf{Data Availability Statement}

Raw data were generated at Imperial College London. Derived data supporting the findings of this
study are available from the corresponding author upon request. The code for the analysis is

Funding

Jakub Vohryzek is supported by EU H2020 Future and Emerging Technologies (FET)
Proactive project Neurotwin grant agreement no. 101017716, Morten L. Kringelbach is
supported by the European Research Council Consolidator Grant: CAREGIVING (615539),
Pettit Foundation, Carlsberg Foundation and Center for Music in the Brain, funded by the
Danish National Research Foundation (DNRF117). Joana Cabral is supported by Portuguese
Foundation for Science and Technology CEECIND/03325/2017, UIDB/50026/2020 and
UIDP/50026/2020, Portugal. Gustavo Deco 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] SGA1) and 785907
(HBP SGA2); and by the Catalan Agency for Management of University and Research Grants
Programme 2017 SGR 1545.

Competing Interest

Robin Carhart-Harris reports receiving consulting fees from Beckley Psytech. The other
authors have declared that no competing interests exist.

References


Figure 1. Study Overview. A) Experimental Analysis. Probabilistic Metastable Substates were computed for each condition using leading eigenvector dynamics analysis (LEiDA). Regional fMRI timeseries were first converted to analytical signal, followed by computation of the leading eigenvector of the phase coherence matrix at every timepoint. An unsupervised k-means algorithm was deployed to cluster the eigenvectors into a three substate solution. The PMS is defined as the probability distribution of substates, obtained for each individual scan and averaged within each condition (Substate 1 = S1, Substate 2 = S2, Substate 3 = S3). B) Model Fitting. Large-scale brain model parameters were optimised to fit the PMS before treatment separately for responders and non-responders. C) Dynamic Sensitivity Analysis. In silico bilateral perturbations were performed to find the optimal protocol to transition to the PMS characteristic of a healthy brain state (described by responders’ (as predicted by the 5-week QIDS endpoint) one-day post-treatment brains). D) Dynamic Sensitivity Evaluation. Perturbations are applied separately in each pair of bilateral brain regions by varying the intensity of oscillations as defined by the bifurcation parameter a.

Figure 2. A) Experimental Analysis. Probability of occurrence (or Fractional Occupancy) of each metastable substate detected with LEiDA for the three-substate clustering solution. Significant differences were observed in the substate 3 between responders before and after treatment (p = 0.0258, signed rank-sum permutation test, green star for significance, number of responders 6 and number of non-responders 9, individual subjects plotted for the experimental analysis A), responders and non-responders after treatment (p = 0.0141, rank-sum permutation test, green star for significance, number of responders 6 and number of non-responders 9, individual subjects plotted for the experimental analysis A), and no significant differences were found between responders and non-responders before treatment. B) Model Fitting of the responder and non-responder models as a function of the global coupling parameter G, with optimal fits at G = 0.185 (Kullback-Liebler divergence = 0.0064) and G = 0.165 (KL-divergence = 0.0187) respectively. We report the median across subjects in each group with interquartile range (IQL). C) Experimental and Simulated PMS. Experimental PMS for responders and non-responders before treatment (left), their simulated counterparts at optimal G (middle), and experimental PMS for responders and non-responders after treatment (right).

Figure 3. Evaluation of Dynamic Sensitivity Analysis. A) Perturbation to induce a transition to a healthy state. Each homological pair of brain regions was perturbed by varying the bifurcation parameter a, which modulates the intrinsic oscillatory behavior of the dynamical units. The more a is positive, the larger the amplitude...
of intrinsic oscillations, whereas for negative \( a \) the units decay to a fixed point equilibrium and the local dynamics is dominated by noise. The performance of the perturbations is evaluated by computing the KL divergence between the simulated PMS and the empirical PMS from patients who recovered after treatment with psilocybin. Optimal intensity of \( a = 0.07 \) was achieved for the responder group (red rectangles). **B) Perturbation to induce a transition to a depressive state.** A transition to the depressive state showed worse or no effect at varying values of the bifurcation parameter \( a \). This is expected since the models were adjusted to patients in the depressive state before treatment.

**Figure 4. Subset of regions working in responders but not in non-responders.** A) Rank ordered absolute difference of KL divergence between perturbations of the responder and non-responder models before treatment at a stimulation intensity of \( a = 0.07 \). Inset brain rendering of the ten brain regions with the highest difference: Temporal Superior pole, Rolandic operculum, Fusiform gyrus, Supplementary Motor Area, Parietal Inferior gyrus, Angular gyrus, Supramarginal gyrus, Frontal Inferior gyrus (opercular), Frontal Middle gyrus (orbital) and the Parahippocampal gyrus. B) Cortical rendering and flat maps showing the distribution of all Kullback-Leibler (KL) divergence differences.

**Figure 5 Ability to promote a transition relates to density of specific serotonin receptors:** For each pair of homologous brain regions, the ability to promote a transition is plotted against the receptor map densities of: A) 5-HT\(_{2a} \) (Spearman correlation \( \rho = 0.227, p = 0.032 \)). B) 5HT\(_{1a} \) (\( \rho = 0.284, p = 0.007 \)). C) In contrast, the other 5-HT receptors with non-significant results: 5-HT\(_{2b} \) (\( \rho = 0.064, p = 0.055 \)), 5-HT\(_{4} \) (\( \rho = 0.055, p = 0.607 \)), as well as the serotonin transporter the 5-HTT (\( \rho = -0.172, p = 0.106 \)).
A Experimental Analysis

- Substate 1
- Substate 2
- Substate 3

RB - Responders Before, RA - Responders After
NB - Non-Responders Before, NA - Non-Responders After

* p-val < 0.05

B Model Fitting

- Probabilistic Metastable Substate (PMS)

- KL Divergence vs. Global Coupling (G)

C Experimental and Simulated PMS

- Experimental Responders Before
  - G = 0.185, KL-div = 0.0064

- Experimental Responders After

- Experimental
  - G = 0.165, KL-div = 0.0064

- Simulated
  - G = 0.185, KL-div = 0.0064
A Perturbation **healthy** state

Comparison: Responders to **healthy** state

- KL Divergence
  - Range: 0.05 to 0.2
  - Best values highlighted

Comparison: Non-Responders to **healthy** state

- KL Divergence
  - Range: 0.05 to 0.2
  - Best values highlighted

B Perturbation to **depressive** state

Comparison: Responders to **depressive** state

- KL Divergence
  - Range: 0.05 to 0.2
  - Best values highlighted

Comparison: Non-Responders to **depressive** state

- KL Divergence
  - Range: 0.05 to 0.2
  - Best values highlighted
Subset of regions working in responders but not non-responders

1. Superior Temporal Pole
2. Rolandic operculum
3. Fusiform gyrus
4. Supplementary Motor Area
5. Parietal Inferior gyrus
6. Angular gyrus
7. Supramarginal gyrus
8. Frontal Inferior gyrus (opercular)
9. Frontal Middle gyrus (orbital)
10. Parahippocampal gyrus
A  

**5HT$_{2a}$**

\[ \rho = 0.227, \quad p = 0.032 \]

Ability to force transition
(smaller = more able)

---

B  

**5HT$_{1a}$**

\[ \rho = 0.284, \quad p = 0.007 \]

Ability to force transition
(smaller = more able)

---

C  

**5HT$_{1b}$**

\[ \rho = 0.064, \quad p = 0.550 \]

**5HT$_4$**

\[ \rho = 0.055, \quad p = 0.607 \]

**5HTT**

\[ \rho = -0.172, \quad p = 0.106 \]

Ability to force transition
(smaller = more able)
Whole-brain modelling for predicting psilocybin response for treatment of depression

A Experimental Analysis

1. Parcellation
2. Functional MRI Phases
3. Imaginary axis
4. Real axis
5. Time (s)
6. Band-passed fMRI signal
7. Imaginary
8. Real
9. Brain Areas
10. Probability Metastable Substates

B Model Fitting

1. Experimental Responders
2. Simulated Responders
3. Stuart-Landau oscillators
4. S1, S2, S3
5. Experimental
6. Simulated

C Dynamic Sensitivity Analysis

1. Perturbed States
2. Healthy State
3. Comparison: Responders to Healthy State
4. Comparison: Non-Responders to Healthy State
5. KL Divergence
6. Stimulation Intensity
7. Brain Regions
8. S1, S2, S3
9. Before
10. After

D Dynamic Sensitivity Evaluation

1. Hopf bifurcation
2. y [Im(z)]
3. x [Re(z)]
4. Comparison: Responders to Healthy State
5. Comparison: Non-Responders to Healthy State
6. KL Divergence
7. Best

https://mc.manuscriptcentral.com/braincom