communications biology

Article

https://doi.org/10.1038/s42003-024-06852-9

Whole brain modelling for simulating pharmacological interventions on patients with disorders of consciousness

I. Mindlin[1] ✉, R. Herzog[1], L. Belloli[1,2], D. Manasova[1,3], M. Monge-Asensio[4], J. Vohryzek[4], A. Escrichs[4], N. Alnagger[5,6], P. Núñez[5,6], O. Gosseries[5,6], M. L. Kringelbach[7,8,9], G. Deco[4,10], E. Tagliazucchi[2,11,12], L. Naccache[1], B. Rohaut[1,13], J. D. Sitt[1,14] ✉ & Y. Sanz Perl[1,2,4,14] ✉

Disorders of consciousness (DoC) represent a challenging and complex group of neurological conditions characterised by profound disturbances in consciousness. The current range of treatments for DoC is limited. This has sparked growing interest in developing new treatments, including the use of psychedelic drugs. Nevertheless, clinical investigations and the mechanisms behind them are methodologically and ethically constrained. To tackle these limitations, we combined biologically plausible whole-brain models with deep learning techniques to characterise the low-dimensional space of DoC patients. We investigated the effects of model pharmacological interventions by including the whole-brain dynamical consequences of the enhanced neuromodulatory level of different neurotransmitters, and providing geometrical interpretation in the low-dimensional space. Our findings show that serotonergic and opioid receptors effectively shifted the DoC models towards a dynamical behaviour associated with a healthier state, and that these improvements correlated with the mean density of the activated receptors throughout the brain. These findings mark an important step towards the development of treatments not only for DoC but also for a broader spectrum of brain diseases. Our method offers a promising avenue for exploring the therapeutic potential of pharmacological interventions within the ethical and methodological confines of clinical research.

Despite the lack of broad definitions of consciousness as a global brain state, it is accepted that it can be lost or diminished when we fall asleep, or as the effect of drugs, such as those employed to induce general anaesthesia. Consciousness is also impaired in pathological conditions, such as during coma or in post-comatose disorders of consciousness (DoC). The two primary categories within the spectrum of DoC are unresponsive wakefulness syndrome (UWS) which is also coined vegetative state (VS), and minimally conscious state (MCS). UWS is characterised by preserved arousal without any behavioural evidence of consciousness, while MCS represents a condition with richer cognitive processes that are not limited to reflexive processes. Observation of MCS patients demonstrates that—in contrast with UWS/VS patients—cortical networks still play a role in their overt behaviour[1,2]. DoC are complex neurological diseases with heterogeneous etiologies and presentations, often resulting in a substantial misdiagnosis rate with potentially devastating implications[3].

In the last two decades, the incorporation of neuroimaging techniques resulted in significant improvements regarding the diagnosis and prognosis of DoC patients[4]. Data-driven machine learning approaches are capable of identifying brain activity patterns that predict different levels of impaired consciousness, with an accuracy comparable to that of expert

[1]Institut du Cerveau - Paris Brain Institute - ICM, Inserm, CNRS, Sorbonne Université, Paris, 75013, France. [2]Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Ministry of Science, Technology and Innovation, Buenos Aires, Argentina. [3]Université Paris Cité, Paris, France. [4]Center for Brain and Cognition, Computational Neuroscience Group, Universitat Pompeu Fabra, Barcelona, Spain. [5]Coma Science Group, GIGA Consciousness, University of Liège, Liège, Belgium. [6]Centre du cerveau, CHU of Liège, Liège, Belgium. [7]Centre for Eudaimonia and Human Flourishing, Linacre College, University of Oxford, Oxford, UK. [8]Department of Psychiatry, University of Oxford, Oxford, UK. [9]Center for Music in the Brain, Department of Clinical Medicine, Aarhus University, Aarhus, Denmark. [10]Institució Catalana de la Recerca i Estudis Avançats (ICREA), Barcelona, Spain. [11]Latin American Brain Health Institute (BrainLat), Universidad Adolfo Ibáñez, Santiago, Chile. [12]Cognitive Neuroscience Center, Universidad de San Andrés, Buenos Aires, Argentina. [13]APHP, Hôpital de la Pitié Salpêtrière, DMU Neurosciences, Neuro ICU, Sorbonne Université, Paris, France. [14]These authors jointly supervised this work: J. D. Sitt, Y. Sanz Perl.
✉e-mail: ivan.mindlin@icm-institute.org; jacobo.sitt@icm-institute.org; yonatan.sanz@upf.edu

neurologists[5–7]. One downside of these approaches is that they are not based on generative models, therefore they cannot be used to test causal mechanisms underlying the observed differences between groups[8]. In turn, the insufficient understanding of how consciousness is lost in brain-injured patients hinders the development of novel therapeutic interventions[9].

Currently, the range of treatments for DoC includes Zolpidem and Amantadine as pharmacological options[10], as well as invasive[11] and non-invasive electrical magnetic stimulation methods, such as transcranial magnetic stimulation (TMS)[12]. The pharmacodynamics of the pharmacological interventions are heterogeneous. Zolpidem is a sedative-hypnotic that interestingly has a selective agonism to the GABA receptor subtype ω1 which is strongly present in basal ganglia and striatum provoking an inhibition of these inhibitory areas. It restores consciousness in chronic patients due to this paradoxical effect[13], and its effect wears off along with the drug[14]. In contrast, Amantadine is a N-methyl-D-aspartate receptor antagonist and indirect dopamine agonist capable of inducing both short (5-10 days) and long-term (4-6 weeks) improvements[15], and it was incorporated in DoC practical guidelines for clinicians[16]. Unfortunately, available drugs have modest therapeutic efficacy and act via unknown specific mechanisms, thus hindering the development of new options, while clinical trials are difficult to organise due to a low number of participants and other limitations intrinsic to patients with conscious impairments[9].

In recent years, serotonergic psychedelics have been proposed as a potential new treatment to accelerate the recovery of DoC patients[17]. This proposal was not based on pharmacological considerations, but on the measured effects of psychedelics on brain activity, and the interpretation assigned to these changes by theories of consciousness[18]. Psychedelic drugs are suggestive to increase brain complexity measured using either Shannon Entropy or Lempel-Ziv complexity[19], both metrics known to decrease during episodes of diminished awareness (deep sleep, general anaesthesia, and DoC)[20]. At the same time, over the past years the use of psychedelics for treating mental health disorders has shown promising results in conditions such as depression[21–24] and obsessive-compulsive disorder[25]. It is speculated that psychedelics can result in functional rearrangements mediated by serotonin 2 A (5-HT2A) receptor agonism, which modulates the emergent whole-brain dynamics, generating principal psychoactive effects of these drugs[26]. Importantly, the density of this receptor is maximal in key sub-cortical and high-level cortical association areas implicated with conscious information processing[27]. Taken as a whole, these findings support the proposal by Scott and Carhart-Harris of exploring classic psychedelics as new avenues for the treatment of DoC.

The current consensus is that psychedelic drugs are very safe substances when consumed by healthy individuals under adequate conditions[28]. However, a deeper analysis reveals distinct ethical concerns for the use of psychedelics in non-communicative patients[29]. Computational modelling offers an alternative to evaluate the therapeutic action of serotonergic psychedelics without facing the ethical challenges of human in vivo experimentation. While this research cannot replace proper clinical trials, it can be used to support the feasibility of this intervention, adding to evidence from other sources, such as animal experimentation.

Here we proposed a framework combining deep learning and computational whole-brain modelling to provide mechanistic understanding of possible pharmacological treatments for DoC, including serotonergic treatments, in a controlled and ethical manner. We used a biophysically-grounded model based on coupled mean fields representing the dynamics of brain regions at the macroscale, composed by excitatory and inhibitory neural populations[30]. The strength of connection between nodes is informed through anatomical structural connectome informed by diffusion MRI (dMRI). Using this Dynamic Mean Field (DMF) model, we simulated agonism at multiple receptors by altering the nonlinear response of neuronal populations to synaptic input, weighted by the local density of normalised receptors informed by Positron Emission Tomography (PET) data. This integrative multimodal approach in computational modelling has already demonstrated its effectiveness in capturing the intricate interplay between coupled whole-brain networks and neurotransmitter systems[31]. To assess the effects of simulated interventions, we used a variational autoencoder (VAE) trained on the output of the DMF model fitted to the pathological and healthy control states. The VAE learns the defining attributes of the data in an unsupervised manner and projects it into a low-dimensional latent space, where different brain states are clustered in a continuous and organised way[32–34]. The perturbations can then be interpreted as trajectories within this space, which we characterise using well-established metrics sensitive to the level of conscious awareness.

Results

Methodological overview

The procedure followed in this work is represented in Fig. 1. First, we implemented a biologically realistic DMF model[30] fitted to fMRI data from DoC patients and healthy controls. The DMF models brain dynamics as the emergent activity of local recurrent excitatory/inhibitory connections, with long-range excitatory connections between cortical regions, scaled by the global connectivity parameter (G) and the structural connectivity. The structural connectivity (SC) is determined by the amount of white fibre passing through each pair of regions and was obtained from an independent cohort of healthy subjects with high-quality DTI data from the HCP dataset. The SC used in this work is obtained by averaging the individual SC from this dataset. Local inhibition is controlled by the Feedback Inhibitory Current (FIC), which offsets inter-areal excitation and clamps the average excitatory firing rate around 3 Hz. The G parameter, and thus indirectly the FIC, are tuned to maximise the similarity to the empirical brain functional connectivity (FC), as computed from the fMRI Blood-Oxygen-Level-Dependent (BOLD) signals. Finally, the expression of neurotransmitter receptors on each region is obtained from PET maps, and the susceptibility of each region to the activation of these receptors is controlled via a global scaling parameter. Since every region has its own level of receptor expression, this global scaling parameter will introduce regionally heterogeneous neuromodulation. In past studies, this procedure was successfully used to simulate different states of consciousness, such as propofol-induced anaesthesia[35] and the psychedelic state elicited by LSD[36].

After tuning the model, we used the inferred parameters for each condition (G and FIC) to generate surrogate FC matrices as a data synthesis procedure[37] to train a variational autoencoder (VAE), resulting in a two-dimensional encoding of each FC matrix. We then characterised this low-dimensional space in terms of different metrics sensitive to changes in consciousness and explored the effects of pharmacological interventions. As a result, changes in receptor scalings yield the potential to reflect global shifts in neurotransmission signalling and can explain whole-brain dynamics in response to a drug intervention.

We implemented this procedure with 8 different receptors (serotonergic 5HT2A, 5HT1A, 5HT6; dopaminergic D1, D2; dopamine transporter (DAT); opiate μ, Histamine H3). We selected serotonergic receptors to test the effect of psychedelic drugs, whose main target is the 5HT2A receptor[28], while the choice of dopaminergic receptors stemmed from their relevance to current treatments for DoC (e.g., Amantadine). Including opiate and histamine receptors completed the spectrum of potential stimulation targets, resulting in a comprehensive array of diverse receptors for analysis. These receptors were normalised (see Supplementary Fig. 1) to account for their overall presence in the brain, independent of the absolute PET values.

As a final step, we encoded the pharmacological intervention in the two-dimensional latent space, parametrised by the scaling that modulates the receptor density maps. This combined framework allowed us to represent the interventions as trajectories in the latent space and compute geometrical properties that characterised the intervention, such as distance to the healthy controls, the distance from the original non-perturbed model, as well as changes in brain complexity, network integration, and correlation with the anatomical connectivity.

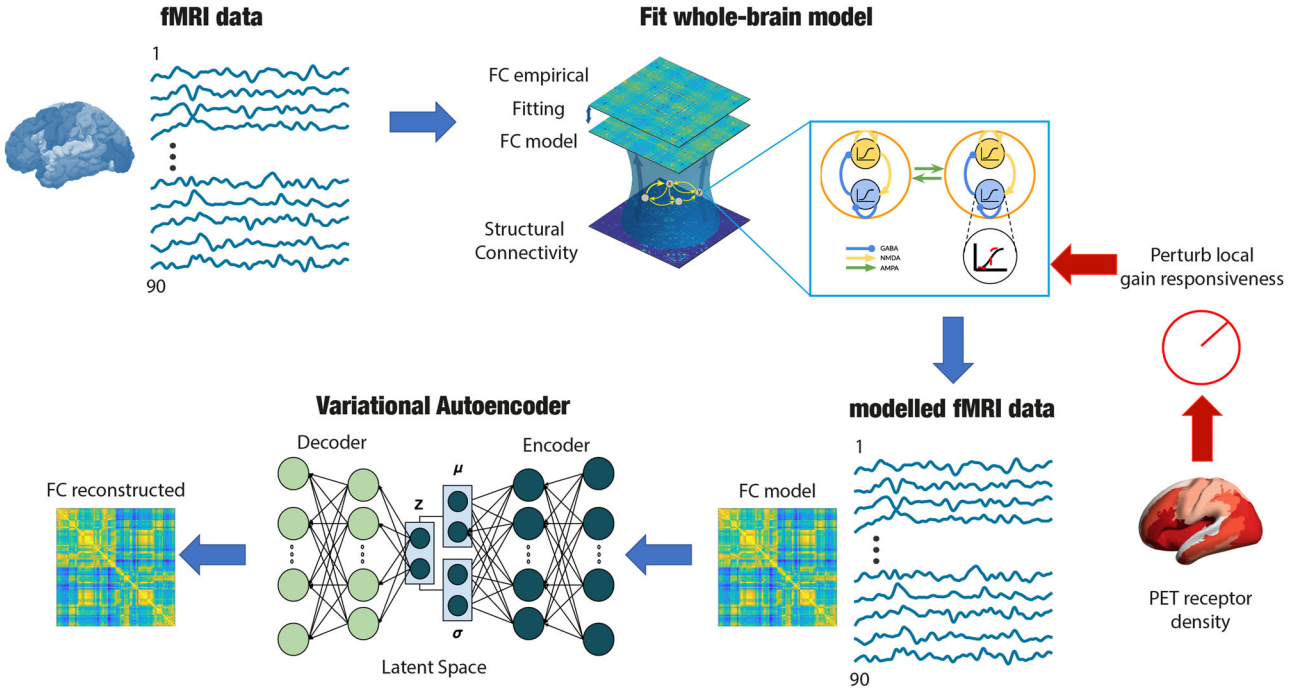

Fig. 1 | Overview of the whole brain model, variational autoencoder and perturbation pipeline. A biophysical model simulates the local dynamics of a 90 region parcellation, with structural connectivity provided by DTI data of healthy subjects. Model parameters are adjusted to fit FC matrices derived from fMRI data of each condition. Pharmacological perturbations are simulated by changing the synaptic scaling of each region weighted by PET receptor maps, which provide a measure of local receptor density. Subsequently, a variational autoencoder (VAE) is trained to reconstruct the input data, resulting in a low-dimensional latent space which facilitates the visualisation of the perturbations, while also providing a geometric interpretation of their effect.

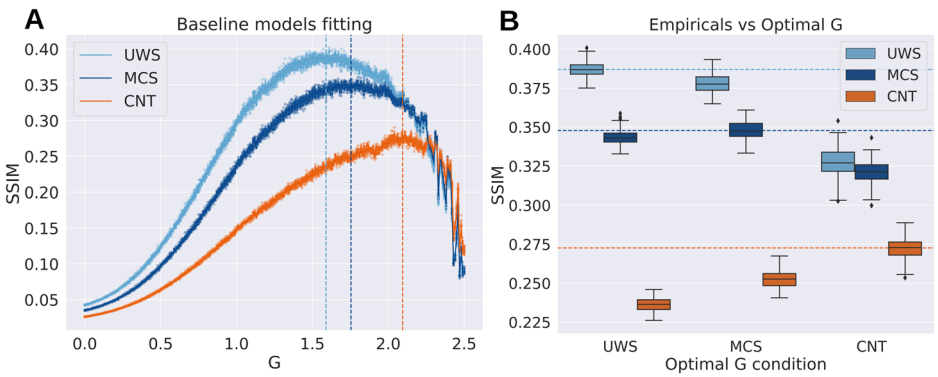

Fig. 2 | Model fitting and validation of the baseline models. A Model fit curves where the parameter G is optimised to maximise the SSIM to the empirical FC. The optimal values of G for each condition (mean + -std: GUWS = 1.59 + −0.04; GMCS = 1.75 + −0.07; GCNT = 2.1 + −0.06) align with existing literature, where lower *G* is associated with reduced consciousness. Error bars show standard error for each G value. **B** The optimal model for each condition exhibits significantly better SSIM values to its corresponding empirical dataset compared to the rest, as shown in the dashed line for each condition.

Baseline model fitting

To explore the effect of perturbing via different receptors, we first created a *baseline* model for healthy controls (CNT) and each patient group, namely UWS and MCS. For each value of G, we conducted repeated simulations for the number of subjects in each group and then also repeated the training process 20 times to optimise the SSIM between simulation and our target data at group level. The SSIM is a similarity measure that balances correlation and euclidean distance between the inputs[38], in this case the simulated and the empirical FC. The range of G values were defined such that the model always remained in a biologically plausible regime, below hyper-excited and hyper-correlated activity. In Fig. 2A, the plots demonstrate the fitting performances for each group. In all three groups, a point of instability is reached, resulting in the breakdown of the goodness of fit. Prior to this instance, all the models attain their best fit. We repeated this exploration of G values 15 times. Then, we calculated the optimal value of G for each condition by taking the mean optimal G obtained across iterations (mean + -std: $G_{\mathrm{UWS}} = 1.59 + -0.04$; $G_{MCS} = 1.75 + -0.07$; $G_{CNT} = 2.1 + -0.06$). The low variability in the results across iterations highlights the robustness of our models. These G values determine how we model each condition with the DMF at the baseline.

As a next step, we assessed if these models were distinguishable between levels of consciousness. To this end, we generated 3000 simulations and calculated the SSIM between their FC and the empirical FC of each condition. We compared the SSIM values obtained between simulations and empirical FCs for each condition and measured the effect size using Cohen's d. The models simulations were significantly different across states (Empirical UWS/Simulations CNTvsMCS Cohen's d = 23.45 MCSvsUWS Cohen's d = 8.75 CNTvsUWS Cohen's d = 31.96 Empirical MCS/CNTvsMCS Cohen's d = 17.24 CNTvsUWS Cohen's d = 22.60 MCSvsUWS Cohen's d = −5.39 Empirical CNT/MCSvsUWS Cohen's d = −0.68 CNTvsMCS Cohen's d = 7.20 CNTvsUWS Cohen's d = 6.87). Additionally, each model had the highest similarity with the condition they were fitted to ($SSIM_{uws} = 0.38$, $SSIM_{MCS} = 0.34$, $SSIM_{CNT} = 0.27$), as shown in Fig. 2B. As the *G* parameter was decreased, regions did not interact between themselves,

reducing integration of the network and stabilising the overall dynamics to those determined by the anatomical structure. This situation was seen for unconscious brain states, whereas the opposite could be seen for conscious controls. Therefore the resulting optimal values of G found for each condition are consistent with this observation ($G_{UWS} < G_{MCS} < G_{CNT}$).

Latent space embedding of baseline models

After fitting the baseline models, we employed 3000 simulations from each condition to train a variational autoencoder (VAE) for reconstructing the input data, following previous work[33]. The VAE is a type of generative model that combines the strengths of traditional autoencoders with the probabilistic framework of variational inference. The variational component allows the model to learn the statistical distribution of the data in the latent space rather than just mapping the data to a fixed point. By learning the statistical distribution of the input data, the VAE constructs a continuous and smooth latent space[32], called *manifold*, that captures the underlying structure of the simulated FC. This results in a meaningful representation of the data, where similar inputs are mapped closer together and dissimilar inputs are mapped farther apart in the latent space. The use of the 2-dimensional latent space allowed us to visualise a *perturbational landscape*, summarising the impact of all possible pharmacological perturbations applied to the DoC models.

Once the VAE was trained, we encoded a test set to observe how the different conditions organised in the latent space, as depicted in Fig. 3A. The different conditions formed separate clusters with very small overlap, which means that conditions were distinguishable within the latent space embedding. We quantified this by training a Support Vector Machine that successfully classified the projected clusters (See Supplementary Table 1). Additionally, there appeared an ordering along an axis starting from lower consciousness (UWS), passing through an intermediate stage (MCS), and finally to the fully conscious state (CNT). Fig. 3C–E. The highlighted area shows the value at the centroid of each condition. As seen in Fig. 3C, the embedded FC matrices from the control group appear in regions that exhibit higher Shannon Entropy (H) than the corresponding DoC regions (Median, IQR HUWS = 4.04, 0.1, HMCS = 4.17,0.07, HCNT = 4.32, 0.04 UWSvsMCS Cohen's d = −1.99 MCSvsCNT Cohen's d = −2.80 UWSvsCNT Cohen's d = −4.13). We calculated the H value on the node strength vector of the FC matrix, following the methodology established in previous studies[39]. The mean value of FC (mFC) matrices has been related to the depthness of unconsciousness[33] due to the hierarchical decline of this value as awareness is lost (Median, IQR mFC_{UWS} = 0.21, 0.03, mFC_{MCS} = 0.25, 0.03, mFC_{CNT} = 0.32, 0.03 UWSvsMCS Cohen's d = −1.81 MCSvsCNT Cohen's d = −2.82 UWSvsCNT Cohen's d = −4.76). As shown by previous research, loss of consciousness abolishes functional interactions that are not directly mediated by structural links, thus making FC patterns of DoC patients more similar to the underlying structural connectivity (SC) pattern[40,41]. Fig. 3E shows these values for our three conditions (Median, IQR $SC\text{-}CORR_{UWS}$ = 0.70, 0.02, $SC\text{-}CORR_{MCS}$ = 0.67, 0.01, $SC\text{-}CORR_{CNT}$ = 0.64,0.015 UWSvsMCS Cohen's d = 2.08 MCSvsCNT Cohen's d = 3.47 UWSvsCNT Cohen's d = 5.25).

Pharmacological perturbations produce transitions away from baseline dynamics

Once we obtained an interpretable latent space of our baseline simulations, we simulated pharmacological perturbations of the DoC models using the normalised PET density maps of different receptors. As mentioned before, these maps give a density value for each node that is then scaled by the *scaling* parameter (see Methods). The perturbation consists of increasing this value in a range between 0 (no receptor intervention) to 1 with a step of 0.01. Embeddings of the perturbations applied to MCS and UWS groups are shown in Fig. 4A, B, respectively. For each receptor map we embedded in the latent space the simulated FC after applying the scaling value in a given step. The resulting trajectory comes from encoding the simulated FC matrices for all steps in the scaling range. Different receptor trajectories reach different distances and the proximity to the control centroid is limited by its starting point (Fig. 4C, D). Another interesting observation is that all the trajectories seem to move approximately in the same direction, i.e., in parallel to the line that unites the three groups from states of more diminished consciousness towards the group of wakeful conscious controls. Despite lacking data on subjects undergoing pharmacological interventions, we simulated 5HT2A activation in healthy individuals, observing a slight deviation from the "wakeful" axis (Supplementary Fig. 2) when introducing this perturbation. This suggests that psychedelics may exacerbate some of the defining characteristics of the conscious wakeful state.

These results show that despite exhibiting different anatomical heterogeneities, the perturbed dynamics exist only in certain regions in latent space, while others are forbidden, thus narrowing the repertoire of possible transitions. The main difference between the trajectories is the distance from the end point to the starting point (Longest and shortest distance achieved for each condition: $Max\text{-}MCS_{5HT1A}$ = 3.49, $Max\text{-}UWS_{5HT2A}$ = 4.61, $Min\text{-}MCS_{D2}$ = 1.62, $Min\text{-}UWS_{D2}$ = 0.74).

Relationship between distribution of receptors and perturbation trajectories

We investigated the relationship between the final distance reached by the trajectory and two attributes of the receptor maps: the mean density of the normalised map and a structural divergence (SD) metric (see Methods). Briefly, the SD is the euclidean distance between the density map and the node strength, which represents the average structural connection strength from each node to the rest. Since we are summing the overall difference between density and connectivity in each region, a low SD will imply that the density map and the node strength are similar (small differences), suggesting that the receptor has high density values in strongly connected nodes, or hubs, and vice versa.

We found a strong correlation between the mean density of the normalised map and the final distance of the trajectory (Fig. 5A, B, MCS Pearson correlation: 0.89, $p = 0.003$; UWS Pearson correlation: 0.96, $p = 0.0001$). The difference in correlation values can be attributed to the curvature of the trajectory after crossing the CNT region in MCS perturbations, which leads to changes in linear distance from the starting point. Furthermore, Fig. 5C, D demonstrates that SD is not correlated with the length of the trajectory. To reaffirm this result, we generated a randomised version of a receptor map that preserved spatial autocorrelation[42] and the mean density. We then perturbed our patient models with this null version of the map. In Supplementary Fig. 3 we can see that the distances from the starting point are indistinguishable. The combination of these results suggests that the neuromodulation of a given receptor displaces the dynamics of a baseline DoC model independently of the topology of the stimulated nodes. Instead, it primarily depends on the receptor presence throughout the brain.

Additionally, we assessed how intensely each receptor should be stimulated to shift brain dynamics. Different drugs may require different relative dosages to alter brain dynamics depending on their pharmacokinetic and pharmacodynamic properties. Ideally, good candidates for treatments should exert their action with small doses to avoid undesired side effects due to the stimulation of off-target receptors. For this purpose, we estimated how quickly the perturbation could displace the model towards CNT-like dynamics. Trajectories may reach similar distances, yet this can occur at different speeds. In mathematical terms we are interested in quantifying the non-linearity of the increase in travelled distance. We calculated the dot product between each projected simulation and an axis drawn between the starting point of the models and the CNT centroid (Fig. 5E, F). To quantify the non-linearity we applied a quadratic regression fit to the cumulative distance along this axis as the scaling increases. If the quadratic coefficient is null, the change is constant. If it is negative the change will be large at first but then decrease as the scaling increases, suggesting a larger effect for a small scaling value. We label "perturbational efficiency" (PE) the negative quadratic coefficient of the best fit, which represents the aforementioned rate. We observe that 5HT2A has the highest PE, even though it has a similar perturbational distance than the other serotonin receptor subtype, 5HT1A.

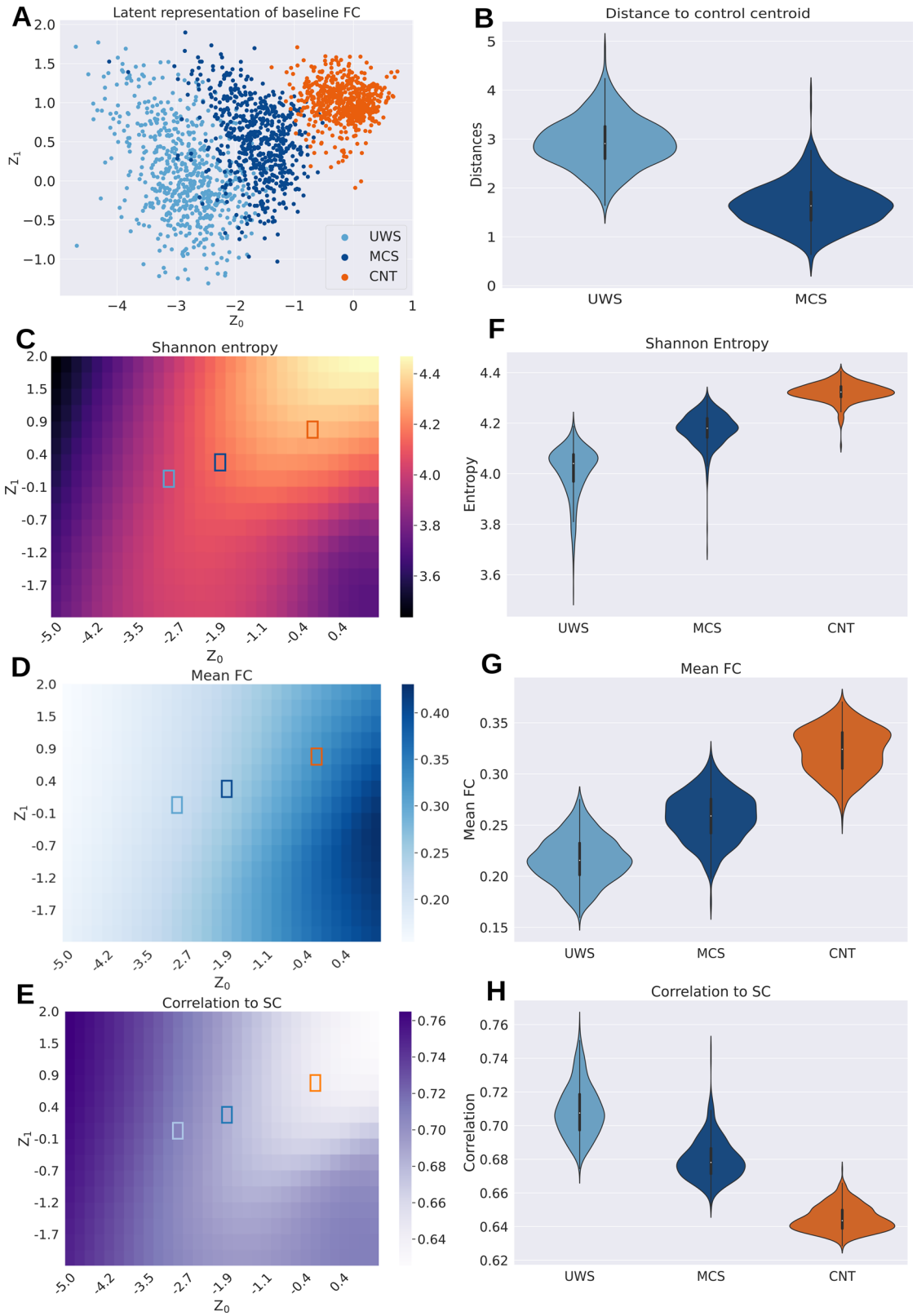

Fig. 3 | Groups are distinguished and characterised by their latent space embedding in an ordered manner. A Scatter plot of FC matrices obtained from the three baseline models for the three groups of participants. The groups are clearly separated within the embedded regions. Embedded FC matrices from Disorders of Consciousness (DoC) conditions are distinguished by (**B**) their distance from the centroid of the FC matrices from the healthy control (CNT) condition, which is the largest for unresponsive wakefulness syndrome (UWS) and smallest for minimally conscious state (MCS). We sampled FC matrices from a grid to calculate different metrics over the latent space: (**C**) Entropy, (**D**) mean FC value and (**E**) correlation to structural connectivity. Each square corresponds to a point from that grid and has therefore an associated value. Highlighted squares indicate the position of the centroids of each group in this space. The values of these metrics in the regions corresponding to each condition are consistent with their interpretation concerning consciousness. **F**–**H** Distribution of the previous measures when sampling the embedded points from (**A**).

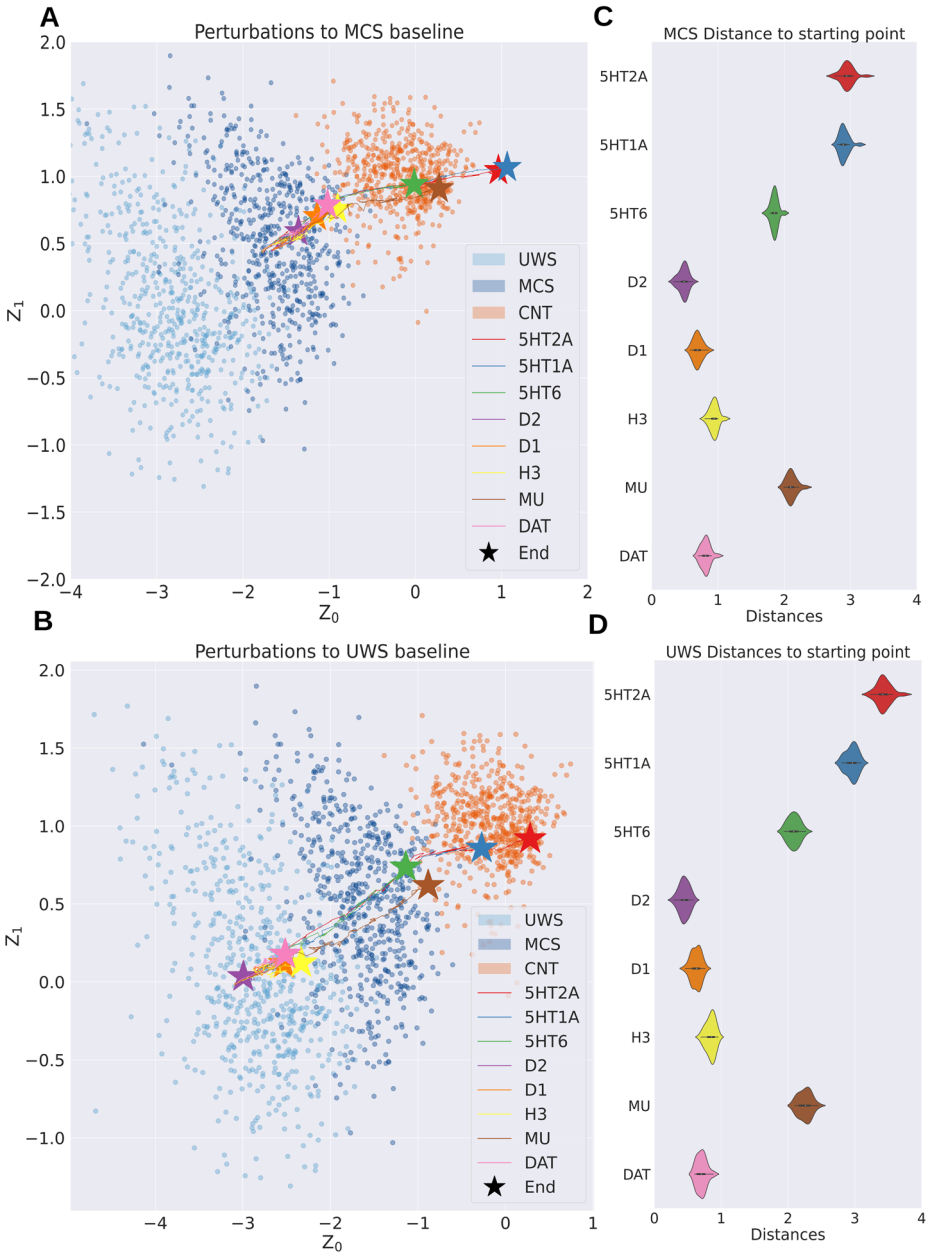

Fig. 4 | Hierarchy of target neurotransmitter receptors based on how their simulated activation displaces whole-brain dynamics towards conscious wakefulness. Panels (**A**) and (**B**) display embedded trajectories for perturbing the MCS and UWS baseline models, respectively, at the different receptors. All perturbations displace the embedded FC matrices along the same axis in latent space, indicating significant variations in the final distance covered across the entire perturbation range. This is presented in panels (**C**) and (**D**), where the covered distances are similar, with differences in the maximum distance reached. Specifically, for UWS, the perturbations did not deviate as far from their starting point, as was observed for the MCS models. Simulated perturbations at the serotonergic receptors and the MU receptor presented the most significant impact on the models.

Discussion

DoC encompasses a set of devastating conditions, affecting not only the lives of the patients themselves but also causing emotional distress among their families and caretakers. The diagnosis of DoC patients is very challenging, also resulting in a difficult evaluation of potential treatment options. This and other issues, such as heterogeneous etiologies and small available populations, hinder the systematic exploration of new pharmacological alternatives. Currently used drugs became available as a consequence of indirect assumptions, instead of being optimised specifically for the treatment of DoC patients. The purpose of this study is to demonstrate the feasibility of using biophysically-grounded whole-brain models to inform the effect of targeted receptor stimulation in the clinical status of DoC patients. We were able to map different pharmacological interventions into two-dimensional space trajectories, from which we derived features indicative of treatment feasibility and efficacy. Moreover, successful perturbations not only recovered conscious-like dynamics, but also restored to baseline levels of the three metrics previously introduced: Shannon Entropy, correlation to structural connectivity and mean FC value as a proxy for integration.

Our results highlight that stimulation of two serotonin receptor subtypes (2A and 1A) restores dynamics indicative of conscious wakefulness. Interpreted within the context of previous proposals to investigate psychedelics as treatment options in DoC[17], this result can be taken as an additional suggestive element favouring further steps, such as the exploration of animal models. Interestingly, even though the 5HT2A receptor is the main target of psychedelic drugs (i.e., all psychedelics are full or partial agonists at this receptor), we also found comparable results for the 5HT1A receptor. Psychedelic compounds are divided into two main families depending on their chemical structure: substituted tryptamines (including N,N-Dimethyltryptamine and psilocybin) and substituted phenethylamines (such as mescaline). Of these families, tryptamines present significant agonism at serotonin 5HT2A/1A receptors[28]. This raises the possibility that the potential therapeutic effect of psychedelics on DoC patients is not universal for all psychedelics, but can change from compound to

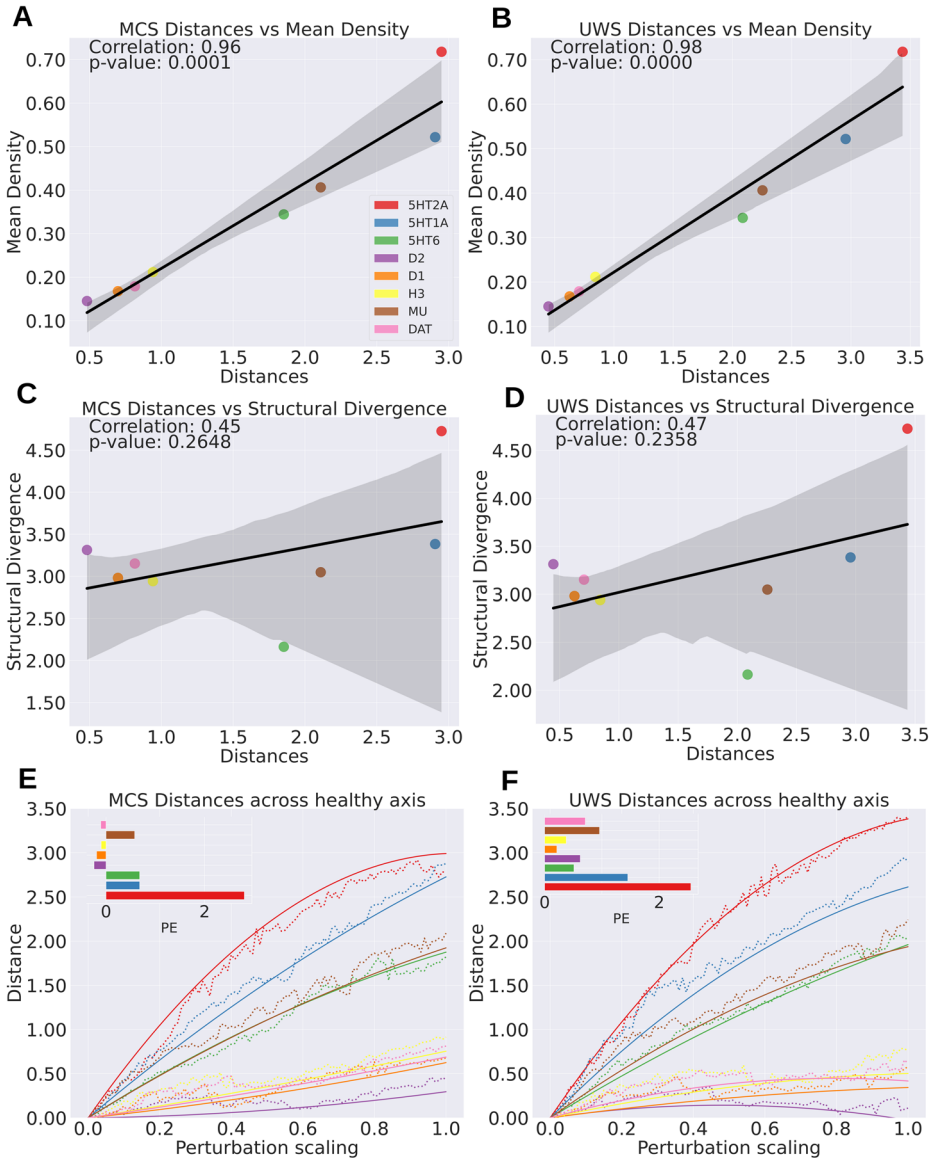

Fig. 5 | Geometric properties show that perturbation trajectories are related to the density distribution of stimulated receptors. Panels (**A**) and (**B**) show that the maximum distance reached from the starting point exhibits a strong linear correlation with the mean receptor density. In panels (**C**) and (**D**), in contrast, we show that there is no significant correlation between the maximum distance reached and the "structural divergence" (representing the spatial relationship between receptor maps and the weight vector of connections). This finding highlights the extent to which the model dynamics can be shifted independently of the spatial arrangement of receptors. **E**, **F** Dotted lines show the perturbational distance accumulated as the *scaling* increases. Solid lines indicate the quadratic regression fitted to the cumulative distances. The inset shows the "perturbational efficiency" (PE) for each receptor. The 5HT2A receptor has the highest PE despite reaching a similar overall perturbational distance than the 5HT1A.

compound. It is also important to note that not all 5HT2A receptor agonists induce psychedelic effects[43], thus it could be possible that a non-psychedelic serotonergic agonist is eventually capable of improving the clinical status of DoC patients.

Our model also predicted that stimulation of opioid μ receptors should have a positive impact on DoC patients. This result is consistent with a recent publication demonstrating that use of opioid drugs during surgery correlates with consciousness improvement[44]. A similar role of μ receptor stimulation is suggested by the study of neurorehabilitation and recovery of brain-injured patients after severe COVID-19 infections, since for MCS patients the amount of oxycodone (an opioid analgesic drug) received during treatment correlated with their recovery indices after the infection[45]. It has been speculated that opioid compounds could have neuroprotective properties, which could be beneficial for brain-injured patients[46]. These studies should be taken with care, however, as they report observational findings instead of the results of clinical trials designed to evaluate the safety and efficacy of opioids in DoC patients.

The validity of the perturbational analysis depends on the capacity of the baseline biophysical models to capture the relevant features of whole-brain dynamics. The results of the model fitting are consistent with previous studies, such as the work of Lopez-Gonzales and colleagues, showing that a less biophysically-realistic whole-brain model optimises the reproduction of fMRI DoC patient dynamics with low coupling (G) values[47]. The same behaviour was demonstrated using the same model fitted to statistical observables derived from the turbulent brain dynamics regime[48]. In our case, the optimal G values correlated with the expected level of consciousness per group (i.e., UWS < < MCS < CNT), in accordance with previous work using this model[49]. Interestingly, this behaviour seems to be invariant with respect to the choice of model and its optimisation targets, thus being indicative of a neurobiological principle.

Another key point of our analysis is the data-driven synthesis of training samples for the VAE. The use of data synthesis or augmentation techniques is a standard procedure when training deep learning algorithms; in this case, data synthesis was obtained as the output of a generative dynamical system trained to capture the whole-brain FC of participants[37,50]. In addition, using the output of the whole-brain model to train the VAE provided a dataset with fixed underlying parameters and controlled initial conditions, which is difficult to obtain from empirical recordings. While the VAE was not strictly necessary to fit the data and simulate the perturbations, it provided a framework to interpret and visualise the results. Instead of reducing our study to evaluate if a given perturbation can induce a healthy-like state, observing them as trajectories in a continuous space opens more

possibilities for analysis. Additionally it facilitates interpretations considering the high-dimensional nature of fMRI data[34], endowing each point in latent space with the value of relevant metrics, namely entropy, mean FC value, and correlation to structure[40,41,51]. Thus, the construction of the latent space allowed us to demonstrate that the three groups (UWS, MCS and CNT) appeared organised within a one-dimensional direction representing the extent of consciousness impairment, which is aligned with previous results obtained using phenomenological models[33]. As expected, this direction was also characterised by a decrease in whole-brain entropy and mean FC, as well as an increase in correlation to anatomical connectivity, of the sampled FC matrices, affirming that the chosen metrics are sensitive to the level of impaired consciousness of the patients.

Previous research investigated the capacity of simulated external perturbations (e.g., TMS and tDCS) to transition towards states of increased consciousness[33,52]. Here, we adopted biophysical models capable of representing the effects of pharmacological perturbations, resulting in a first exploration of simulated drug treatment in DoC patients. In combination with receptor density PET maps of different neuromodulatory systems[53], we introduced heterogeneities in the model by changing the slope of the synaptic scaling function, weighted by local receptor density estimates[36,54]. It is important to note that the *scaling* allows us to understand large-scale neurotransmission in terms of the regional heterogeneity of susceptibility to endogenous and exogenous modulatory signals. Consequently, changes in the *scaling* yield the potential to reflect global shifts in neurotransmission signalling and explain whole-brain dynamics in response to a drug intervention, but these changes are not directly modelling the dosage of a drug. For each value of *scaling* we projected the generated FC into our latent space. As the *scaling* increased, the projected simulations moved from the lower consciousness regions to that corresponding to the healthy controls in a continuous manner. Leveraging the VAE latent space representation, these trajectories could be quantified using simple geometrical properties, as well as by the features previously assigned to each point of the space. We found that perturbing some receptors induced a significant shift in the dynamics of the DoC model as determined by the travelled distance as the *scaling* parameter increased. A key aspect of this work is that although global coupling and neuromodulation account for different mechanisms, when we perturb pharmacologically in our model we can see an effect relatable to global coupling. Even more, although we cannot biologically increase global coupling in a patient, we can modulate responsiveness through receptor stimulation. Of note, the least effective in terms of reaching a maximum change in the dynamics, were the dopaminergic receptors, which are currently used as treatment options[10,15]. This discrepancy could arise due to downstream effects of dopaminergic stimulation not included in our model, or due to the fact that dopaminergic drugs used to treat DoC are non-specific with regards to the targeted dopamine receptor subtype. Future work should attempt to characterise the effect of multi-target stimulation in DoC patients.

Our results show that the mean density of the normalised receptor is highly correlated with the maximum distance reached after the perturbation suggesting an effect driven by the widespread presence of the receptor in the brain; however, this result was not found for the local node strength, suggesting that the underlying topological structure has a limited effect on the outcome of the perturbation. PET measurements may generate different magnitudes of values depending on the chosen region of reference, therefore our result is not linked with an absolute density between receptors, but rather the relative normalised density distribution. Additionally, we measured the impact of different receptors by determining how much scaling was required to shift brain dynamics towards a healthy state. Even though some receptors eventually reached the same overall perturbational distance, they did so at different speeds. Notably, the 5HT2A receptor, the main target of serotonergic psychedelics, presented the highest efficiency in this regard.

The present work contributes to demonstrating the usefulness of in silico experimentation to investigate perturbations to transition between different brain states. It has been recently stated that this type of modelling approach can be a crucial step towards personalised treatments[31,55]. Leveraging the concept of *digital twins*[56] and the idea of *phase 0 clinical trials*[55,57], these computational models offer unique opportunities to explore and predict the effects of therapeutic interventions. Through virtual simulations of patient-specific brain dynamics, "digital twins" provide a testing ground for different treatment strategies, opening doors for personalised and targeted approaches to manage DoC. Additionally, *phase 0 clinical trials* present a platform for early exploratory studies, allowing researchers to assess the safety and efficacy of potential treatments in a limited patient cohort before scaling up to larger trials.

A main limitation of our work is the lack of interaction between neuromodulatory systems in the model. Future work should address this with a more sophisticated modelling of the neuromodulation, for example by perturbing multiple receptors from the same neuromodulatory system. For instance, in the case of LSD, despite its main action being exerted via the serotonergic system, it also exhibits agonism for dopaminergic receptors[58]. Other large scale data-driven works have shown that most mind-altering drugs can be understood in terms of contributions from multi-neurotransmitter systems[59,60]. In the same direction, the impact of the heterogeneities in the model should be explored via different parameters depending on the drug that is modelled. For instance, sedative compounds such as ketamine and memantine are postulated to exert their effects by modulating the balance between excitatory and inhibitory mechanisms. This hypothesis can be experimentally evaluated by manipulating the FIC parameter. It is important to note that absolute density of a PET map is not biologically meaningful due to PET measurements (binding potential) being a relative measure. We addressed this issue by normalising each map and then scaling the modulation of each map within the models which allow us to interpret the results, nevertheless the interpretation of the absolute value of each map in each region is not possible.

Additionally, there are interesting directions to take with respect to non-pharmacological perturbations such as central thalamic Deep Brain Stimulation, which requires implanting electrodes, or non-invasive repetitive transcranial magnetic stimulation. This would also require using a whole-brain parcellation with higher subcortical resolution, such as the Melbourn Subcortex Atlas[61]. Although there are theoretical benefits for subcortical stimulation, the clinical results are not conclusive or still in early stages[62], which could be benefited from in silico experimentation.

Another limitation of our work is the lack of structural connectivity included in the DoC patient model construction. The aberrant function presented by the brains of DoC patients is ultimately underwritten by structural damage. As the model fitted to patient data estimates the structural connectivity from a set of healthy controls, this assumption may omit key aspects of simulating patient brain function of individual patients and assessing the result of introducing pharmacological perturbations. The use of the average SC for modelling the three conditions is justified by the objective of demonstrating that modulations of serotonergic receptors induce transition from DoC patients towards healthy states across participants, without shifting our focus to individual participants. The latter possibility is undermined by the current lack of data at the level of individual subjects. Note that the receptor maps are measured at the group level in an independent cohort, moreover, the same is valid for the SC matrix. Faced with this limitation, we decided to train a model to represent on average each condition. The use of high quality SC was justified by this objective, following multiple previous works by our group and others adopting the same approach. In Supplementary Fig. 4 we show that individually fitted models using the group averaged SC yield an inferior fit. Nevertheless, we consider that investigating how the modulation of dynamics via neurotransmitter receptor activation at the individual level is crucial for a personalised approach. As stated in by Luppi and colleagues[55], given the heterogeneity of DoC it is imperative to create models that are fitted individually. Future publications should aim to improve the accuracy of such models via incorporating patient structural connectivity in their generation. Another unexplored avenue of research is the parametrization of the hemodynamic response model to take into consideration the variations in DoC brain vasculature that would affect BOLD signals. Beyond the particular case of

our work, we believe it would be an interesting step to create a multi-modal biophysical model to profit from abundant EEG data acquired from DoC patients. This type of data has been proven useful to diagnose DoC patients[6,7] and could contribute improved temporal resolution which is lacking from fMRI data.

Our study is based on the hypothesis that manipulating brain activity to reflect a state of heightened complexity and entropy will be reflected in (at least a partial) recovery of consciousness in DoC patients, since brain activity of conscious healthy individuals shows these signatures. However, there is no clear causal relationship between complexity-based neuroimaging markers and levels of consciousness, and such a relationship cannot be demonstrated solely by modelling studies. Instead, our work should be taken as a proof of plausibility in the context of future empirical studies aimed to investigate the causal link between complexity markers and consciousness in DoC patients.

We provided in silico support for the use of serotonergic compounds in the treatment of DoC patients, while also introducing a computational framework to investigate the effect of pharmacological stimulation in neuropsychiatric conditions. Drug development is a lengthy and expensive process, and only a small percentage of the assayed compounds are eventually adopted in clinical practice. The drugs that are currently used to treat DoC patients were not developed for this objective, instead, they were repurposed after their efficacy was discovered. It must be emphasised that our method cannot replace the evidence generated by clinical trials. However, adding whole-brain biophysical models to the arsenal of methods available for the discovery of novel pharmacological treatments could facilitate the concentration of efforts on promising leads, even in cases where the ethical implications could outweigh the unknown benefits to the patients.

Methods

Patients

Our study encompassed a total of 11 patients in MCS (with 5 females; mean age ± s.d., 47.25 ± 20.76 years), 10 patients in UWS (with 4 females; mean age ± s.d., 39.25 ± 16.30 years), and 13 healthy controls (including seven females; mean age ± s.d., 42.54 ± 13.64 years), as detailed in ref. 48. The clinical evaluation and Coma Recovery Scale-Revised (CRS-R) scoring[63] were conducted by trained medical professionals to ascertain the participants' levels of consciousness. A diagnosis of MCS was assigned to patients displaying behaviours potentially indicative of awareness, such as visual tracking, localisation of noxious stimulation, or consistent response to commands. Conversely, patients were categorised as UWS if they exhibited arousal (eye-opening) without any indications of awareness, never displaying purposeful voluntary movements. This study received approval from the local ethics committee Comité de Protection des Personnes Ile de France 1 (Paris, France) under the designation 'Recherche en soins courants' (NEURODOC protocol, no. 2013-A01385-40). Informed consent was obtained from the relatives of the patients, and all investigations adhered to the principles of the Declaration of Helsinki and the regulations of France.

Anatomical connectivity

We employed diffusion MRI (dMRI) data from a cohort of 16 healthy right-handed participants (5 women, mean age: 24.8 ± 2.5) gathered at Aarhus University, Denmark. The research protocol was approved by the internal research board at CFIN, Aarhus University, and received ethical clearance from the Research Ethics Committee of the Central Denmark Region (De Videnskabsetiske Komiteer for Region Midtjylland). All participants provided written informed consent before participating in the study.

Anatomical connectivity acquisition

The imaging data was acquired during a single session using a 3T Siemens Skyra scanner at CFIN, Aarhus University. The structural MRI T1 scan employed the following parameters: voxel size of 1 mm^3; reconstructed matrix size of 256 ×256; echo time (TE) of 3.8 ms, and repetition time (TR) of 2300 ms.

For the dMRI data collection, a TR of 9000 ms, TE of 84 ms, flip angle of 90 degrees, reconstructed matrix size of 106 ×106, voxel size of 1.98 ×1.98 mm with a slice thickness of 2 mm, and a bandwidth of 1745 Hz/Px were used. The dataset was recorded with 62 optimal nonlinear diffusion gradient directions at $b = 1500$ s/mm^2. Additionally, one non-diffusion weighted image ($b = 0$) was acquired per every 10 diffusion-weighted images. The acquisition of dMRI images employed both anterior-to-posterior phase encoding direction: one in anterior to posterior and the opposite in the rest.

Resting-state fMRI acquisition

MRI images of both patients and healthy subjects were acquired using a 3T General Electric Signa System. T2*-weighted whole-brain resting-state images were obtained through a gradient-echo EPI sequence with axial orientation (200 volumes, 48 slices, slice thickness: 3 mm, TR/TE: 2400 ms/30 ms, voxel size: 3.4375 × 3.4375 × 3.4375 mm, flip angle: 90°, FOV: 220 mm^2).

Probabilistic tractography analysis for anatomical data

In this study, we utilised previously obtained structural connectivity data of healthy subjects from ref. 64. Briefly, the averaged whole-brain structural connectivity matrix involved a three-step process: first, defining regions based on the AAL template used in functional MRI data; second, estimating connections (edges) between nodes using probabilistic tractography within the whole-brain network; third, averaging data across participants. In brief, the FSL toolbox's linear registration tool (www.fmrib.ox.ac.uk/fsl, FMRIB, Oxford)[65] was used to co-register the EPI image with the T1-weighted structural image. The T1-weighted image was co-registered to the T1 template of ICBM152 in MNI space. Combining and reversing these transformations, we applied them to map the AAL template[66] from MNI space to the EPI native space, preserving labelling values through nearest-neighbour interpolation for accurate brain parcellations in individual native space. For further details please refer to the original work[64].

Resting state pre-processing and FC estimation

Resting state data preprocessing was conducted using FSL, following the methodology outlined[48]. In summary, resting state fMRI data were processed using MELODIC (multivariate exploratory linear optimised decomposition into independent components)[67]. The process involved discarding the initial five volumes, motion correction using MCFLIRT[65], brain extraction using the brain extraction tool (BET)[68], application of spatial smoothing with a 5 mm FWHM Gaussian kernel, rigid-body registration, high pass filtering with a cutoff of 0.01 Hz, and application of single-session independent component analysis (ICA) with automatic dimensionality estimation. Subsequently, for patients, lesion-driven artefacts were identified and regressed out, along with noise components, independently for each subject, utilising FIX (FMRIB's ICA-based X-noiseifier)[69]. Lastly, FSL tools were employed to co-register the images and extract the time series from the AAL atlas[66] for each subject in MNI space, encompassing 90 cortical brain regions. Subsequently, the mean time series for each parcellated region were extracted, followed by the calculation of Pearson correlations between the time series of each pair of regional areas to derive the interregional FC matrices.

Parcellation

Drawing from our earlier comprehensive whole-brain investigations, we adopted the AAL atlas, specifically focusing on the 90 cortical and subcortical non-cerebellar brain regions[66]. This atlas served as the foundation for integrating all structural, functional, and neuromodulation data. Leveraging FSL tools, we harnessed the publicly accessible receptor density maps in MNI space to derive the mean receptor density for each distinct AAL region in every individual.

Whole brain modelling

In this study, we employed a network model to simulate spontaneous brain activity at the whole-brain level, where individual brain areas were represented as nodes, and the white matter connections between them were depicted as links. The dynamic mean field (DMF) model, as proposed in ref. 30, was utilised to describe the activity in each brain area. This DMF model offers a reductionist approach, summarising the activity of interconnected excitatory (E) and inhibitory (I) spiking neurons into a simplified set of dynamical equations. Within the DMF framework, excitatory synaptic currents $I_i^{(E)}$ are mediated by NMDA receptors, while inhibitory currents $I_i^{(I)}$ are mediated by GABA receptors.

Each brain area i in the DMF model consists of mutually interconnected E and I neuronal pools. Additionally, inter-area coupling between two areas n and p is exclusively established at the E-to-E level and is adjusted based on the structural connectivity C_{ij} (for details, refer to Methods - Anatomical Connectivity). This configuration enables the simulation of brain activity and interactions between different regions, providing insights into the dynamics of large-scale neural networks.

More specifically, the DMF model at the whole-brain level is expressed by the following system of coupled differential equations:

$$I_i^{(E)} = I_i^{ext} + W_E I_0 + w_+ J_{NMDA} S_i^{(E)} + G J_{NMDA} \sum_j C_{ij} S_j^{(E)} - J_i S_i^{(I)} \quad (1)$$

$$I_i^{(I)} = W_I I_0 + J_{NMDA} S_i^{(E)} - S_i^{(I)} \quad (2)$$

$$r_i^{(E)} = H^{(E)}(I_i^{(E)}) = \frac{M_i(a_E I_i^{(E)} - b_E)}{1 - exp(-d_E M_i(a_E I_i^{(E)} - b_E))} \quad (3)$$

$$r_i^{(I)} = H^{(I)}(I_i^{(I)}) = \frac{M_i a_I I_i^{(I)} - b_I}{1 - exp(-d_I(a_I I_i^{(I)} - b_I))} \quad (4)$$

$$M_i = sD_i + 1 \quad (5)$$

$$\frac{dS_i^{(E)}(t)}{dt} = \frac{dS_i^{(E)}}{\tau_E} + (1 - S_i^{(E)})\gamma r_i^{(E)} + \sigma v_i(t) \quad (6)$$

$$\frac{dS_i^{(I)}(t)}{dt} = \frac{dS_i^{(I)}}{\tau_I} + r_i^{(I)} + \sigma v_i(t) \quad (7)$$

In our computational model, each brain area n was characterised by excitatory (E) and inhibitory (I) pools of neurons, where $I_i^{(E,I)}$ (in nA) represents the total input current, and $r_i^{(E,I)}$ (in Hz) represents the firing rate. Additionally, $S_i^{(E,I)}$ denotes the synaptic gating variable. The neuronal nonlinear response functions, $H^{(E,I)}$, were applied to convert the total input currents received by the E and I pools into firing rates, $r_i^{(E,I)}$, based on the input-output function proposed by Abbott and Chance[70]. The model employed specific gain factors, threshold currents, and constants to determine the shape of the curvature of H around the threshold.

In this study we used receptor density maps from healthy subjects estimated using PET tracer studies obtained by Hansen et al.[53]. All PET images were registered to the MNI-ICBM 152 nonlinear 2009 (version c, asymmetric) template and subsequently parcellated to the 90 region AAL atlas[66]. More details on the acquisitions and limitations of this dataset can be found in ref. 53. This processed dataset provided us with a quantitative measure of receptor densities in each AAL region denoted as D_i. These density values were normalised using min-max normalisation:

$$D' = D/(\max(D) - \min(D))$$

$$D_{norm} = D' - \max(D) + 1$$

.Leveraging D_{norm-i}, we modulated the firing rates $r_i^{(E,I)}$ of the excitatory and inhibitory neuronal pools in each brain region, drawing inspiration from[70]. Receptors were thus used to modulate the gain of the neuronal response function $H^{(E,I)}$ in each brain area.

The parameters of the DMF model were calibrated to mimic resting-state conditions, ensuring that each isolated node exhibited the typical noisy spontaneous activity with a low firing rate ($r < 3$ Hz) observed in electrophysiology experiments. Additionally, the inhibition weight, J_i, was adjusted for each node i to achieve Feedback Inhibition Control (FIC). This regulation, described in ref. 32, ensured that the average firing rate of the excitatory pools $r_i^{(E)}$ remained clamped at 3 Hz even when receiving excitatory input from connected areas. The FIC was shown to lead to a better prediction of resting-state functional connectivity (FC) and more realistic evoked activity.

The synaptic gating variable for excitatory pools, $S_i^{(E)}$, was governed by NMDA receptors, with specific decay time constant tNMDA = 0.1 s and a constant g = 0.641. On the other hand, the average synaptic gating in inhibitory pools depended on GABA receptors with a decay time constant tGABA = 0.01 s. The overall effective external input was represented by $I_0 = 0.382$ nA, with specific weights $W_E = 1$ and $W_I = 0.7$ for excitatory and inhibitory pools, respectively. The model further considered recurrent excitation with a weight of $w_+ = 1.4$ and excitatory synaptic coupling with weight $J_{NMDA} = 0.15$ nA. Gaussian noise, v_i, with an amplitude of $\sigma = 0.01$ nA was introduced in the system.

In our whole-brain network model, we considered N = 90 brain areas after parcellation of the structural and functional MRI data. Each area n received excitatory input from all structurally connected areas p into its excitatory pool, weighted by the connectivity matrix Cnp derived from dMRI data. All inter-area E-to-E connections were equally scaled by a global coupling factor G. This global scaling factor was adjusted to optimise the system's working point, where the simulated activity best matched the empirical resting-state activity of participants under placebo conditions. To explore different scenarios, simulations were run for a range of G values between 0 and 2.5 with an increment of 0.025 and a time step of 1 ms. Each G value underwent 20 simulations of 192 s duration, mirroring the empirical resting-state scans of 10,11 and 13 participants for the UWS, MCS and CNT conditions respectively.

Generated BOLD signal

To map the simulated mean field activity to a BOLD signal, we adopted the generalised hemodynamic model proposed by Stephan et al.[71]. The BOLD signal in each brain area i was computed based on the firing rate of the excitatory pools $r_i^{(E)}$. In this model, an increase in the firing rate led to an increase in a vasodilatory signal, s_i, which was further subject to autoregulatory feedback. The changes in blood inflow, f_i, were proportionally influenced by this signal, resulting in alterations in blood volume v_i and deoxyhemoglobin content q_i. The relationships between these biophysical variables were governed by the following equations:

$$\frac{ds_i(t)}{dt} = 0.5 r_i^{(E)} + 3 - s_i k_i - \gamma_i(f_i - 1) \quad (8)$$

$$\frac{df_i(t)}{dt} = s_i \quad (9)$$

$$\tau_i \frac{dv_i(t)}{dt} = f_i - v_i^{1/\alpha} \quad (10)$$

$$\tau_i \frac{dq_i(t)}{dt} = \frac{f_i(1 - (1 - \rho_i)^{1/f_i})}{\rho_i} - \frac{q_i v_i^{1/\alpha}}{v_i} \quad (11)$$

where ρ is the resting oxygen extraction fraction, τ is a time constant and α that represents the resistance of the veins. Finally, the BOLD signal in each area i, B_i, is a static nonlinear function of volume, v_i, and deoxyhemoglobin content, q_i, that comprises a volume-weighted sum of extra- and

intravascular signals:

$$B_n = V_0[(k_1(1-q_i) + k_2(1-q_i/v_i) + k_3(1-v_i)] \quad (12)$$

All biophysical parameters were taken as in[71]. To concentrate on the frequency range where resting-state activity appears the most functionally relevant, both empirical and simulated BOLD signals were band pass filtered between 0.01 and 0.1 Hz. As with the empirical data we derived the inter-regional FC matrices between each pair of nodes by calculating the Pearson correlations of their time series.

Randomised receptor map

To confirm that the shift towards healthy-like dynamics was not driven by structure we created a randomised variation of the 5HT2A receptor map while preserving spatial autocorrelation. For this we used the Neuromaps library[72] which implements various null models. Concretely we used the 'moran' function which uses Moran spectral randomisation[42] to create the null variations of the receptor map. We used min-max normalisation rescaled accordingly to obtain the same mean as the original receptor map.

Metrics

Shannon Entropy: To calculate the Shannon entropy of the functional connectivity (FC) matrix, we divided the connectivity values into 30 equal-width bins, with each bin representing a range of connectivity values. The frequency of FC values falling into each bin was counted to create a histogram, which was subsequently normalised by dividing each bin's count by the total number of FC values, yielding a probability distribution. Shannon entropy H was then calculated using the formula

$$H = -\sum_i^n \frac{1}{n} P(i) log(P(i))$$

where $P(i)$ is the probability of the FC values falling into the i-th bin and n is the total number of bins (30). Only non-zero probabilities were included in the summation to avoid undefined values from the logarithm of zero. This procedure quantified the entropy of the FC matrix, providing a measure of the complexity and variability of the functional connectivity pattern.

Mean FC: It's been shown that the decrease in this value is related to the level of unconsciousness[33]. Particularly, the nodes that exert the biggest changes are the ones that have inter modular connections suggesting a change in the integrative capacity of the network. This value was calculated by taking the overall mean of the FC matrix.

Variational Autoencoders

We employed a Variational Autoencoder (VAE)[32] to encode the functional connectivity matrices C_{ij} into a low-dimensional representation. VAEs are neural networks that map inputs to probability distributions in a latent space, allowing for regularisation during training to generate meaningful outputs after decoding the latent space coordinates. The VAE architecture (depicted in Fig. 1) comprises three components: the encoder network, the middle variational layer, and the decoder network.

The encoder network is a deep neural network utilising rectified linear units (ReLU) as activation functions and includes two dense layers. This part of the network bottlenecks into the two-dimensional variational layer, represented by units z_1 and z_2, which span the latent space. The encoder applies a nonlinear transformation to map the functional connectivity matrices (C_{ij}) into Gaussian probability distributions in the latent space. Conversely, the decoder network mirrors the encoder architecture and reconstructs matrices C_{ij} from samples of these distributions.

The VAE is trained using error backpropagation via gradient descent to minimise a loss function consisting of two terms. The first term is a standard reconstruction error computed from the units in the output layer of the decoder. The second term is a regularisation term measured as the Kullback-Leibler divergence between the distribution in latent space and a standard Gaussian distribution. This regularisation term ensures continuity and completeness in the latent space, ensuring that similar values are decoded into similar outputs and that these outputs represent meaningful combinations of the encoded inputs.

For training the VAE model, we generated 9000 correlation matrices (C_{ij}) corresponding to healthy control, UWS and MCS conditions. The model's hyperparameters were optimised using a training set, which was created by randomly splitting the dataset into 80% training and 20% test sets. The training procedure involved using batches with 256 samples and 50 training epochs, with the Adam optimiser and the loss function described in the previous paragraph.

Statistics and reproducibility

Effect size to compare model simulations were calculated using Cohen's d. Perturbational effects in the latent space were evaluated using Pearson correlation, quantifying the linear relationship between variables. A Wilcoxon signed rank test was used in Supplementary Fig. 3.

Data availability

Source data to reproduce figures is available on Figshare (https://doi.org/10.6084/m9.figshare.26728804)[73] along with the code (See Code availability). The disorder of consciousness datasets contain information from a clinical population and are not publicly available due to constraints imposed by the approved ethics protocol. Data can be shared upon request to the authors.

Code availability

All code is publicly available in https://doi.org/10.5281/zenodo.13330349[74].

Received: 31 January 2024; Accepted: 5 September 2024;
Published online: 19 September 2024

References

1. Giacino, J. T. et al. The minimally conscious state: definition and diagnostic criteria. *Neurology* **58**, 349–353 (2002).
2. Naccache, L. Minimally conscious state or cortically mediated state? *Brain* **141**, 949–960 (2018).
3. Fins, J. J. When no one notices: disorders of consciousness and the chronic vegetative state. *Hastings Cent. Rep.* **49**, 14–17 (2019).
4. Owen, A. M. Improving diagnosis and prognosis in disorders of consciousness. *Brain* **143**, 1050–1053 (2020).
5. Comanducci, A. et al. Clinical and advanced neurophysiology in the prognostic and diagnostic evaluation of disorders of consciousness: review of an IFCN-endorsed expert group. *Clin. Neurophysiol.* **131**, 2736–2765 (2020).
6. Hermann, B. et al. Multimodal FDG-PET and EEG assessment improves diagnosis and prognostication of disorders of consciousness. *NeuroImage Clin.* **30**, 102601 (2021).
7. Sitt, J. D. et al. Large scale screening of neural signatures of consciousness in patients in a vegetative or minimally conscious state. *Brain* **137**, 2258–2270 (2014).
8. Deco, G. & Kringelbach, M. L. Great expectations: using whole-brain computational connectomics for understanding neuropsychiatric disorders. *Neuron* **84**, 892–905 (2014).
9. Marino, M. H. & Whyte, J. Treatment trials in disorders of consciousness: challenges and future directions. *Brain Sci.* **12**, 569 (2022).
10. Giacino, J. T. et al. Placebo-controlled trial of amantadine for severe traumatic brain injury. *N. Engl. J. Med.* **366**, 819–826 (2012).
11. Schiff, N. D. et al. Behavioural improvements with thalamic stimulation after severe traumatic brain injury. *Nature* **448**, 600–603 (2007).
12. Bourdillon, P., Hermann, B., Sitt, J. D. & Naccache, L. Electromagnetic brain stimulation in patients with disorders of consciousness. *Front. Neurosci.* **13**, 445903 (2019).
13. Bomalaski, M. N., Claflin, E. S., Townsend, W. & Peterson, M. D. Zolpidem for the treatment of neurologic disorders: a systematic review. *JAMA Neurol.* **74**, 1130–1139 (2017).

14. Whyte, J. et al. Zolpidem and restoration of consciousness. *Am. J. Phys. Med. Rehabil.* **93**, 101–113 (2014).
15. Rühl, L. et al. Amantadine treatment is associated with improved consciousness in patients with non-traumatic brain injury. *J. Neurol. Neurosurg. Psychiatry* **93**, 582–587 (2022).
16. Giacino, J. T. et al. Practice guideline update recommendations summary: disorders of consciousness: Report of the Guideline Development, Dissemination, and Implementation Subcommittee of the American Academy of Neurology; the American Congress of Rehabilitation Medicine; and the National Institute on Disability, Independent Living, and Rehabilitation Research. *Neurology* **91**, 450–460 (2018).
17. Scott, G. & Carhart-Harris, R. L. Psychedelics as a treatment for disorders of consciousness. *Neurosci Conscious* **2019**, niz003 (2019).
18. Carhart-Harris, R. L. et al. The entropic brain: a theory of conscious states informed by neuroimaging research with psychedelic drugs. *Front. Hum. Neurosci.* **8**, 20 (2014).
19. Tagliazucchi, E., Carhart-Harris, R., Leech, R., Nutt, D. & Chialvo, D. R. Enhanced repertoire of brain dynamical states during the psychedelic experience. *Hum. Brain Mapp.* **35**, 5442–5456 (2014).
20. Koch, C., Massimini, M., Boly, M. & Tononi, G. Neural correlates of consciousness: progress and problems. *Nat. Rev. Neurosci.* **17**, 307–321 (2016).
21. Carhart-Harris, R. et al. Trial of psilocybin versus escitalopram for depression. *N. Engl. J. Med.* **384**, 1402–1411 (2021).
22. Davis, A. K. et al. Effects of psilocybin-assisted therapy on major depressive disorder: a randomized clinical trial. *JAMA Psychiatry* **78**, 481–489 (2021).
23. Goodwin, G. M. et al. Single-dose psilocybin for a treatment-resistant episode of major depression. *N. Engl. J. Med.* **387**, 1637–1648 (2022).
24. Palhano-Fontes, F. et al. Rapid antidepressant effects of the psychedelic ayahuasca in treatment-resistant depression: a randomized placebo-controlled trial. *Psychol. Med.* **49**, 655–663 (2019).
25. Moreno, F. A., Wiegand, C. B., Taitano, E. K. & Delgado, P. L. Safety, tolerability, and efficacy of psilocybin in 9 patients with obsessive-compulsive disorder. *J. Clin. Psychiatry* **67**, 1735–1740 (2006).
26. Ly, C. et al. Psychedelics promote structural and functional neural plasticity. *Cell Rep.* **23**, 3170–3182 (2018).
27. Guldenmund, P., Vanhaudenhuyse, A., Boly, M., Laureys, S. & Soddu, A. A default mode of brain function in altered states of consciousness. *Arch. Ital. Biol.* **150**, 107–121 (2012).
28. Nichols,D. E. Psychedelics. *Pharmacol Rev.* **68**, 264–355 (2016).
29. Peterson, A., Tagliazucchi, E. & Weijer, C. The ethics of psychedelic research in disorders of consciousness. *Neurosci. Conscious* **2019**, niz013 (2019).
30. Deco, G. et al. How local excitation–inhibition ratio impacts the whole brain dynamics. *J. Neurosci.* **34**, 7886–7898 (2014).
31. Kringelbach, M. L. et al. Dynamic coupling of whole-brain neuronal and neurotransmitter systems. *Proc. Natl Acad. Sci.* **117**, 9566–9576 (2020).
32. Kingma, D. P. & Welling, M. Auto-encoding variational bayes. Preprint at https://arxiv.org/abs/1312.6114 (2013).
33. Perl, Y. S. et al. Low-dimensional organization of global brain states of reduced consciousness. *Cell Rep.* **42**, 112491 (2023).
34. Perl, Y. S. et al. Generative embeddings of brain collective dynamics using variational autoencoders. *Phys. Rev. Lett.* **125**, 238101 (2020).
35. Luppi, A. I. et al. Whole-brain modelling identifies distinct but convergent paths to unconsciousness in anaesthesia and disorders of consciousness. *Commun. Biol.* **5**, 384 (2022).
36. Deco, G. et al. Whole-brain multimodal neuroimaging model using serotonin receptor maps explains non-linear functional effects of LSD. *Curr. Biol.* **28**, 3065–3074.e6 (2018).
37. Perl, Y. S. et al. Data augmentation based on dynamical systems for the classification of brain states. *Chaos Solitons Fractals* **139**, 110069 (2020).
38. Wang, Z., Bovik, A. C., Sheikh, H. R. & Simoncelli, E. P. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **13**, 600–612 (2004).
39. Viol, A. et al. Shannon entropy of brain functional complex networks under the influence of the psychedelic Ayahuasca. *Sci. Rep.* **7**, 7388 (2017).
40. Barttfeld, P. et al. Signature of consciousness in the dynamics of resting-state brain activity. *Proc. Natl Acad. Sci.* **112**, 887–892 (2015).
41. Demertzi, A. et al. Human consciousness is supported by dynamic complex patterns of brain signal coordination. *Sci. Adv.* **5**, eaat7603 (2019).
42. Wagner, H. H. & Dray, S. Generating spatially constrained null models for irregularly spaced data using Moran spectral randomization methods. *Methods Ecol. Evol.* **6**, 1169–1178 (2015).
43. López-Giménez, J. F. & González-Maeso, J. Hallucinogens and Serotonin 5-HT2A receptor-mediated signaling pathways. *Curr. Top. Behav. Neurosci.* **36**, 45–73 (2018).
44. Ge, Q. et al. Opioid-induced short-term consciousness improvement in patients with disorders of consciousness. *Front. Neurosci.* **17**, 1117655 (2023).
45. Gurin, L. et al. Early neurorehabilitation and recovery from disorders of consciousness after severe COVID-19. *Neurocrit. Care.* **36**, 357–371 (2022).
46. Vaidya, B., Sifat, A. E., Karamyan, V. T. & Abbruscato, T. J. The neuroprotective role of the brain opioid system in stroke injury. *Drug Discov. Today* **23**, 1385–1395 (2018).
47. López-González, A. et al. Loss of consciousness reduces the stability of brain hubs and the heterogeneity of brain dynamics. *Commun. Biol.* **4**, 1–15 (2021).
48. Escrichs, A. et al. Unifying turbulent dynamics framework distinguishes different brain states. *Commun. Biol.* **5**, 1–13 (2022).
49. Luppi, A. I. et al. Reduced emergent character of neural dynamics in patients with a disrupted connectome. *NeuroImage* **269**, 119926 (2023).
50. Tubaro, P. L. & Mindlin, G. B. A dynamical system as the source of augmentation in a deep learning problem. *Chaos Solitons Fractals X.* **2**, 100012 (2019).
51. Tononi, G. & Edelman, G. M. Consciousness and complexity. *Science* **282**, 1846–1851 (1998).
52. Perl, Y. S. et al. Perturbations in dynamical models of whole-brain activity dissociate between the level and stability of consciousness. *PLOS Comput. Biol.* **17**, e1009139 (2021).
53. Hansen, J. Y. et al. Mapping neurotransmitter systems to the structural and functional organization of the human neocortex. *Nat. Neurosci.* **25**, 1569–1581 (2022).
54. Herzog, R. et al. A whole-brain model of the neural entropy increase elicited by psychedelic drugs. *Sci. Rep.* **13**, 6244 (2023).
55. Luppi, A. I. et al. Computational modelling in disorders of consciousness: closing the gap towards personalised models for restoring consciousness. *NeuroImage* **275**, 120162 (2023).
56. Vohryzek, J. et al. Dynamic sensitivity analysis: Defining personalised strategies to drive brain state transitions via whole brain modelling. *Comput. Struct. Biotechnol. J.* **21**, 335–345 (2023).
57. Dagnino, P. et al. Re-awakening the brain: forcing transitions in disorders of consciousness by external in silico perturbation. *PLoS Comput. Biol.* **20**, e1011350 (2024).
58. Marona-Lewicka, D. & Nichols, D. E. Further evidence that the delayed temporal dopaminergic effects of LSD are mediated by a mechanism different than the first temporal phase of action. *Pharm. Biochem. Behav.* **87**, 453–461 (2007).

59. Ballentine, G., Friedman, S. F. & Bzdok, D. Trips and neurotransmitters: discovering principled patterns across 6850 hallucinogenic experiences. *Sci. Adv.* **8**, eabl6989 (2022).
60. Luppi, A. I. et al. In vivo mapping of pharmacologically induced functional reorganization onto the human brain's neurotransmitter landscape. *Sci. Adv.* **9**, eadf8332 (2023).
61. Tian, Y., Margulies, D. S., Breakspear, M. & Zalesky, A. Topographic organization of the human subcortex unveiled with functional connectivity gradients. *Nat. Neurosci.* **23**, 1421–1432 (2020).
62. Vanhoecke, J. & Hariz, M. Deep brain stimulation for disorders of consciousness: Systematic review of cases and ethics. *Brain Stimul.* **10**, 1013–1023 (2017).
63. Kalmar, K. & Giacino, J. T. The JFK coma recovery scale-revised. *Neuropsychol. Rehabil.* **15**, 454–460 (2005).
64. Deco, G., Kringelbach, M. L., Jirsa, V. K. & Ritter, P. The dynamics of resting fluctuations in the brain: metastability and its dynamical cortical core. *Sci. Rep.* **7**, 3095 (2017).
65. Jenkinson, M., Bannister, P., Brady, M. & Smith, S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage* **17**, 825–841 (2002).
66. Tzourio-Mazoyer, N. et al. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage* **15**, 273–289 (2002).
67. Beckmann, C. F. & Smith, S. M. Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Trans. Med. Imaging* **23**, 137–152 (2004).
68. Smith, S. M. Fast robust automated brain extraction. *Hum. Brain Mapp.* **17**, 143–155 (2002).
69. Griffanti, L. et al. ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging. *NeuroImage* **95**, 232–247 (2014).
70. Abbott, L. F. & Chance, F. S. Drivers and modulators from push-pull and balanced synaptic input. *Prog. Brain Res.* **149**, 147–155 (2005).
71. Stephan, K. E., Weiskopf, N., Drysdale, P. M., Robinson, P. A. & Friston, K. J. Comparing hemodynamic models with DCM. *NeuroImage* **38**, 387–401 (2007).
72. Markello, R. D. et al. Neuromaps: structural and functional interpretation of brain maps. *Nat. Methods* **19**, 1472–1479 (2022).
73. Mindlin I. Data for figures. https://doi.org/10.6084/m9.figshare.26728804 (2024).
74. Mindlin I. Picardian14/pharmacological-perturbations: Communications Biology release. *Zenodo* (2024).

Acknowledgements

The authors express gratitude to all the individuals who participated in the studies. Additionally, we extend our thanks to the dedication and support provided by the clinicians at the Neuro ICU, DMU Neurosciences, APHP Sorbonne Université, Hôpital de la Pitié Salpêtrière, Paris, France. Moreover, we acknowledge the invaluable contributions of patient families whose consent and understanding significantly contribute to the advancement of the field. We finally thank the financial support of the FLAG-ERA research funding organisation (project ModelDXConsciousness). Y.S.P.is supported by European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant 896354, and 'ERDF A way of making Europe', ERDF, EU, Project NEurological MEchanismS of Injury, and Sleep-like cellular dynamics (NEMESIS; ref. 101071900) funded by the EU ERC Synergy Horizon Europe. M.L.K. is supported by the Center for Music in the Brain, funded by the Danish National Research Foundation (DNRF117), and Centre for Eudaimonia and Human Flourishing at Linacre College funded by the Pettit and Carlsberg Foundations. ET is supported by grants FONCYT-PICT (2019-02294), CONICET-PIP (11220210100800CO), and ANID/FONDECYT Regular (1220995). O.G. is research associate and N.A. is research fellow at FNRS. G.D. is supported by grant no. PID2022-136216NB-I00 funded by MICIU/AEI/10.13039/501100011033 and by 'ERDF A way of making Europe', ERDF, EU, Project NEurological MEchanismS of Injury, and Sleep-like cellular dynamics (NEMESIS; ref. 101071900) funded by the EU ERC Synergy Horizon Europe, and AGAUR research support grant (ref. 2021 SGR 00917) funded by the Department of Research and Universities of the Generalitat of Catalunya.

Author contributions

Mindlin I: Conceptualisation, Methodology, Software, Formal analysis, Investigation, Writing—Original Draft, Visualisation. Herzog R: Software, Conceptualisation, Writing—Review & Editing. Belloli L: Conceptualisation, Writing—Review & Editing. Manasova D Software, Conceptualisation, Writing—Review & Editing, Monge-Asensio M: Software, Resources, Writing - Review & Editing, Vohryzek J: Resources, Writing - Review & Editing. Anira Escrichs: Resources, Writing—Review & Editing. Alnagger N: Writing—Review & Editing. Núñez Novo P: Writing—Review & Editing. Gosseries O: Writing—Review & Editing, Morten L. Kringelbach: Writing—Review & Editing. Deco G: Conceptualisation, Writing—Review & Editing. Tagliazucchi E: Conceptualisation, Writing—Review & Editing, Naccache L: Resources, Writing—Review & Editing. Rohaut B: Resources, Writing—Review & Editing. Sitt* JD: Supervision, Conceptualisation, Project Administration, Funding Acquisition, Writing—Review & Editing. Sanz Perl* Y Supervision, Methodology, Software, Conceptualisation, Project Administration, Writing—Review & Editing.

Competing interests

The authors declare to have no conflict of interest. E.T. is an Editorial Board Member for Communications Biology, but was not involved in the editorial review of, nor the decision to publish this article.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-024-06852-9.

Correspondence and requests for materials should be addressed to I. Mindlin, J. D. Sitt or Y. Sanz Perl.

Peer review information *Communications Biology* thanks Amy Kuceyeski and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Manuel Breuer. A peer review file is available.