nature portfolio

Peer Review File

Whole brain modelling for simulating pharmacological interventions on patients with Disorders of Consciousness

Reviewers' comments:

Reviewer #1 (Remarks to the Author):

This paper presents a super interesting, rigorous computational modeling approach to investigate how pharmacological agents may impact brain functon(and perhaps consciousness) in patient with disorders of consciousness. Their results are convincing for the most part, and the potential clinical application of this model in designing or identifying drugs that could aid in recovery of consciousness in specific patients is potentially life-changing. Overall, this paper is potentially very impactful and paradigm shifting in the development of neurological treatments. However, I think more attention needs to be paid to individual-level variability (if we are to use this to model individual patients) and the spatial specificity of the receptor maps in the simulated drug models.

1) It is unclear if the SC and FC matrices of the subjects in each group were averaged, and then the model fitted (I am assuming this is the case for the rest of the review). If so, there may be individual-level heterogeneity of damage/injury that is smoothed out. it would be good to replicate the main findings on an individual level model approach - this is relevant clinically if we are to apply this to single subjects who have noisy SC and FC matrices.

2) Relatedly, it would be interesting to see the individual-level variability in the generated models - when you have noisier data you may have worse fit and more variance in the simulations. perhaps adding in a few examples of generated data from each individual would be interesting to compare against the population level model simulations (fig 3)

3) Since we know what happens in controls who take serotonergic, domanergic and histiminergic drugs, it would be a good validation to apply the receptor simulation studies to the controls as well to make sure the model predicts well that modulation. relatedly, would it not be important to show you can move in the other direction - from conscious control to unconsciousness. perhaps this can be done my applying a simulation of anesthesia to control model.

4) The paper makes the claim that "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 density throughout the brain." but this is not what they have tested with a lack of correlation between distance from start and spatial density of SC. Rather, they should test if a random spin-null spatial pattern of receptor maps (spin test) with receptor density held constant would also result in a trajectory along the same line. This would more fully test if (for example) the 5HT2A map is (or is not) optimally organized to maximize distance from start (similar to what was found in simulation studies via network control theory in Singleton et al. 2022, Nature Communications).

5) It may not be that the trajectories are "forbidden" as this simulation is only done using one SC and one FC - perhaps the individual variability in the SC and FC result in drastically varied predicted trajectories (see previous comment)

6) The pertubation scaling also appears to exactly reflect receptor density - again, this analysis may benefit from null spin-test modeling where the receptor density is held identical and the regional receptor maps are spun to obtain a better measure of how the spatial organization of the actual receptors is impacting the model. it is important to show that these results are not just all reflecting average receptor density (and if they are, then that is important information as well!)

For transparency, I will sign the review - Amy Kuceyeski

Reviewer #2 (Remarks to the Author):

This is a very interesting an ambitious paper that sets out to devise computational treatments for disorders of consciousness based on real patients data and the maps of receptors in the human brain, using a sophisticated brain model and deep learning.
I would be enthusiastic about this work, were it not for two main criticisms. First, many method choices need to be more thoroughly justified and compared with the authors' own previous work, especially Deco 2018 Current Biology which is conceptually similar but the authors changed many things without telling the reader why (fitting function, addition of the deep learning VAE). I add some details below.

Second, the authors show that their results are entirely dependent on mean density, and not at all on where the receptor is more or less expressed, since the success of each PET map at inducing transitions from DoC to control is extremely correlated with the PET map's mean value. This is very unfortunate because mean value of a PET radiotracer map such as these is biologically meaningless. Binding Potential is a relative measurement rather than an absolute quantification of the expression of a receptor. Comparisons between maps with different tracers are not possible. At the moment, the results = do not support the paper's title, because they are really about mean value of a PET tracer, not about identity of the receptor. I provide more detail below. It is possible that I am wrong, in fact I hope so. If so I would welcome an explanation, perhaps I have overlooked some aspect of the modelling framework that removes this concern, but I do not see this addressed. But if I am right, then my only suggestion would be to use some normalization to remove the difference in mean density and variability to make the map really comparable, and repeat the work. For this reason I recommend major revisions.

Some points of lesser importance:
Reference styles are inconsistent at times.
Figure 2A: the light blue is hardly visible.
Line 396: 5HT1A/1A should be 2A?

About the Abstract:
Clinically, DoC are defined in terms of responsiveness.

About the Introduction:

Please provide additional information and references about Cortically Mediated State, and why MCS is not a “minimally conscious” state? This part was unclear.

Can the authors provide more details about how zolpidem works, beside simply stating that it is paradoxical?

Do any theory of consciousness aside from Carhartt-Haris actually state that psychedelics increase consciousness? Bayne and Carter 2018 Neuroscience of Consciousness state quite clearly that psychedelics do not increase consciousness.

What is the relevance for DoC of the use of 5HT2A agonists for OCD and depression?

To this reviewer’s knowledge, the density of 5HT2A is maximal in V1 cortex, which is primary region, the opposite of association areas mentioned by the authors (although it is also high there). How can these facts be reconciled?

The role of the VAE is unclear: why add this extra step to the existing setup of Deco 2018? The explanation for making the procedure even more complicated should be thoroughly justified. The authors need to show what benefit is obtained from stacking deep learning with the mean field model. The authors acknowledge in the Discussion that “the VAE was not strictly necessary to fit the data and simulate the perturbations, it provided a framework to facilitate the interpretation and visualisation of the results” but for me, it made the process more obscure, adding an extra black box element with unclear benefit.

About the Results

Deco 2018 used a different procedure for fitting the same model, showing that FC was a poor choice. Why go back to a provably inferior fitting target?

Figure 2B is puzzling. It appears that simulations based on CNT connectome match the CNT FC better than they match the MCS or UWS condition, as the authors state. However, the simulations based on MCS and UWS connectomes are both better at matching the controls’ FC, than the simulation based on CNT’s own connectome. So, simulation from patient data are the ones that best match control data? This casts doubt on the model quality for CNT.
VAE: why need two dimensions, if the model only has one free dimension (the global coupling)?

LZ: what is the meaning of LZ complexity of a FC matrix (and why do the authors talk of “pixel” in the Methods)? The entropic theory and papers based on it talk about LZ of brain signals (so in time), not of connectivity. I believe that LZ of a matrix is not invariant to reordering of the rows and columns (for example, from left-right-left-right to left hemisphere then right hemisphere). But the information that the matrix conveys about the brain remains the same. Can the authors explain their choice of LZ for an FC matrix, and demonstrate that the result is the same if regions in the matrix are reordered? An ordering-dependent result would seem meaningless. Also, the DMF can produce BOLD signals. Why not use the LZ of the BOLD signals?

Measure of integration: what is the meaning of shortest paths on a functional matrix? Links in a functional matrix are correlations, so what is the meaning of a path made of correlations? Every region has a correlation with every other region. I know that this is often done in the literature, but it does not make it any more meaningful.

Pharmacology. It is nice that the UWS trajectory goes through the MCS space before getting to control. But how do the authors explain that some receptors induce much greater progress but all are along the same trajectory? This seems a bit strange.
If I am correct the pharmacology as implemented is equivalent to making the model progressively more excitable. This is opposite of the reduction in G parameter. Are these two just being played against each other? What would be the trajectory, if instead of the pharmacological manipulation, the authors simply increased the global coupling in the UWS or MCS model?

Can the authors show brain maps of each receptor map on the brain? It is known that some tracers have very different PET intensity in cortex and subcortex. Readers should be able to see the maps for themselves.

I am not familiar with the structural divergence. Could the authors provide scatter plots of each PET map against structural node strength, and the corresponding correlation between the two? This is a more easily interpretable measure, and the visual demonstration would be very helpful for the reader.

It is very unfortunate that the authors found that the topology of the stimulated nodes does not matter: what matters is the mean density ("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 density throughout the brain"). The authors do not seem to realise it, but this is a major problem for their entire study. Of course a map with larger values will show bigger effects under the authors' implementation (could the authors confirm this?). Unfortunately, absolute density of a PET map is not biologically meaningful because PET measurements (binding potential) are relative, and do not measure absolute expression of a receptor in a region. Comparisons between regions are still possible for each individual PET tracer, but the authors show that regional differences do not matter. However, comparisons of density between different tracers (so, different maps), are meaningless.

Given differences in mean density between maps, I do not know how to interpret the results about "perturbational efficiency". However, this section is also not very well explained, I think the reader should be given more explanation (why a quadratic regression, for example?).

About the Discussion

This section is almost entirely based on the assumption that the results have a meaningful biological interpretation in terms of receptor identity, which unfortunately they do not if they just depend on mean map density. For example, the interpretation of dopamine receptors as least

effective (and attempt to justify this) is not appropriate: those are just the maps with least density. As far as I can tell, it is just happenstance that the 5HT2A map, the one that the authors were more interested in, is also the one with the greatest mean density.

The Limitations section is appropriate, but omits to discuss that dependence on the mean value of the PET map is a critical problem that makes the results uninterpretable from a biological point.

Reviewer #3 (Remarks to the Author):

There has recently been discussion about the possibility that serotonergic psychedelics could be useful for restoring awareness of patients with disorders of consciousness. This hypothesis is based on the well-replicated finding that psychedelics push the brain towards a higher-entropy dynamical regime (as measured with complexity metrics like the Lempel-Ziv complexity), while patients with DoC typically have lower-entropy dynamics. Given the obvious ethical issues around giving brain-damaged patients mind-altering drugs, in this paper Mindlin et al., take a computational approach: by fitting in silico models of brain activity to data from patients with DoC, they can explore the effects of perturbing different receptors. They find that activation of serotonin receptors (as would be done by a psychedelic) creates patterns in the DoC brain that are more similar to normal, conscious brains.

This is a fascinating piece of science, and one that I would be happy to see added to the literature. I do have some conceptual/technical questions I would like the authors to respond to, and demand a small change to the network analysis, but by and large, I think that this is a fine paper

- It seems like the use of fMRI-based data is a very serious confound here, since fMRI data is based on a vascular signal. Serotonergic psychedelic drugs are powerful cranial vasoconstrictors (a fact that is almost never addressed in the literature), and so it is unclear to me how many of the results of fMRI-based analyses of psychedelics (which are cited as justification for this study) are compromised by changes to vascular tone following something like LSD or psilocybin. Similarly, the global atrophy seen in DoC brains also has vascular impacts. I do not think we can assume that the hemodynamic response function is identical in normal consciousness, psychedelia, and DoC. This seems like a project that would be much better suited for use with EEG or MEG data, rather than fMRI data.

- Path-based measures on FC networks (used here to measure integration) should never be used. It is mathematically incoherent. For a measure of integration, I would suggest the total correlation, which can be computed from the covariance matrix directly. Other measures of polyadic dependence could also work, but I feel very strongly that the use of path-based metrics on correlation matrices is invalid.

- I feel like there's something of a cart-horse thing happening here. The changes to fMRI activity seen during DoC or psychedelia are fingerprints of clinical differences, but they are not necessarily

causal. The logic of this paper seems to invert cause and effect a bit: if psychedelics cause increased complexity, then increasing complexity should cause emergence from DoC. However, I don't think we can say that the loss of complexity in DoC patient data is the cause of the failure of consciousness (unless you subscribe to IIT, which I emphatically do not). There cliché that when a metric becomes a target, it becomes a bad metric is apropos here. I'd like to see the authors draw a much stronger distinction between cause, effect, and correlation in this context.

Minor comments:

- Where is the CNT acronym introduced? The first time I see it is on line 199 without explanation.

Response to the reviewers of Communication Biology COMMSBIO-24-0669A

We thank the Editor and Reviewers for their insightful and constructive criticism. We have systematically addressed all their concerns, commentaries and suggestions. We have performed additional analyses whenever it was needed to settle the concerns. In doing so, we have included new supplementary material and made substantial changes to the main text of our manuscript. Most of these changes are also reproduced here for the sake of clarity. We hope that our efforts led to a revision deemed suitable for publication.

—----------

Reviewer #1 (Remarks to the Author):

This paper presents a super interesting, rigorous computational modeling approach to investigate how pharmacological agents may impact brain functon(and perhaps consciousness) in patient with disorders of consciousness. Their results are convincing for the most part, and the potential clinical application of this model in designing or identifying drugs that could aid in recovery of consciousness in specific patients is potentially life-changing. Overall, this paper is potentially very impactful and paradigm shifting in the development of neurological treatments. However, I think more attention needs to be paid to individual-level variability (if we are to use this to model individual patients) and the spatial specificity of the receptor maps in the simulated drug models.

We are thankful to the Reviewer for the positive reception of our work and for the useful comments.

1) It is unclear if the SC and FC matrices of the subjects in each group were averaged, and then the model fitted (I am assuming this is the case for the rest of the review). If so, there may be individual-level heterogeneity of damage/injury that is smoothed out. it would be good to replicate the main findings on an individual level model approach - this is relevant clinically if we are to apply this to single subjects who have noisy SC and FC matrices.

We thank the Reviewer for pointing this out. We understand that our explanation about the use of average SC and FC in the model and the rationale behind it could have been suboptimal, and we appreciate the opportunity to clarify it now. Indeed, the SC was obtained as the average of an independent sample with high quality DTI measurements, and the FC was obtained by averaging the correlation matrices from individual participants. We have modified this part of the manuscript as following:

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.

We understand that modelling the individual FC matrices could have been preferable. However, this was not possible due to limitations of the currently available data (see our reply to a concern raised below for a more detailed answer). Therefore, we have edited the revised version of the manuscript to include this point in the list of limitations of our study:

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.

2) Relatedly, it would be interesting to see the individual-level variability in the generated models - when you have noisier data you may have worse fit and more variance in the simulations. perhaps adding in a few examples of generated data from each individual would be interesting to compare against the population level model simulations (fig 3)

We appreciate the Reviewer's suggestion, which is closely related to the first point. Following her advice, we created individual-level models with the same procedure as followed for the group level analysis. As you can see in the Supp Fig 1A-C (included below as Figure R1), the obtained SSIM (i.e. our choice of goodness of fit metric) values from each condition at the group level were lower than the values corresponding to the goodness of fit values obtained from the group level analysis. This implies that the model poorly fitted the data at the level of individual participants, likely due to a combination of higher noise (which is reduced when averaging the FC across participants) and lack of SC corresponding to each individual participant. We modified the manuscript to inform this comparison with the group average version:

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 Supp Fig 1 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.

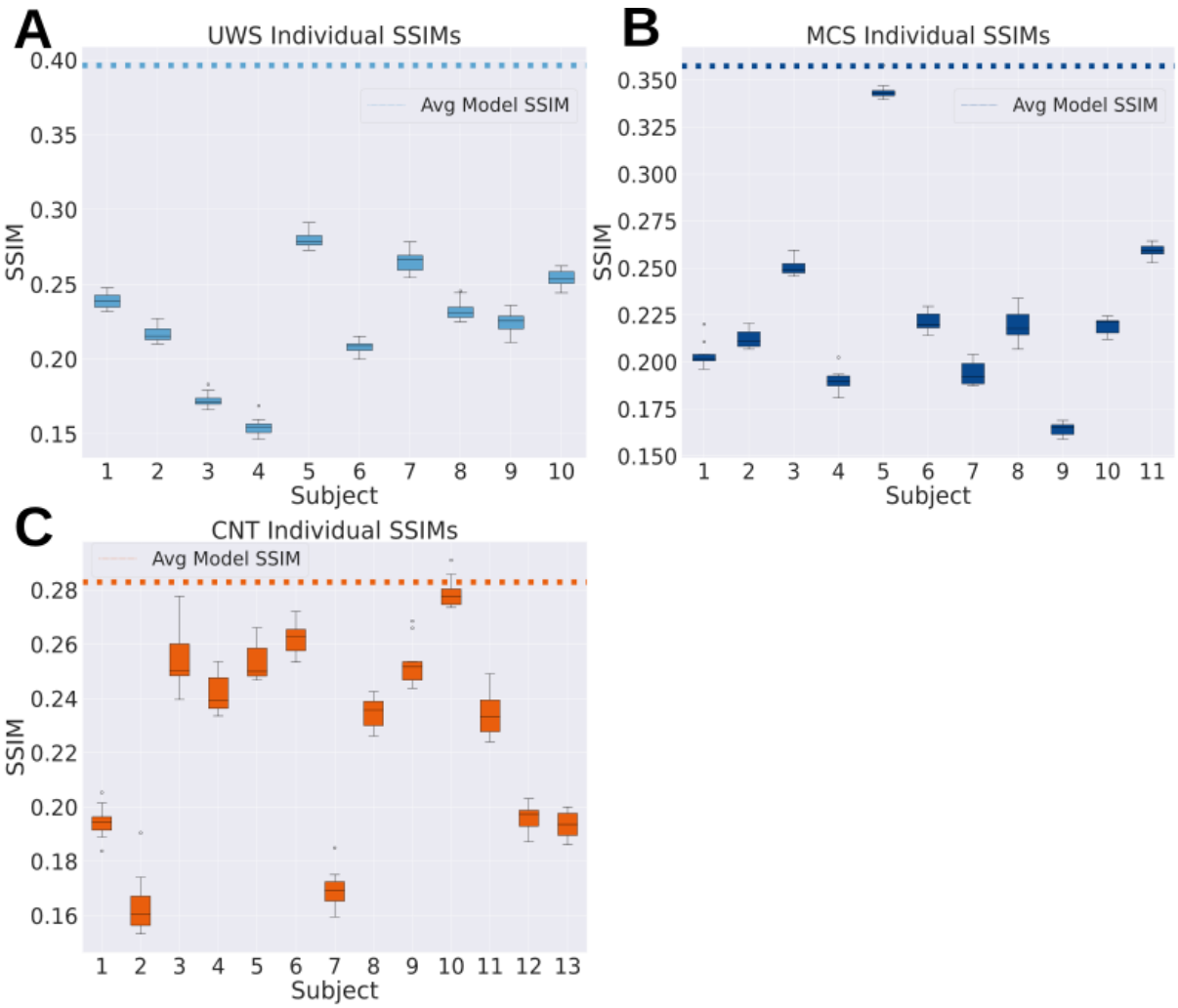


Fig R1. Individual vs. group level model fitting performance **A)** Comparison between the SSIM of each subject-level model vs. the group level model performance for UWS patients. **B,C)** Same as panel A but for MCS and CNT groups, respectively.

3) Since we know what happens in controls who take serotonergic, domanergic and histiminergic drugs, it would be a good validation to apply the receptor simulation studies to the controls as well to make sure the model predicts well that modulation. relatedly, would it not be important to show you can move in the other direction - from conscious control to unconsciousness. perhaps this can be done my applying a simulation of anesthesia to control model.

The Reviewer makes a reasonable inquiry about the validation of the DMF model for simulating pharmacological stimulation. In (1), it is shown that by neuromodulating a model of whole-brain activity corresponding to healthy individuals model through a GABA-ergic map, the resulting dynamics matches those seen in the empirical data corresponding to propofol anaesthesia. Similarly in (2), the same methods are used to fit the DMF model to a cohort of subjects after under acute effects of LSD. Because we do not possess data from our subjects in these altered states we cannot perform a similar fitting, however, we have added the results obtained after neuromodulating the t model fitted to healthy participants at the 5HT2A receptor map (Supp Fig 2 and Fig R2 of this document). It is interesting to observe that these perturbations break away from the "awareness" axis along which the other trajectories moved, starting from DoC conditions. We have included the following explanation of these results the introduction section of the revised manuscript:

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 (71) and the psychedelic state elicited by LSD (51).

Concerning the perturbation of the model fitted to healthy participants, we added the following paragraph:

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 of subjects undergoing pharmacological interventions, we simulated 5HT2A activation in healthy individuals, observing a slight deviation from the "wakeful" axis (Supp Fig 2) when introducing this perturbation. This suggests that psychedelics may exacerbate some of the defining characteristics of the conscious wakeful state.

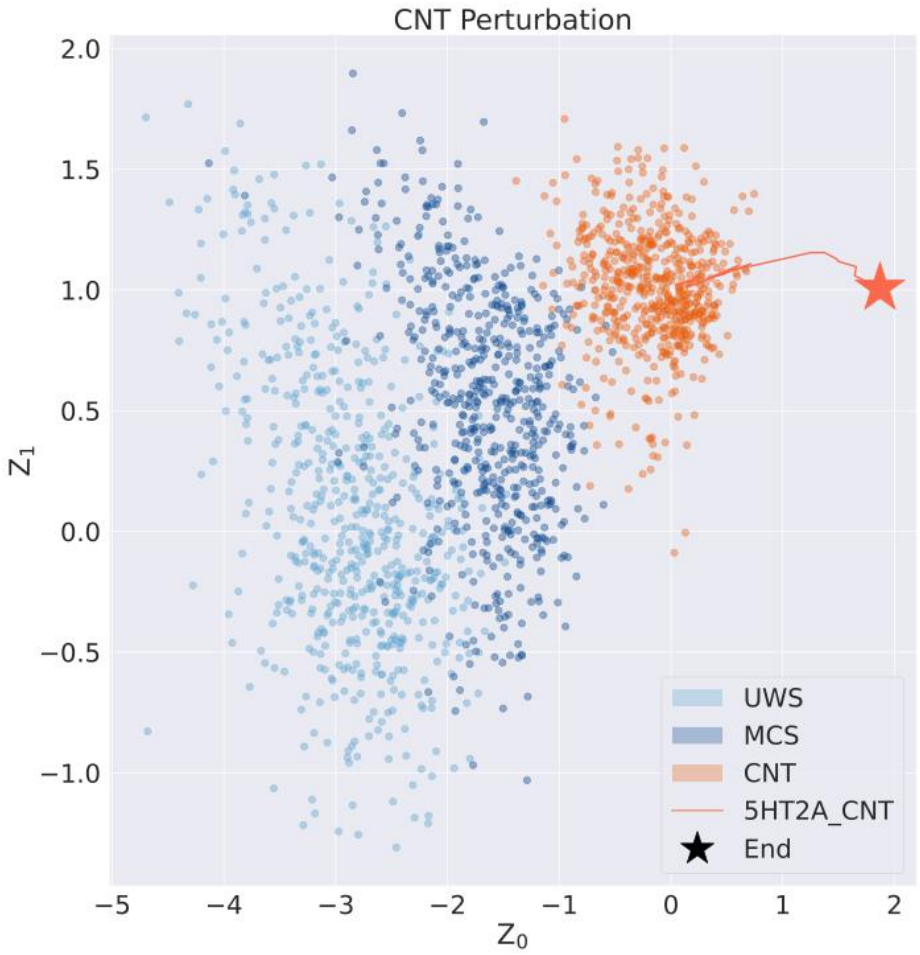


Fig R2. Modelling the effects in brain dynamics during the psychedelic state by modulating the serotonergic receptors parametrized by the scaling parameter. Interestingly, this intervention is represented in the 2D latent space as a trajectory that moves towards the direction opposite to that of reduced states of consciousness.

4) The paper makes the claim that "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 density throughout the brain." but this is not what they have tested with a lack of correlation between distance from start and spatial density of SC. Rather, they should test if a random spin-null spatial pattern of receptor maps (spin test) with receptor density held constant would also result in a trajectory along the same line. This would more fully test if (for example) the 5HT2A map is (or is not) optimally organized to maximize distance from start (similar to what was found in simulation studies via network control theory in Singleton et al. 2022, Nature Communications).

The Reviewer makes a reasonable request to add an adequate null model of the receptor maps to confirm the hypothesis of the network topology not being the driving factor of our observed effect. We simulated the neuromodulation using a randomised version of the 5HT2A receptor map but with the same mean density and spatial autocorrelation, and showed that the same similarity to the healthy

model is reached (Fig R3). We have included the following explanation of these results in the main manuscript text:

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, Figs 5C-D demonstrate 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 (72) and the mean density. We then perturbed our patient models with this null version of the map. In Supp Fig 4 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.

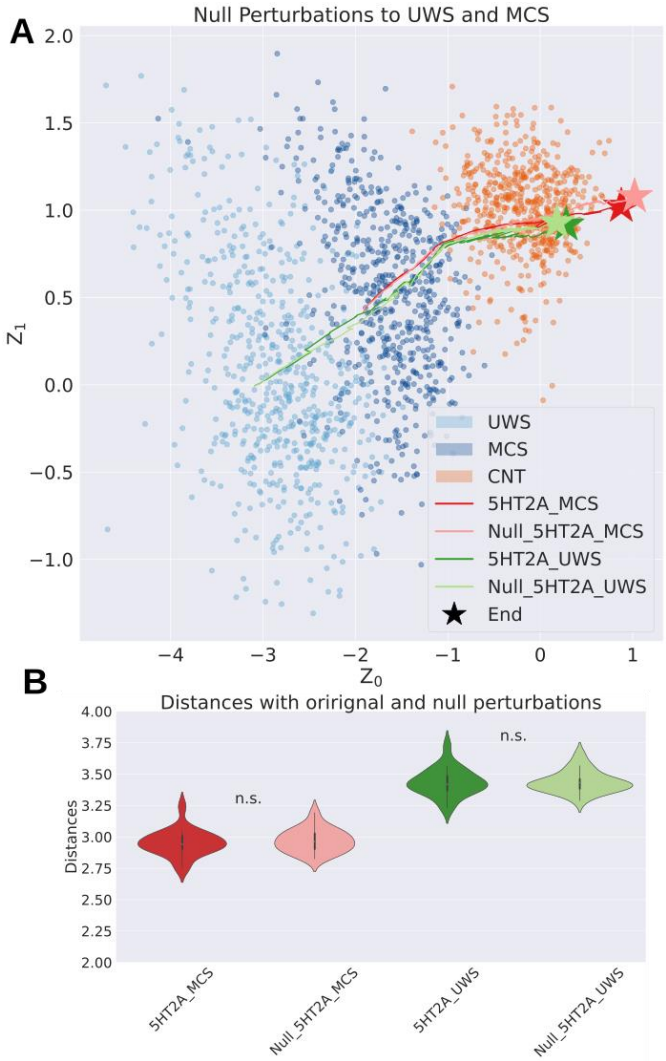


Fig R3. Controlling the effect of the neuromodulation on DoC patients by replicating the results using null maps conserving the autocorrelation, mean and standard deviation of original receptor maps.

5) It may not be that the trajectories are "forbidden" as this simulation is only done using one SC and one FC - perhaps the individual variability in the SC and FC result in drastically varied predicted trajectories (see previous comment)

The reviewer has raised a good point regarding the differences that could be observed at individual level approach. Unfortunately, considering the results presented already shown in Fig R1, we are not confident that the models fitted to individual data are a sufficiently accurate representation of the empirical FC, thus their low dimensional representation and the introduced perturbations would be very difficult to interpret. As we previously stated, we fully agree with the Reviewer that an individual assessment is relevant and could constitute a natural follow up for our study, however, a roadblock exists in the lack of individual high quality data to guarantee a good fit at that level. Two possible directions to overcome this obstacle in future research are to either improve the model, e.g. implement an optimization of the effective connectivity, as in (23,24), or obtain a cohort with high quality DTI to obtain individual level structural connectivity.

6) The pertubation scaling also appears to exactly reflect receptor density - again, this analysis may benefit from null spin-test modeling where the receptor density is held identical and the regional receptor maps are spun to obtain a better measure of how the spatial organization of the actual receptors is impacting the model. it is important to show that these results are not just all reflecting average receptor density (and if they are, then that is important information as well!)

Similarly we address this point in the supplementary analysis for point 4.

—----------

Reviewer #2 (Remarks to the Author):

This is a very interesting an ambitious paper that sets out to devise computational treatments for disorders of consciousness based on real patients data and the maps of receptors in the human brain, using a sophisticated brain model and deep learning.
I would be enthusiastic about this work, were it not for two main criticisms. First, many method choices need to be more thoroughly justified and compared with the authors' own previous work, especially Deco 2018 Current Biology which is conceptually similar but the authors changed many things without telling the reader why (fitting function, addition of the deep learning VAE). I add some details below.

Second, the authors show that their results are entirely dependent on mean density, and not at all on where the receptor is more or less expressed, since the success of

each PET map at inducing transitions from DoC to control is extremely correlated with the PET map's mean value. This is very unfortunate because mean value of a PET radiotracer map such as these is biologically meaningless. Binding Potential is a relative measurement rather than an absolute quantification of the expression of a receptor. Comparisons between maps with different tracers are not possible. At the moment, the results = do not support the paper's title, because they are really about mean value of a PET tracer, not about identity of the receptor. I provide more detail below. It is possible that I am wrong, in fact I hope so. If so I would welcome an explanation, perhaps I have overlooked some aspect of the modelling framework that removes this concern, but I do not see this addressed. But if I am right, then my only suggestion would be to use some normalization to remove the difference in mean density and variability to make the map really comparable, and repeat the work. For this reason I recommend major revisions.

We appreciate the Reviewer's comments on our work. Following their inquiries and suggestions we have made modifications on the manuscript and performed additional analysis to address their concerns.

1) Some points of lesser importance:
Reference styles are inconsistent at times.

We have found the inconsistencies and we have corrected them

2) Figure 2A: the light blue is hardly visible.

We have chosen darker colours for all baseline models to help visualisation

3)Line 396: 5HT1A/1A should be 2A?

Yes, it has been corrected

3) About the Abstract:
Clinically, DoC are defined in terms of responsiveness.

We corrected the abstract to consider the Reviewer's comment:

Disorders of consciousness (DoC) are a group of neurological conditions characterised by their unresponsiveness

4) About the Introduction:
Please provide additional information and references about Cortically Mediated State, and why MCS is not a "minimally conscious" state? This part was unclear.

We recognize that this section was not clear in the manuscript. In (22) Naccache proposes that due to the problems that can be raised with the use of the word 'conscious' in the MCS, a more suitable interpretation framework for clinical assessment can be reached if it is defined as "Cortically Mediated State". We believe that this point is expendable in the context of our study. Therefore, we have removed the following sentence from the ms:

Even if being in a MCS does not necessarily correspond to being in a "minimally conscious" state, it does certainly correspond to a Cortically Mediated State (CMS).

5) Can the authors provide more details about how zolpidem works, beside simply stating that it is paradoxical?

Yes, we can see the there was a lack of explanation regarding the mechanism of action of Zolpidem. We have added the following to the manuscript:

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 (12), and its effect wears off along with the drug (13)

6) Do any theory of consciousness aside from Carhartt-Haris actually state that psychedelics increase consciousness? Bayne and Carter 2018 Neuroscience of Consciousness state quite clearly that psychedelics do not increase consciousness.

The hypothesis of psychedelics as potential treatments for DoC is only partially supported by the theoretical developments from Carhart-Harris and colleagues. Instead, there are multiple converging reasons, both based on theoretical considerations and experimental results. Psychedelics are known to increase the entropy of brain signals recorded with different modalities (fMRI/EEG/MEG) (3,4) , which is opposite to the effect seen in states of reduced conscious awareness (including DoC patients) (5–7). Moreover, the anatomical distribution of serotonin 5HT2A receptors partially overlaps with fronto-parietal regions (8,9) showing diminished brain activity and metabolism during unconsciousness (10). Importantly, the activation of these receptors increases cell excitability and metabolic activity, also opposing what is seen in Doc patients (11). Psychedelics may also facilitate cross-talk between brain regions, resulting in a breakdown of functional modularity (12) , whereas states of reduced awareness may be characterised by diminished functional integration (13) . While the analysis presented by Bayne and Carter (2018) convincingly argues that psychedelics cannot be seen as "consciousness enhancers", their analysis concerns the possibility of inducing "higher" levels of consciousness relative to the normal state of conscious wakefulness. However, the

possibility of restoring normal levels of consciousness with serotonergic psychedelics is not ruled out by their analysis.

7) What is the relevance for DoC of the use of 5HT2A agonists for OCD and depression?

We thank the reviewer for this interesting question. Theoretically, Carhart-Harris and colleagues posited that different neuropsychiatric disorders may be linked to abnormal stabilisation/destabilisation of brain activity **(14)**. In the extreme case, DoC may be characterised by a persistent stabilisation of brain activity, concomitant with loss of complexity and inter-areal coupling defaulting to the structural connectivity backbone of the brain **(15,16)** . However, other disorders may be seen as less extreme cases of the same dynamical phenomenon. Depression and OCD are clinically characterised by recurrent and persisting thought patterns and/or behaviours, which -at the neural level- may be indicative of a reduced functional repertoire as a consequence of stabilised large-scale brain dynamics.

8) To this reviewer's knowledge, the density of 5HT2A is maximal in V1 cortex, which is primary region, the opposite of association areas mentioned by the authors (although it is also high there). How can these facts be reconciled?

We thank the reviewer for this comment. According to the review published by Saulin and colleagues, the highest density of 5HT2A receptors is found in cortical areas, especially medial frontal cortex, temporo-occipital cortex, as well as piriform and entorhinal cortex. The more recent work of Beliveau and colleagues presents a similar cortical distribution of 5HT2A receptors, also including the posterior cingulate cortex, a key node in the fronto-parietal network. While these regions clearly encompass the visual cortex (including V1 and beyond), they are not limited to the occipital cortex and present a significant overlap with frontal and parietal regions implicated in the neural correlates of consciousness **(10)**. This lack of specificity may translate in a similar lack of specificity in terms of the acute subjective effects of psychedelics, featuring strong visual distortions linked to the activation of occipital 5HT2A receptors **(17)**. However, this does not preclude an effect on the level of consciousness in DoC patients, only that the effect would be selective and may also include visual distortions. As was requested by the reviewer below, we included as a supplementary material renders with the brain distribution of each map.

9) The role of the VAE is unclear: why add this extra step to the existing setup of Deco 2018? The explanation for making the procedure even more complicated should be

thoroughly justified. The authors need to show what benefit is obtained from stacking deep learning with the mean field model. The authors acknowledge in the Discussion that "the VAE was not strictly necessary to fit the data and simulate the perturbations, it provided a framework to facilitate the interpretation and visualisation of the results" but for me, it made the process more obscure, adding an extra black box element with unclear benefit.

We thank the Reviewer for noticing this lack of clarity in our methodological decision of incorporating VAE in our study. In past studies from our group, VAE has been proven to accurately distinguish different consciousness states from high dimensional data by projecting it to a representative latent space (18,19). Moreover, this representation shows that these states of reduced consciousness seem to be organised within a manifold in which one direction can be associated with the reduction of the consciousness (18,19). Thus, the advantage of the VAE is to provide a rich visual/geometric interpretation of how different groups of subjects are related to each other, and how they behave when external perturbations are introduced. Given the large dimensionality of the data, this interpretation would be impossible without the use of VAE or other similar dimensionality reduction technique.

One alternative possibility to evaluate the result of pharmacologically perturbing a DoC model would be to perform a statistical analysis comparing the outcome states after the perturbation. . However, this decision would only inform us on whether the stimulation transitioned the dynamics towards wakefulness. Instead, the use of VAE provides a richer characterization. For example, VAE could detect cases where the neuromodulation might take the model to states that are neither of higher/lower levels of consciousness, provoking altered states of consciousness. Another possibility is that stimulating different receptors might induce the same end result, but not with the same efficiency nor trajectory in the latent space This richer analysis can be performed by creating a latent space using VAE, a continuous map where trajectories represent many different possible brain states. Moreover, given the generative nature of the VAE, each position in the latent space can be characterised in terms of relevant metrics, facilitating the visual understanding of how dynamics are modified as the perturbation intensity is ramped up in the model. This representation allows us to quantify two-dimensional geometry, the effect and efficiency and other details of different interventions. As the Reviewer acutely mentions in point 15 it is interesting that the trajectories follow a straight path towards the healthy state region, suggesting a possible mechanism of recovery. We have made the following adjustment in the introduction to better explain this point:

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 (47), endowing each point in latent space with the value of relevant metrics, namely entropy, mean FC value, and correlation to structure (37,38,48)

And also in the discussion:

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 (30,31,47). 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.

About the Results

10) Deco 2018 used a different procedure for fitting the same model, showing that FC was a poor choice. Why go back to a provably inferior fitting target?

We appreciate the reviewer's comment. It is important to note that in Deco (2018), the authors used the FCD to fit the model mainly for two reasons. On the one hand, they wanted to evaluate the dynamical behaviour of the fMRI signals, and for this purpose, the FCD is a suitable dynamical measure. On the other hand, the FCD fitting in terms of the Kolmogorov-Smirnov distance between the empirical and simulated FCD distributions provides a narrow minimum when exploring G. Conversely, when they evaluated the FC in terms of correlation, they obtained a plateau instead of a narrow maximum, and thus it is not a good metric to determine the optimal working point of the model. Importantly, as demonstrated by the work of Piccinini et al, (20), models are not necessarily good at fitting multiple observables and metrics at the same time. Consequently, in our work we decided to fit the FC, which is the most straightforward choice considering that it is the main measure used to investigate resting state fMRI data and also to evaluate the performance of the models in the literature both before and after Deco's 2018 work. To do so, we evaluated the goodness of fit of the model by computing the SSIM instead of the correlation, which crucially provides a clear maximum that allows us to determine the optimal working point of each model.

11) Figure 2B is puzzling. It appears that simulations based on CNT connectome match the CNT FC better than they match the MCS or UWS condition, as the authors state. However, the simulations based on MCS and UWS connectomes are both better at matching the controls' FC, than the simulation based on CNT's own connectome. So,

simulation from patient data are the ones that best match control data? This casts doubt on the model quality for CNT.

We thank the Reviewer for noticing this issue. We agree that panel B of this figure was misleading. We have changed it so now it can be seen that the optimal G for simulating the CNT condition shows a better fitting than the optimal G for the other two, and the same for MCS and UWS models. Importantly, the brain activity of DoC patients is characterised by a lack of capacity to deviate from the regime dictated by the structural connectome. This is why DoC models show a better fit compared to the CNT brain activity, which is more complex and thus more challenging for the model to reproduce. At the same time, our work is based on simulating interventions on DoC models and CNT models are used as a target, and to address whether that interventions approach CNT condition the relevant fact is that baseline models are specific to each state, i.e., they are distinguishable between themselves in terms of their associated G value and the FC matrices they can produce. Because we observe the neuromodulatory effect in the latent space we used Support Vector Machines (SVM) in the projected baseline FC matrices to show that the resulting clusters were distinguishable. The data from three latent clusters (UWS, MCS, CNT) was concatenated to form the feature matrix and corresponding labels. The dataset was split into training and testing sets using an 80-20 split. The SVM model from the scikit-learn library was used with a default radial basis function (RBF) kernel and was trained on the training set. The model's performance was evaluated on the test set using precision, recall, and F1 score, as provided by scikit-learn. The SVM's parameters were chosen based on default settings without additional hyperparameter tuning. We made the following modification in the manuscript:

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 Table 1)

And added the following supplementary table with the classification performance of the SVM:

	Precision	Recall	F1-Score
UWS	0.9921	0.9921	0.9921
MCS	0.9523	0.9375	0.9448
CNT	0.9158	0.9333	0.9245

Table R1: Performance of a SVM classifier trained to distinguish between the three conditions projected in the 2D latent space.

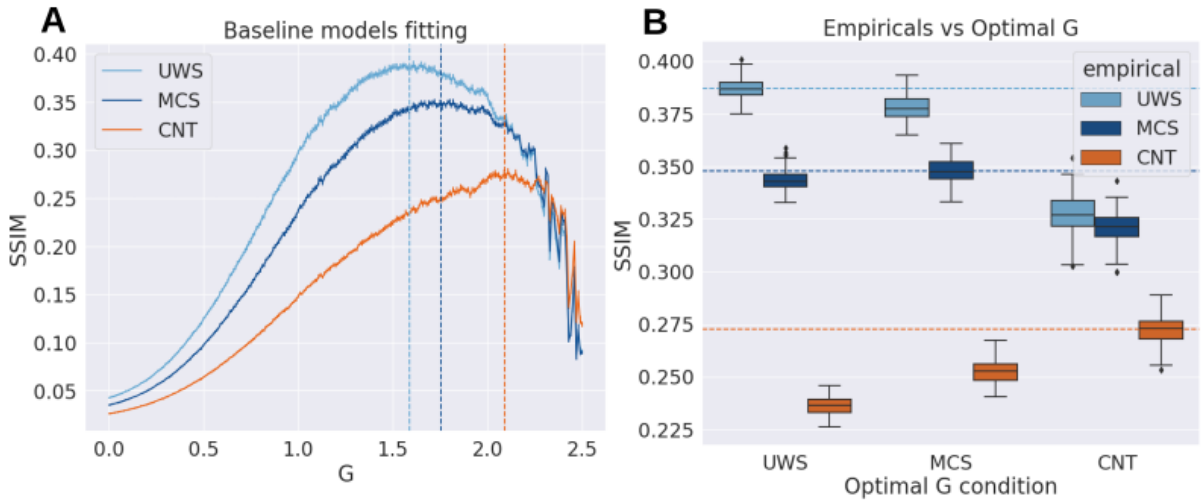


Fig R4. Comparison between the fitting level by each condition. (improvement of Fig 2 of the main ms). **A)** Level of fitting for each condition as a function of G. **B)** Comparison of each model working point. We identified that for each condition each model is statistical different from the the model of other conditions

12) VAE: why need two dimensions, if the model only has one free dimension (the global coupling)?

We thank the Reviewer's question. The model optimisation procedure includes the exploration of the global coupling parameter (G), and in each case the Feedback Inhibitory Control (FIC) is also optimised. The FIC also depends on the level of the simulated neuromodulation, which is controlled by the scaling parameter changing the impact of the regional heterogeneity given by the 90-regions density receptor map. In this sense, to properly represent perturbations on the latent space more than one dimension is needed.

13) LZ: what is the meaning of LZ complexity of a FC matrix (and why do the authors talk of "pixel" in the Methods)? The entropic theory and papers based on it talk about LZ of brain signals (so in time), not of connectivity. I believe that LZ of a matrix is not invariant to reordering of the rows and columns (for example, from left-right-left-right to left hemisphere then right hemisphere). But the information that the matrix conveys about the brain remains the same. Can the authors explain their choice of LZ for an FC matrix, and demonstrate that the result is the same if regions in the matrix are reordered? An ordering-dependent result would seem meaningless. Also, the DMF can produce BOLD signals. Why not use the LZ of the BOLD signals?

The Reviewer makes a valid point regarding what may be considered an unconventional metric to apply on FC matrices. We do not calculate LZ of the BOLD signals because our goal is to use a complexity metric on the FC matrices generated by sampling in our latent space. Also, our model is not trained to optimise the reproduction of time series directly, but to capture their FC instead. Since the LZC may depend on the ordering of the matrix therefore we have decided to replace it by Shannon Entropy of the flattened sampled FC matrix. We have changed this in the figure 3 of the manuscript which is also reproduced in this document as Fig R5

14) Measure of integration: what is the meaning of shortest paths on a functional matrix? Links in a functional matrix are correlations, so what is the meaning of a path made of correlations? Every region has a correlation with every other region. I know that this is often done in the literature, but it does not make it any more meaningful.

The Reviewer raises a valid concern regarding the use of average shortest path length as a proxy for integration. We have changed our analysis to use the mean of the FC matrix, which is a much more straightforward and interpretable measure of network integration. In (19) they show that changes in this value are not homogeneous but rather occur in inter-modular nodes. We have therefore changed Fig 3 and the manuscript accordingly and it is reproduced below:

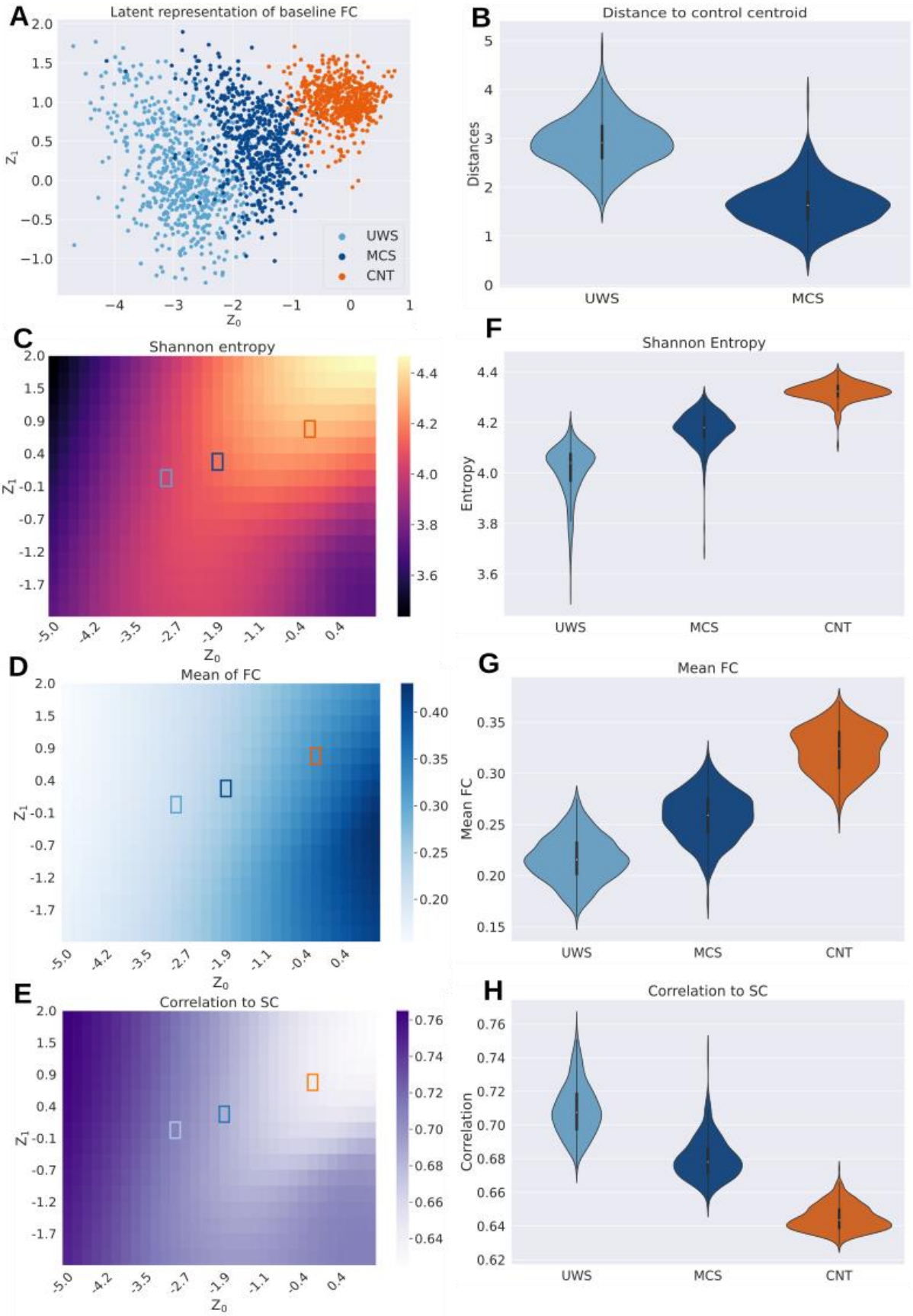


Fig R5. The current version of Figure 3 in the main manuscript where we computed the shannon entropy and the mean FC instead of the LZC and the integration path respectively in each FC generated by decoding a grid in the latent space

15) Pharmacology. It is nice that the UWS trajectory goes through the MCS space before getting to control. But how do the authors explain that some receptors induce much greater progress but all are along the same trajectory? This seems a bit strange. If I am correct the pharmacology as implemented is equivalent to making the model progressively more excitable. This is opposite of the reduction in G parameter. Are these two just being played against each other? What would be the trajectory, if instead of the pharmacological manipulation, the authors simply increased the global coupling in the UWS or MCS model?

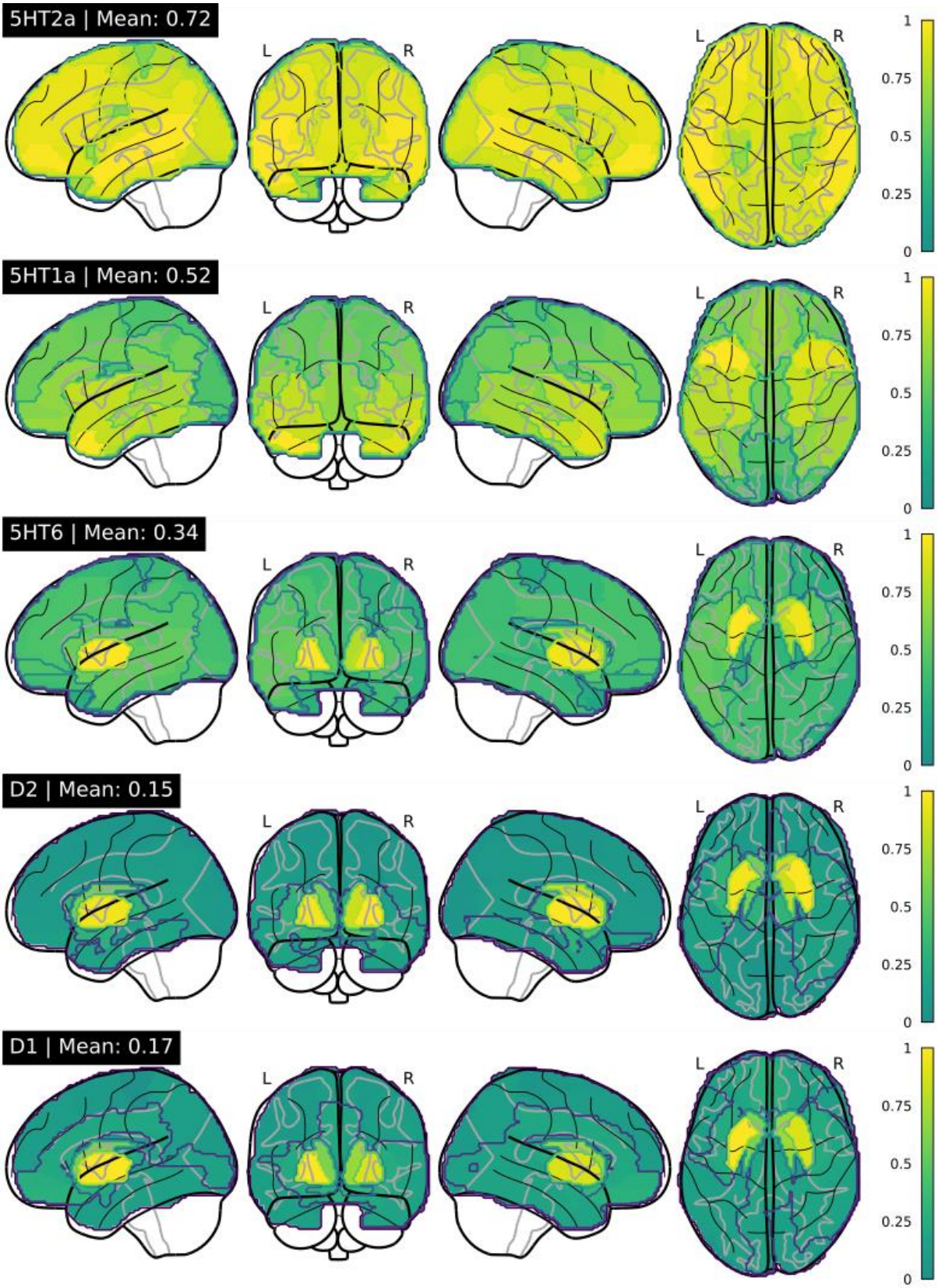
The Reviewer proposes an interesting debate regarding the trajectories of neuromodulation in the perturbational landscape. The structure of the FC data that we generate to train the VAE and to model the pharmacological interventions is determined by the underlying structural connectivity. Also, as we stated in the manuscript, one of the limitations of our work is the fact that we used only one structural connectome for all conditions and perturbations. Consequently, the regions of the latent space that the conditions and trajectories spread are constrained to certain areas. This is also expected considering that the effects of different neurotransmitters on whole-brain dynamics are represented through the same mechanisms, namely modulating the excitatory and inhibitory response function.

Regarding the second question, although we can increase the global coupling in our model, biologically this is not possible in a real experimental setting. What we can do is stimulate different receptor systems to induce changes. This is the reason for choosing a biophysically realistic model for our simulations. Although these are two different mechanisms, we are observing the same effect induced through perturbations on the biologically interpretable parameter. Our result indicates that some receptors can easily replicate the global coupling increase while others cannot.

16) Can the authors show brain maps of each receptor map on the brain? It is known that some tracers have very different PET intensity in cortex and subcortex. Readers should be able to see the maps for themselves.

For transparency we are adding the glass brain plots of the receptor density maps obtained from (21) in Supp Fig 4 and displayed below (Fig R6)

.


5HT2a | Mean: 0.72
5HT1a | Mean: 0.52
5HT6 | Mean: 0.34
D2 | Mean: 0.15
D1 | Mean: 0.17
L
R
1
0.75
0.5
0.25
0

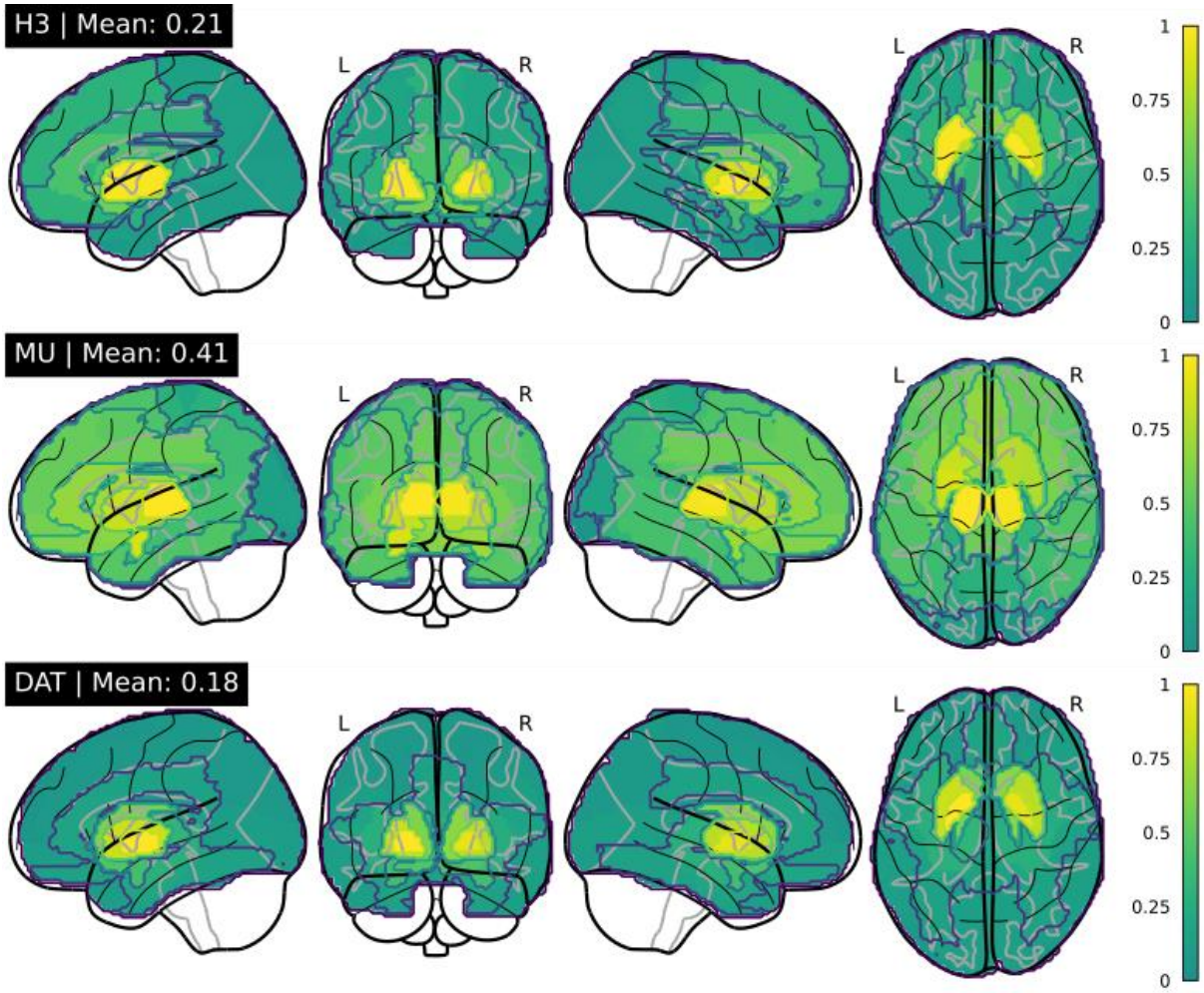


Fig R6. Brain representation of the spatial distribution of the normalised receptor maps used in this work.

17) I am not familiar with the structural divergence. Could the authors provide scatter plots of each PET map against structural node strength, and the corresponding correlation between the two? This is a more easily interpretable measure, and the visual demonstration would be very helpful for the reader.

We understand the Reviewer's request to better explain our chosen metric. We were interested in relating the receptor density to the node strength to assess if high density areas coincided with highly connected areas. Given that the density and node strength did not present significant correlations, we simply calculated the euclidean distance between the two vectors given that this gives the summed differences between density and connectivity. Because this number increases for receptor density maps that are very different from structural connectivity strength, we named this metric 'structural divergence'. To make this point more clear, we modified the manuscript accordingly:

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.

18) It is very unfortunate that the authors found that the topology of the stimulated nodes does not matter: what matters is the mean density ("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 density throughout the brain"). The authors do not seem to realise it, but this is a major problem for their entire study. Of course a map with larger values will show bigger effects under the authors' implementation (could the authors confirm this?). Unfortunately, absolute density of a PET map is not biologically meaningful because PET measurements (binding potential) are relative, and do not measure absolute expression of a receptor in a region. Comparisons between regions are still possible for each individual PET tracer, but the authors show that regional differences do not matter. However, comparisons of density between different tracers (so, different maps), are meaningless.

We understand the reviewer's critique regarding this important aspect of our work. We agree that the absolute density of a PET map is dependent on the details of the PET measurement and thus that it does not have a direct biological interpretation. Moreover, this aspect of the PET data renders direct comparisons between receptor density maps meaningless. To avoid this issue, we normalised each receptor map separately, as can be seen in Fig. R6. In that figure, each normalised map presents a different spatial distribution and different means in relative values. In particular, some maps with high mean values present a widespread anatomical distribution of receptors (e.g., 5HT2A), while others with low mean are quite sparse (e.g., D1). We note that our stimulation results depend on the mean of the normalised map, which in turn, considering the normalisation, is related to the presence of the receptor throughout the brain. In summary, our methodological procedure, which includes the receptor map normalisation and then the modulation of the impact of each receptor map with a scaling parameter, indicates that our results are not driven by the original mean intensity of the PET receptor maps.

This was not expressed accurately in the manuscript therefore we clarified in each instance on the manuscript that we are assessing the normalised receptor map and we made the following modifications in the introduction:

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 (27), 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 Supp Fig 3) to account for their overall presence in the brain, independent of the absolute PET values.

And in the discussion:

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

Additionally, to emphasise this we created spin-null maps that preserved spatial autocorrelation and the mean density of the normalised receptor, but destroyed the topological relationship with brain connectivity (shown in Fig R7, and also included as supplementary figure of the manuscript):

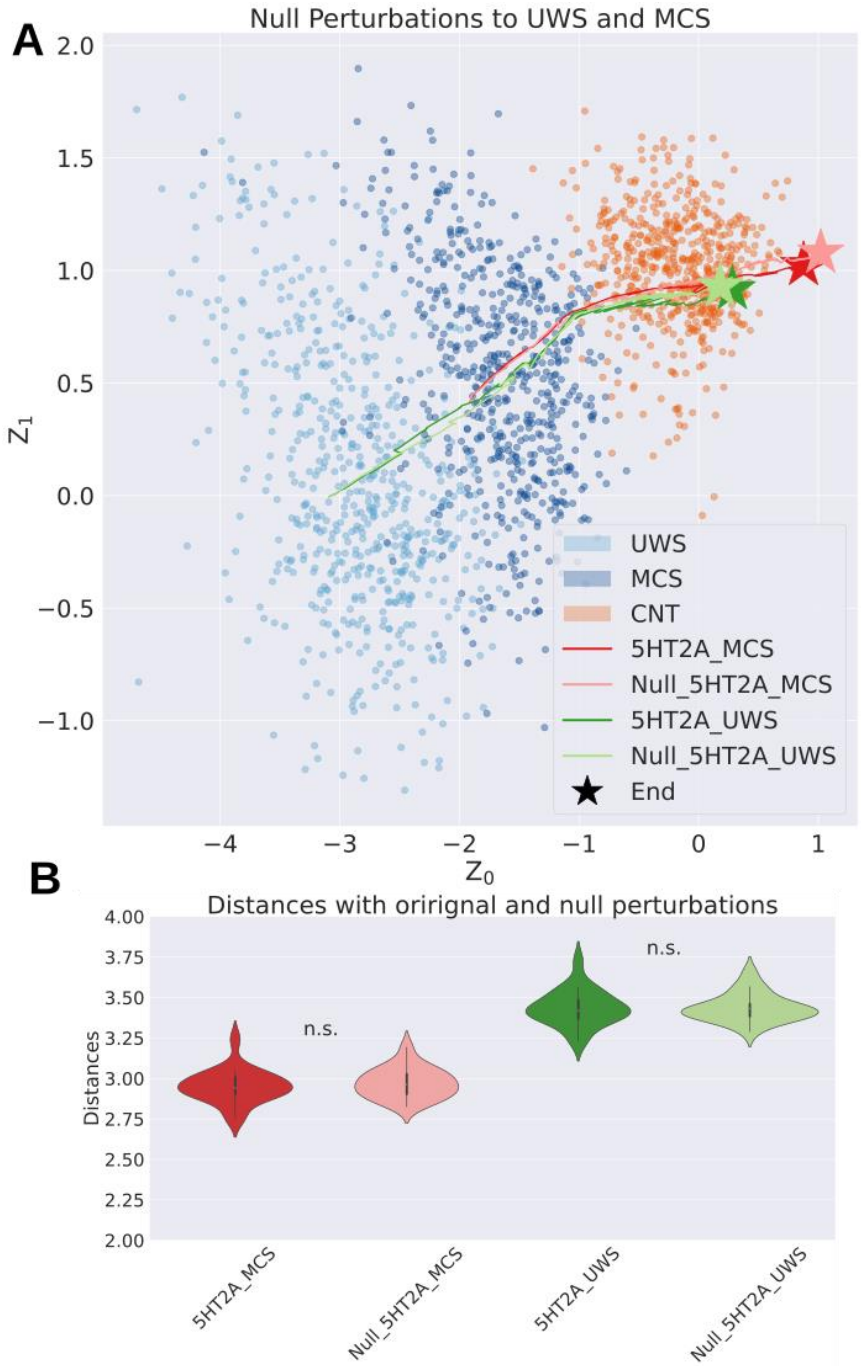


Fig R7. Controlling the effect of the neuromodulation on DoC patients by replicating the results using null maps conserving the autocorrelation, mean and standard deviation of original receptor maps.

19) Given differences in mean density between maps, I do not know how to interpret the results about "perturbational efficiency". However, this section is also not very well explained, I think the reader should be given more explanation (why a quadratic regression, for example?).

We thank the Reviewer for pointing out the lack of clarity regarding perturbational efficiency. We have thus modified the manuscript accordingly:

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.

About the Discussion

20) This section is almost entirely based on the assumption that the results have a meaningful biological interpretation in terms of receptor identity, which unfortunately they do not if they just depend on mean map density. For example, the interpretation of dopamine receptors as least effective (and attempt to justify this) is not appropriate: those are just the maps with least density. As far as I can tell, it is just happenstance that the 5HT2A map, the one that the authors were more interested in, is also the one with the greatest mean density.

We appreciate the Reviewer's comment regarding the interpretation of our results. Their concern regarding the absolute density of a PET map is valid which is why we changed the manuscript accordingly to emphasise that the maps were normalised before perturbing. The modified snippets are shown in points 18, 21 and additionally the supplementary figure stated here as R6.

21) The Limitations section is appropriate, but omits to discuss that dependence on the mean value of the PET map is a critical problem that makes the results uninterpretable from a biological point.

We added in the limitation section the following sentence regarding the limitation of PET measures:

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.
—----------

Reviewer #3 (Remarks to the Author):

There has recently been discussion about the possibility that serotonergic psychedelics could be useful for restoring awareness of patients with disorders of consciousness. This hypothesis is based on the well-replicated finding that psychedelics push the brain towards a higher-entropy dynamical regime (as measured with complexity metrics like the Lempel-Ziv complexity), while patients with DoC typically have lower-entropy dynamics. Given the obvious ethical issues around giving brain-damaged patients mind-altering drugs, in this paper Mindlin et al., take a computational approach: by fitting in silico models of brain activity to data from patients with DoC, they can explore the effects of perturbing different receptors. They find that activation of serotonin receptors (as would be done by a psychedelic) creates patterns in the DoC brain that are more similar to normal, conscious brains.

This is a fascinating piece of science, and one that I would be happy to see added to the literature. I do have some conceptual/technical questions I would like the authors to respond to, and demand a small change to the network analysis, but by and large, I think that this is a fine paper

We thank the Reviewer's enthusiasm with this study and appreciate their constructive critique.

1) It seems like the use of fMRI-based data is a very serious confound here, since fMRI data is based on a vascular signal. Serotonergic psychedelic drugs are powerful cranial vasoconstrictors (a fact that is almost never addressed in the literature), and so it is unclear to me how many of the results of fMRI-based analyses of psychedelics (which are cited as justification for this study) are compromised by changes to vascular tone following something like LSD or psilocybin. Similarly, the global atrophy seen in DoC brains also has vascular impacts. I do not think we can assume that the hemodynamic response function is identical in normal consciousness, psychedelia, and DoC. This seems like a project that would be much better suited for use with EEG or MEG data, rather than fMRI data.

We thank the reviewer for raising this important point. Indeed, it is known that serotonergic psychedelics (as well as several other psychoactive drugs) can alter cerebral blood flow, as they exert an effect in brain vasculature. Unfortunately, this problem is transversal to all neuroimaging studies of serotonergic psychedelics. However, this problem is not manifest in our study, as we did not fit our model to

fMRI data measured under the effects of psychedelics, nor incorporated such recordings in any way in our study. Instead, we simulated the effect of these drugs on brain activity in terms of the well-known and characterised mechanisms of action of these compounds (serotonin 5HT2A receptor activation), applied to fMRI of DoC patients. Additionally, the Balloon-Windkessel hemodynamic response model that we are using is fixed. But we believe it would be relevant for DoC whole-brain modelling to study the effect of global atrophy for BOLD simulation:

Another unexplored avenue of research is the parametrization of the hemodynamic response model to take into consideration the variations DoC brain vasculature that would affect BOLD signals.

2) Path-based measures on FC networks (used here to measure integration) should never be used. It is mathematically incoherent. For a measure of integration, I would suggest the total correlation, which can be computed from the covariance matrix directly. Other measures of polyadic dependence could also work, but I feel very strongly that the use of path-based metrics on correlation matrices is invalid.

We understand the Reviewer's concern regarding the use of Path-based measures on FC matrices. To address this issue we have changed this metric to the mean of the FC matrix. We find this a more direct measure given that in (19) they show that changes in this value are not homogeneous but rather occur in inter-modular nodes. We have therefore changed Fig 3, here as Fig R8:

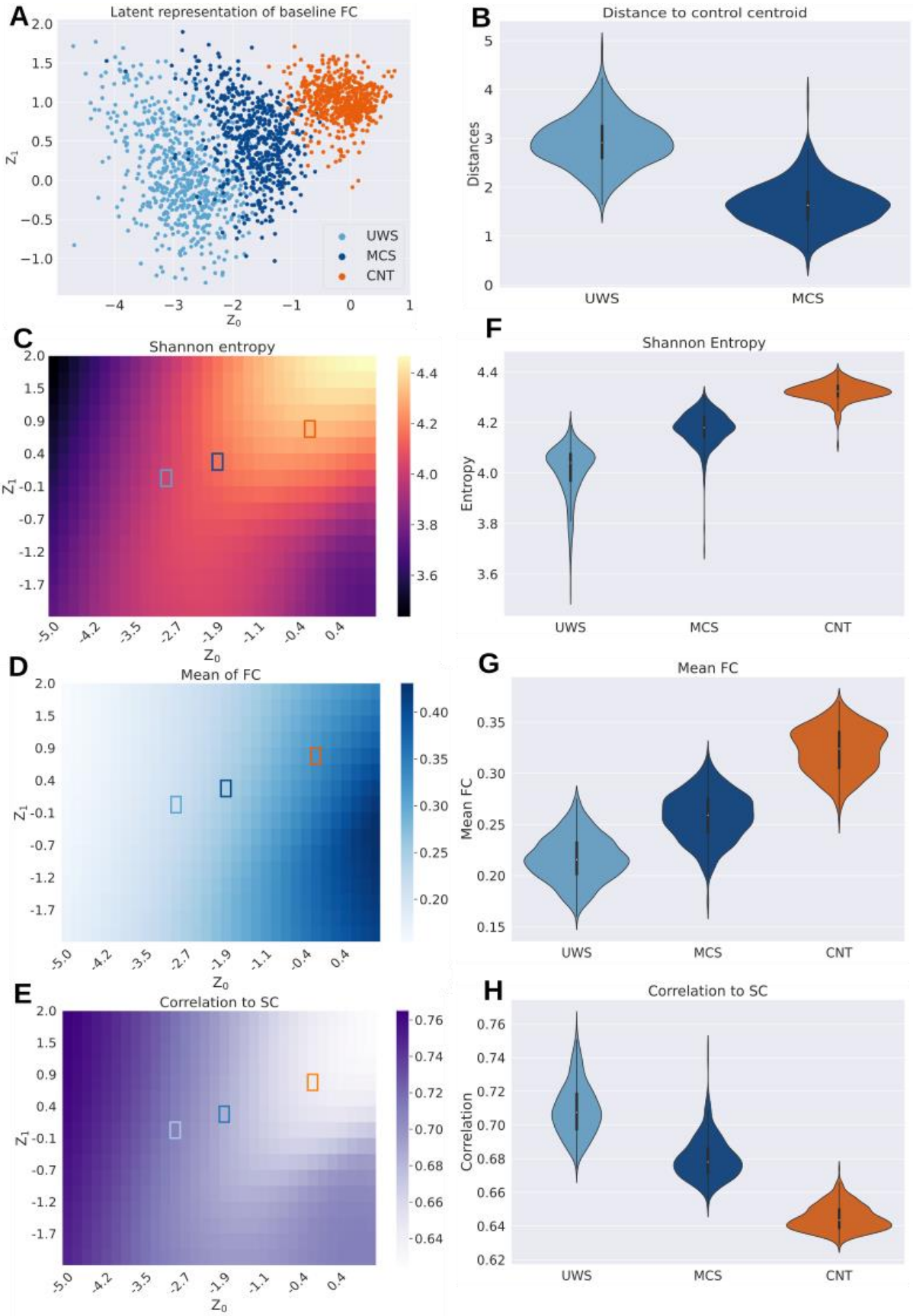


Fig R8. The current version of Figure 3 in the main manuscript where we computed the shannon entropy and the mean FC instead of the LZC and the integration path respectively in each FC generated by decoding a grid in the latent space

3) I feel like there's something of a cart-horse thing happening here. The changes to fMRI activity seen during DoC or psychedelia are fingerprints of clinical differences, but they are not necessarily causal. The logic of this paper seems to invert cause and effect a bit: if psychedelics cause increased complexity, then increasing complexity should cause emergence from DoC. However, I don't think we can say that the loss of complexity in DoC patient data is the cause of the failure of consciousness (unless you subscribe to IIT, which I emphatically do not). There cliché that when a metric becomes a target, it becomes a bad metric is apropos here. I'd like to see the authors draw a much stronger distinction between cause, effect, and correlation in this context.

We appreciate this thoughtful comment made by the reviewer. Indeed, there is not sufficient evidence supporting that complexity (as captured using fMRI measures) has a causal role to determine the level of consciousness. However, there isn't firm evidence supporting the contrary. Thus, we do not believe it can be affirmed that our analysis is based on an erroneous assumption of causality, since the causal relationship between neuroimaging markers and levels of consciousness is not firmly established. Ultimately, this cannot be addressed by modelling studies alone, but also requires new experiments. In this context, we believe that our work is helpful as a proof of plausibility , showing that psychedelics can shift the complexity of fMRI markers in the direction of healthy individuals - whether this entails the recovery of consciousness or not must be investigated empirically.

However, we agree with the reviewer on the need to distinguish between correlation and causality in the context of the hypotheses of our work. We have added the following paragraph to the discussion to provide such distinction.

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. In particular, our work supports the plausibility of increasing these markers with a concrete intervention, i.e. administration of serotonergic psychedelics.

Minor comments:

4) Where is the CNT acronym introduced? The first time I see it is on line 199 without explanation.

The acronym was not introduced before. We thank the reviewer for this correction

References

1. Luppi AI, Mediano PAM, Rosas FE, Allanson J, Pickard JD, Williams GB, et al. Whole-brain modelling identifies distinct but convergent paths to unconsciousness in anaesthesia and disorders of consciousness. Commun Biol. 2022 Apr 20;5(1):1–15.
2. Deco G, Cruzat J, Cabral J, Knudsen GM, Carhart-Harris RL, Whybrow PC, et al. Whole-Brain Multimodal Neuroimaging Model Using Serotonin Receptor Maps Explains Non-linear Functional Effects of LSD. Curr Biol. 2018 Oct 8;28(19):3065-3074.e6.
3. Herzog R, Mediano PAM, Rosas FE, Lodder P, Carhart-Harris R, Perl YS, et al. A whole-brain model of the neural entropy increase elicited by psychedelic drugs. Sci Rep. 2023 Apr 17;13(1):6244.
4. Schartner MM, Carhart-Harris RL, Barrett AB, Seth AK, Muthukumaraswamy SD. Increased spontaneous MEG signal diversity for psychoactive doses of ketamine, LSD and psilocybin. Sci Rep. 2017 Apr 19;7(1):46421.
5. Schartner MM, Pigorini A, Gibbs SA, Arnulfo G, Sarasso S, Barnett L, et al. Global and local complexity of intracranial EEG decreases during NREM sleep. Neurosci Conscious. 2017 Jan 1;2017(1):niw022.
6. Schartner M, Seth A, Noirhomme Q, Boly M, Bruno MA, Laureys S, et al. Complexity of Multi-Dimensional Spontaneous EEG Decreases during Propofol Induced General Anaesthesia. PLOS ONE. 2015 Aug 7;10(8):e0133532.
7. Visani E, Luria G, Sattin D, Rossi Sebastiano D, Ferraro S, Panzica F, et al. Entropy Metrics Correlating with Higher Residual Functioning in Patients with Chronic Disorders of Consciousness. Brain Sci. 2022 Mar 1;12(3):332.
8. Saulin A, Savli M, Lanzenberger R. Serotonin and molecular neuroimaging in humans using PET. Amino Acids. 2012 Jun;42(6):2039–57.
9. Beliveau V, Ganz M, Feng L, Ozenne B, Højgaard L, Fisher PM, et al. A High-Resolution In Vivo Atlas of the Human Brain's Serotonin System. J Neurosci. 2017 Jan 4;37(1):120–8.
10. Rees G, Kreiman G, Koch C. Neural correlates of consciousness in humans. Nat Rev Neurosci. 2002 Apr;3(4):261–70.
11. Carhart-Harris R, Kaelen M, Nutt D. How do hallucinogens work on the brain? The Psychologist. 2014 Sep 1;27:662–5.
12. Tagliazucchi E, Roseman L, Kaelen M, Orban C, Muthukumaraswamy SD, Murphy K, et al. Increased Global Functional Connectivity Correlates with LSD-Induced Ego Dissolution. Curr Biol CB. 2016 Apr 25;26(8):1043–50.

13. Boly M, Perlbarg V, Marrelec G, Schabus M, Laureys S, Doyon J, et al. Hierarchical clustering of brain activity during human nonrapid eye movement sleep. Proc Natl Acad Sci U S A. 2012 Apr 10;109(15):5856–61.
14. Carhart-Harris RL, Leech R, Hellyer PJ, Shanahan M, Feilding A, Tagliazucchi E, et al. The entropic brain: a theory of conscious states informed by neuroimaging research with psychedelic drugs. Front Hum Neurosci. 2014;8:20.
15. Demertzi A, Tagliazucchi E, Dehaene S, Deco G, Barttfeld P, Raimondo F, et al. Human consciousness is supported by dynamic complex patterns of brain signal coordination. Sci Adv. 2019 Feb 6;5(2):eaat7603.
16. Casali AG, Gosseries O, Rosanova M, Boly M, Sarasso S, Casali KR, et al. A theoretically based index of consciousness independent of sensory processing and behavior. Sci Transl Med. 2013 Aug 14;5(198):198ra105.
17. Kometer M, Schmidt A, Jäncke L, Vollenweider FX. Activation of serotonin 2A receptors underlies the psilocybin-induced effects on α oscillations, N170 visual-evoked potentials, and visual hallucinations. J Neurosci Off J Soc Neurosci. 2013 Jun 19;33(25):10544–51.
18. Perl YS, Bocaccio H, Pérez-Ipiña I, Zamberlán F, Piccinini J, Laufs H, et al. Generative Embeddings of Brain Collective Dynamics Using Variational Autoencoders. Phys Rev Lett. 2020 Dec 2;125(23):238101.
19. Perl YS, Pallavicini C, Piccinini J, Demertzi A, Bonhomme V, Martial C, et al. Low-dimensional organization of global brain states of reduced consciousness. Cell Rep. 2023 May 30;42(5):112491.
20. Piccinini J, Ipiñna IP, Laufs H, Kringelbach M, Deco G, Sanz Perl Y, et al. Noise-driven multistability vs deterministic chaos in phenomenological semi-empirical models of whole-brain activity. Chaos Interdiscip J Nonlinear Sci. 2021 Feb 17;31(2):023127.
21. Hansen JY, Shafiei G, Markello RD, Smart K, Cox SML, Nørgaard M, et al. Mapping neurotransmitter systems to the structural and functional organization of the human neocortex. Nat Neurosci. 2022 Nov;25(11):1569–81.
22. Naccache L. Minimally conscious state or cortically mediated state?. Brain. 2018 Apr 1;141(4):949-60.
23. Kringelbach ML, Perl YS, Tagliazucchi E, Deco G. Toward naturalistic neuroscience: Mechanisms underlying the flattening of brain hierarchy in movie-watching compared to rest and task. Science Advances. 2023 Jan 13;9(2):eade6049.
24. Deco G, Lynn CW, Sanz Perl Y, Kringelbach ML. Violations of the fluctuation-dissipation theorem reveal distinct nonequilibrium dynamics of brain states. Physical Review E. 2023 Dec;108(6):064410.

Reviewers' comments:

Reviewer #1 (Remarks to the Author):

The authors have done a complete and thorough job of responding to all of the reviewer comments and the resulting manuscript is very solid and thought-provoking work.

Reviewer #2 (Remarks to the Author):

This is an interesting study and the authors have improved their rigor. I thank them for their responsiveness to reviewer comments. However, before I can endorse publication I still have serious reservations about the language, which is not accurately reflecting the results and contents. “Psychedelics” are mentioned prominently in the title and text but in reality what the paper models is serotonin receptor engagement, which is not the same. The authors also do not properly highlight in the Abstract the weakness that their model treats real and random maps as equally successful, so that anatomy does not matter. Overall, this is a worthwhile modelling study but the authors must be more modest and transparent about what the paper allows them to conclude, tempering the language. The methods additions also need to be described better.

Psychedelics are not actually used in the paper itself. The claim about psychedelics is fine as justification in the Intro, and as speculation in the Discussion, but it is not one of the results of the paper itself and should certainly not appear in the title.

That is because 5HT-2a and 1a agonism is what is simulated, but not every 2a or 1a agonist is a psychedelic, so keeping the title as it is, would be misleading. I urge the authors to use language that is a better indicater of what is in the paper.

Also, thank you for adding null controls with spatial autocorrelation. This has improved the rigour. If I understand correctly, the real and null maps perform equally. So the mean is the only thing that matters, and any map with the same distribution and autocorrelation would be just as good: anatomical location of the receptors does not matter at all, it can be completely random. This is another reason why talking about “serotonergic psychedelics” in the title is misleading: One could equally claim “random maps for the treatment of disorders of consciousness is supported by whole-brain modelling”, because the results show this. The Abstract focuses on the success of serotonin-related maps, but forgets to mention that any random map with same statistical properties works indistinguishably well. This must be reported prominently.

Another inaccurate statement, in the Abstract: “This effect was mediated by the brain-wide density of activated receptors”. This is not correct. Authors must make clear to the reader that PET as used here is not a quantification of the absolute value of receptor presence. It is a relative measure, and different PET maps used different regions as reference. For all we know from these PET maps, the actual number of 5T2a receptors in the brain could be orders of magnitude smaller than D1 or D2 in the brain, even if its normalized mean value is the highest. This is not an issue with normalization,

it's an issue with the data. Any claims about "how much receptor there is in the brain" and similar are simply impossible to make from this kind of PET map. The PET maps are a fine resource and good to use, but conclusions must respect what the data quantify and what they do not.

About the maps. The authors state: "we normalised each receptor map separately, as can be seen in Fig. R6. In that figure, each normalised map presents a different spatial distribution and different means in relative values. In particular, some maps with high mean values present a widespread anatomical distribution of receptors (e.g., 5HT2A), while others with low mean are quite sparse (e.g., D1). We note that our stimulation results depend on the mean of the normalised map, which in turn, considering the normalisation, is related to the presence of the receptor throughout the brain. In summary, our methodological procedure, which includes the receptor map normalisation and then the modulation of the impact of each receptor map with a scaling parameter, indicates that our results are not driven by the original mean intensity of the PET receptor maps."
How exactly did the authors perform normalization? The Methods only report that maps were "normalised ensuring that max = 1". There are many ways to obtain max of 1. Was the minimum subtracted, followed by division by the maximum? Or just division by the maximum? Or sigmoid/tangent function rescaling? This can make a large difference, and it will impact whether the authors' claim that "results are not driven by the original mean intensity of the PET receptor maps" is justified or not. Please provide the mathematical formula. I also recommend to show a histogram of the pre-normalization and post-normalization values next to each brain map (the brain maps were already a good addition, thank you).

Abut the maps: D1 and D2 look almost identical. This is a bit surprising. Could the authors confirm that the correct maps are plotted?

Also, looking at the maps it is evident that all except the serotonin maps are dominated by subcortex. Those dominated by subcortex also have lower means. I think this is what the authors mean when they say "some maps with high mean values present a widespread anatomical distribution of receptors". MU and 5HT6 have a bit higher values in the cortex, and their performance is also in the middle. It seems that the comparison at the end of the day is between cortex-dominated and subcortex-dominated maps, which predicts the mean, which predicts performance. So the title should be about cortically-dominanted receptor maps in general, rather than serotonin.

The explanation: "although we can increase the global coupling in our model, biologically this is not possible in a real experimental setting. What we can do is stimulate different receptor systems to induce changes. This is the reason for choosing a biophysically realistic model for our simulations. Although these are two different mechanisms, we are observing the same effect induced through perturbations on the biologically interpretable parameter. Our result indicates that some receptors can easily replicate the global coupling increase while others cannot." This explanation is very useful, and I strongly recommend the authors to include it in the Discussion to clarify that the pharmacology is being used as a way of countering the lower global coupling observed in the DoC patients.

I think if these clarifications are added, and the language is made to properly reflect what the results show, then this will be a valuable addition to the modelling literature.

Reviewer #3 (Remarks to the Author):

I accept the paper.

Response to the reviewers of Communication Biology COMMSBIO-24-0669A

We thank the Editor and Reviewers for comments. We appreciate the recognition of our added analyses and have accepted their suggestions regarding phrasing of our results. The changes on the manuscript also reproduced here for the sake of clarity. We hope that our efforts led to a revision deemed suitable for publication.

Reviewers' comments:

Reviewer #1 (Remarks to the Author):

The authors have done a complete and thorough job of responding to all of the reviewer comments and the resulting manuscript is very solid and thought-provoking work.

We thank the reviewer for their work.

Reviewer #2 (Remarks to the Author):

This is an interesting study and the authors have improved their rigour. I thank them for their responsiveness to reviewer comments. However, before I can endorse publication I still have serious reservations about the language, which is not accurately reflecting the results and contents. “Psychedelics” are mentioned prominently in the title and text but in reality what the paper models is serotonin receptor engagement, which is not the same. The authors also do not properly highlight in the Abstract the weakness that their model treats real and random maps as equally successful, so that anatomy does not matter. Overall, this is a worthwhile modelling study but the authors must be more modest and transparent about what the paper allows them to conclude, tempering the language. The methods additions also need to be described better.

1) Psychedelics are not actually used in the paper itself. The claim about psychedelics is fine as justification in the Intro, and as speculation in the Discussion, but it is not one of the results of the paper itself and should certainly not appear in the title.

That is because 5HT-2a and 1a agonism is what is simulated, but not every 2a or 1a agonist is a psychedelic, so keeping the title as it is, would be misleading. I urge the authors to use language that is a better indicator of what is in the paper.

We have considered the reviewer's comment regarding the prominence of psychedelics in our work. Although our motivation comes from observing the highest effect when using the 5HT-2a receptor map to perturb, we understand that for a full modelisation of psychedelic action more complexity is required. We have made the following modifications to the manuscript, including the title itself:

Whole brain modelling for simulating pharmacological interventions on patients with Disorders of Consciousness

Abstract:

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. As a novelty, 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.

And removed the following statements from the discussion and conclusions:

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. ~~In particular, our work supports the plausibility of increasing these markers with a concrete intervention, i.e. administration of serotonergic psychedelics.~~

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.~~, as is the case of serotonergic psychedelics.~~

2) Also, thank you for adding null controls with spatial autocorrelation. This has improved the rigour. If I understand correctly, the real and null maps perform equally. So the mean is the only thing that matters, and any map with the same distribution and autocorrelation would be just as good: anatomical location of the receptors does not matter at all, it can be completely random. This is another reason why talking about "serotonergic psychedelics" in the title is misleading: One could equally claim "random maps for the treatment of disorders of consciousness is supported by whole-brain modelling", because the results show this. The Abstract focuses on the success of serotonin-related maps, but forgets to mention that any random map with same statistical properties works indistinguishably well. This must be reported prominently.

We appreciate the reviewers recognition regarding our added analysis on null maps, which we believe has ameliorated our work. The reviewer also makes an acute comment regarding the implications of our null control. Nevertheless, we do report this in the revised manuscript. As we are modelling the pharmacological perturbations the main drive is the overall presence throughout the brain. Of course, this is not a full representation of the mechanism by which neurotransmitters act in the brain, rather the mechanisms captured by our model. Whereas this can be seen as a flaw it is also the richness of modelling: a parsimonious approach through a reduction of the world. With respect to the title, we have made the proper modifications as mentioned in comment 1 to reinforce the contribution of our work to the modelling community.

3) Another inaccurate statement, in the Abstract: "This effect was mediated by the brain-wide density of activated receptors". This is not correct. Authors must make clear to the reader that PET as used here is not a quantification of the absolute value of receptor presence. It is a relative measure, and different PET maps used different regions as reference. For all we know from these PET maps, the actual number of 5T2a receptors in the brain could be orders of magnitude smaller than D1 or D2 in the brain, even if its normalized mean value is the highest. This is not an issue with normalization, it's an issue with the data. Any claims about "how much receptor there is in the brain" and similar are simply impossible to make from this kind of PET map. The PET maps are a fine resource and good to use, but conclusions must respect what the data quantify and what they do not.

We understand the reviewer's concern regarding the interpretation of the normalised receptor values in absolute terms. While our normalisation preserves range it's important to note that these original values come from different references. We made the following modifications in the manuscript to clarify about the relativity of PET values:

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.

4) About the maps. The authors state: "we normalised each receptor map separately, as can be seen in Fig. R6. In that figure, each normalised map presents a different spatial distribution and different means in relative values. In particular, some maps with high mean values present a widespread anatomical distribution of receptors (e.g., 5HT2A), while others with low mean are quite sparse (e.g., D1). We note that our stimulation results depend on the mean of the normalised map, which in turn, considering the normalisation, is related to the presence of the receptor throughout the brain. In summary, our methodological procedure, which includes the receptor map normalisation and then the modulation of the impact of each receptor map with a scaling parameter, indicates that our results are not driven by the original mean intensity of the PET receptor maps."
How exactly did the authors perform normalization? The Methods only report that maps were "normalised ensuring that max = 1". There are many ways to obtain max of 1. Was the minimum subtracted, followed by division by the maximum? Or just division by the maximum? Or sigmoid/tangent function rescaling? This can make a large difference, and it will impact whether the authors' claim that "results are not driven by the original mean intensity of the PET receptor maps" is justified or not. Please provide the mathematical formula. I also recommend to show a histogram of the pre-normalization and post-normalization values next to each brain map (the brain maps were already a good addition, thank you).

We thank the reviewer for noting this clear lack of clarity. We have added the following in the methods section:

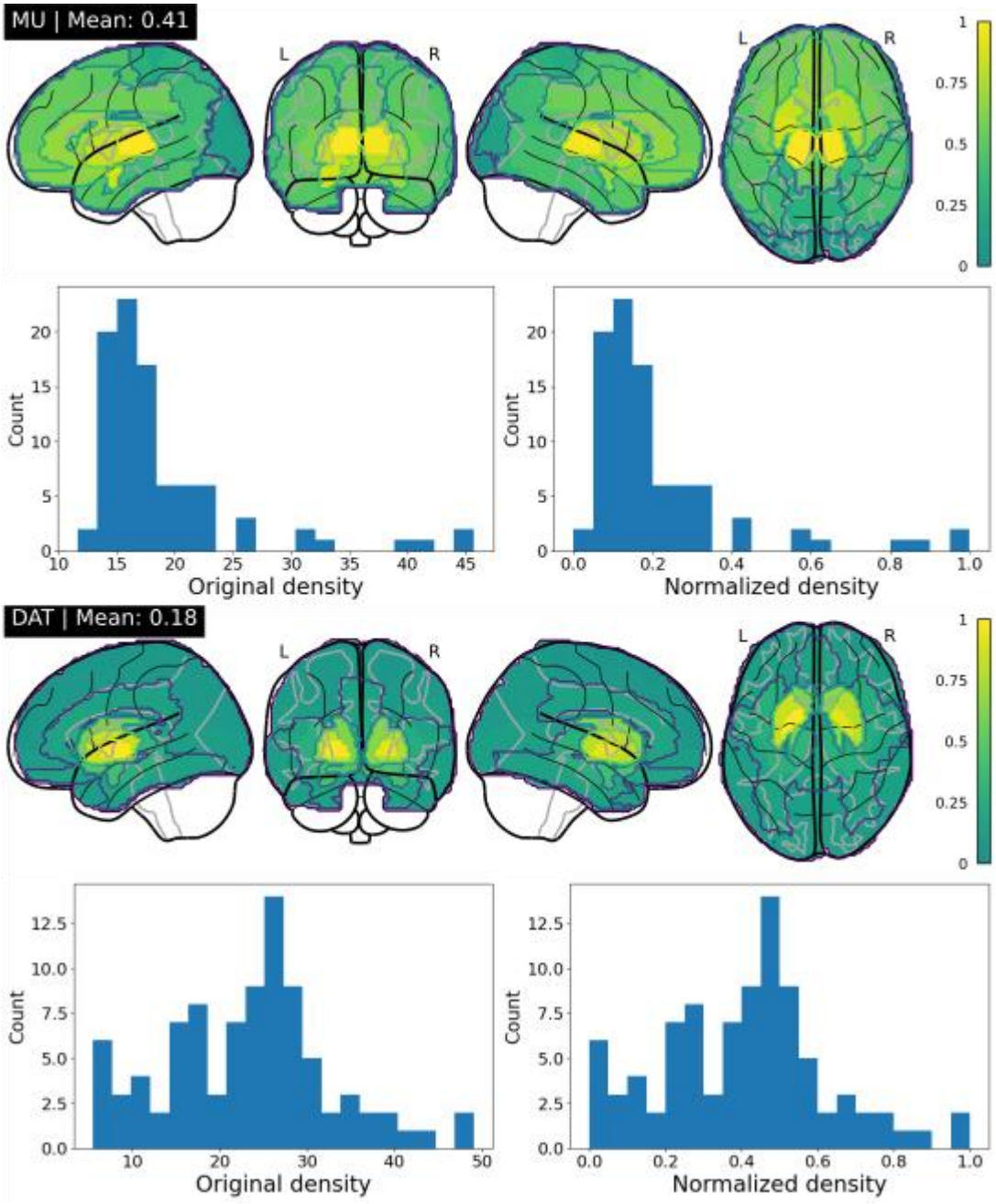
This processed dataset provided us with a quantitative measure of receptor densities in each AAL region denoted as Di. These density values were normalised using min-max normalisation:

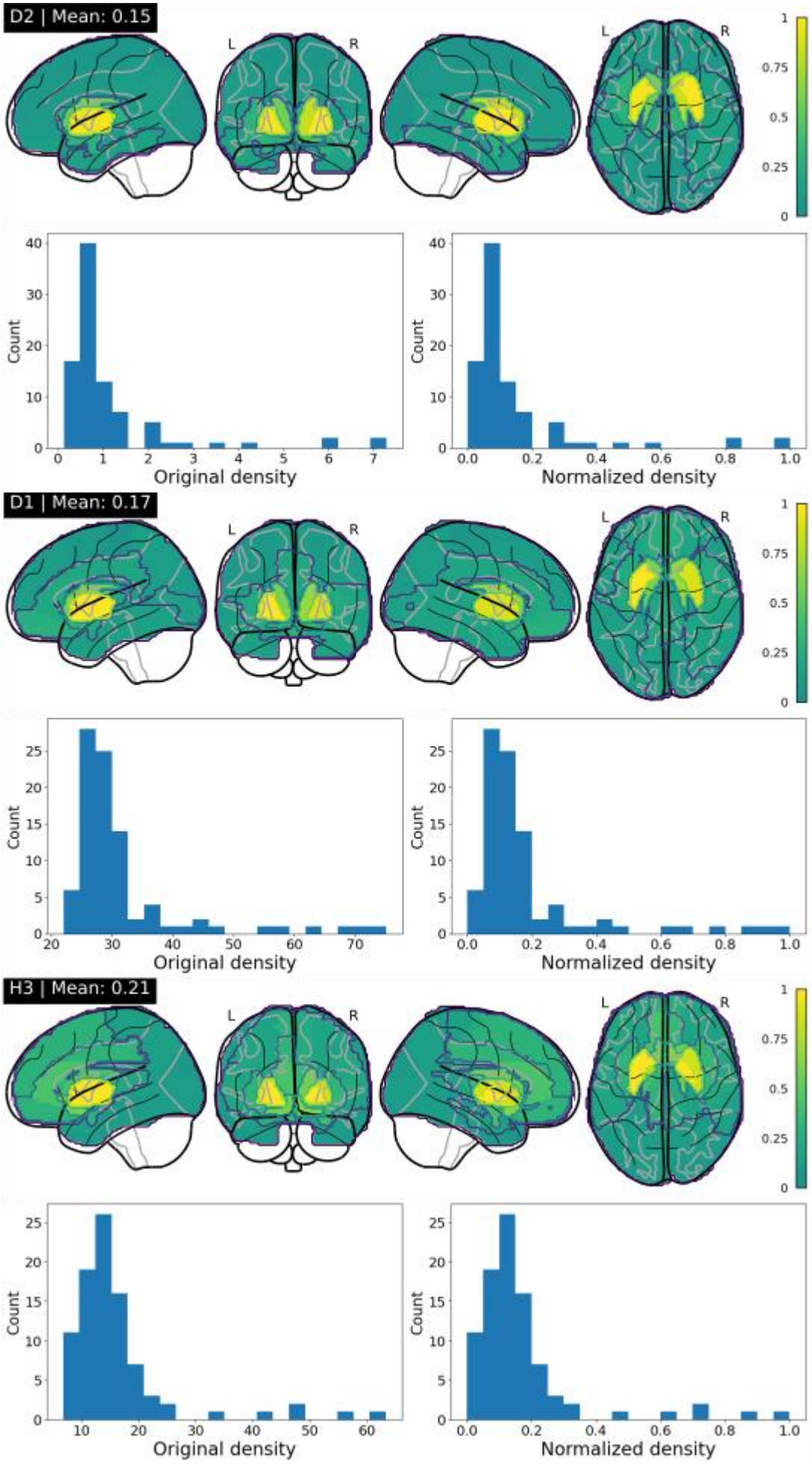
$$\square' = \square/(\square\square\square(\square) - \square\square\square(\square))$$

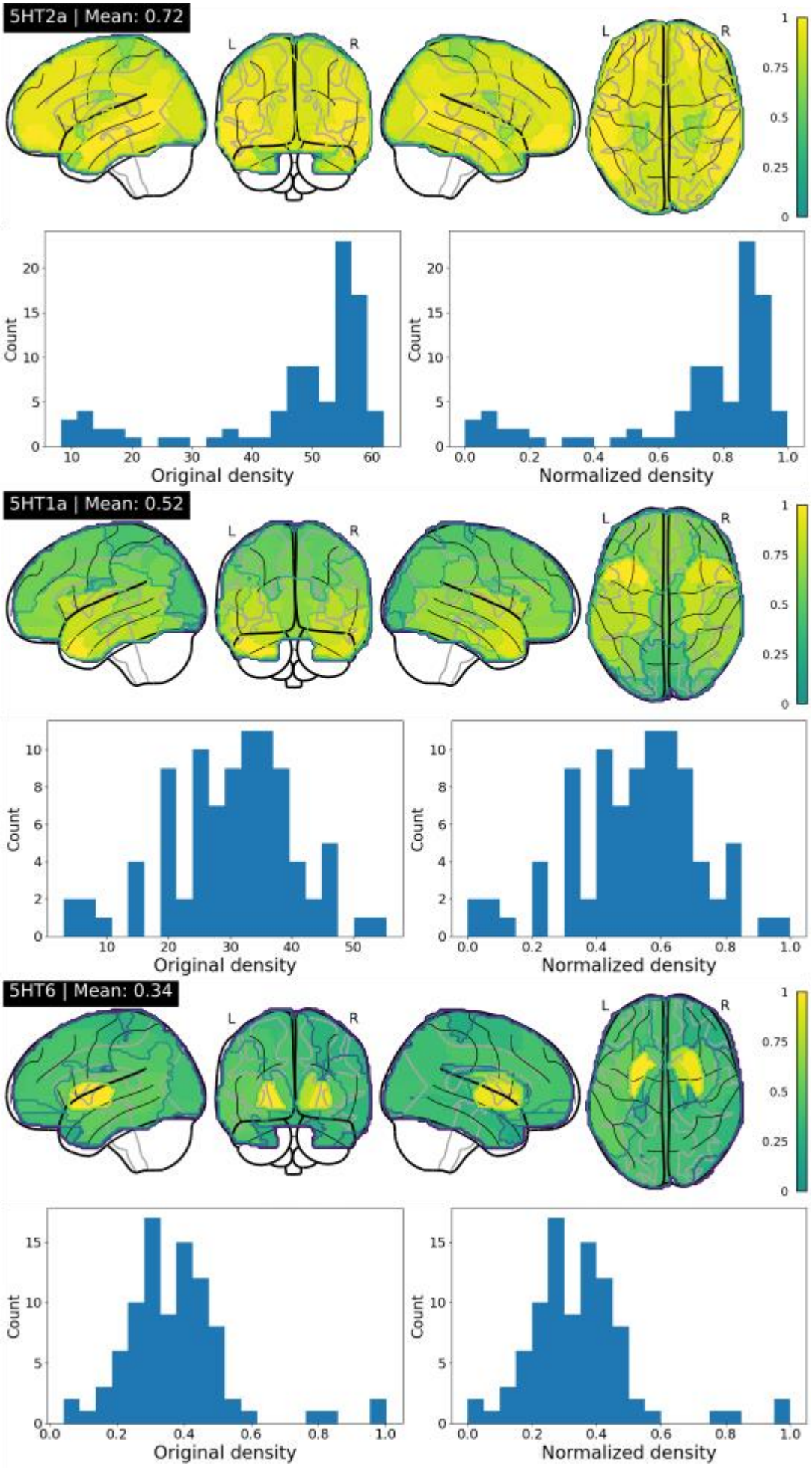
$$\square_{\square\square\square\square} = \square' - \square\square\square(\square) + 1$$

Leveraging Dnorm-i, we modulated the firing rates ri(E,I) of the excitatory and inhibitory neuronal pools in each brain region, drawing inspiration from (68).

And the required histograms are shown in the supplementary figure 3:

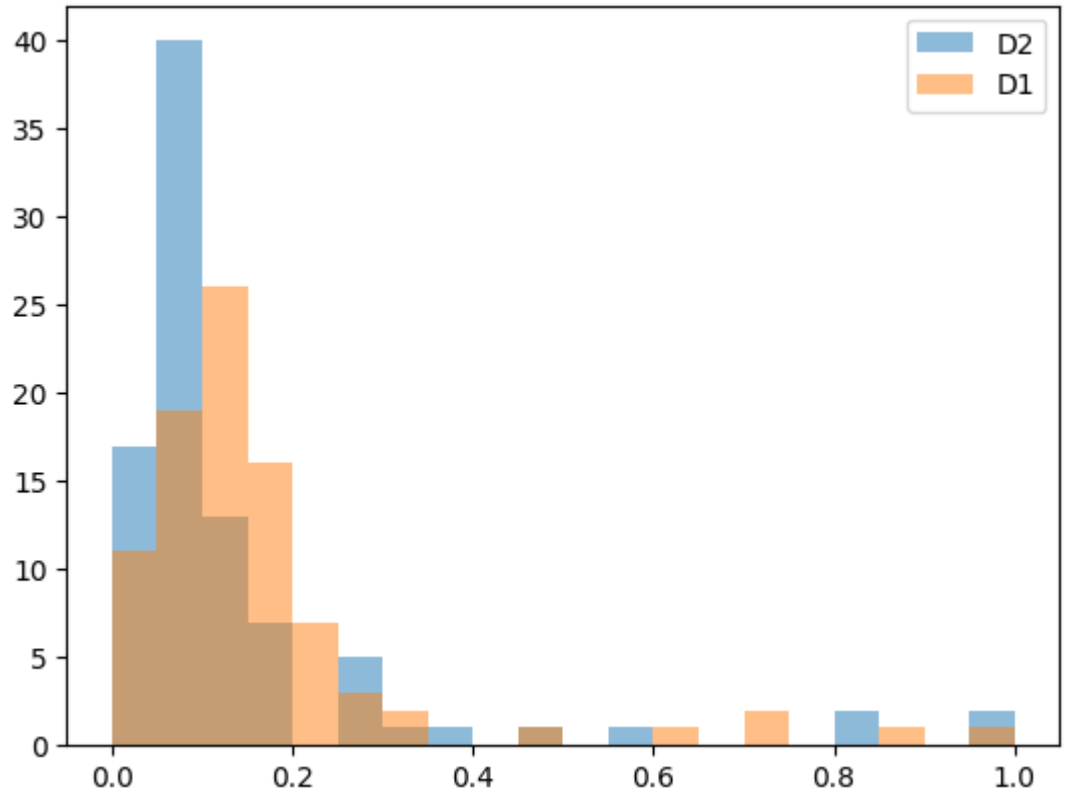

MU | Mean: 0.41
L
R
L
R
1
0.75
0.5
0.25
0
Count
Original density
Normalized density
DAT | Mean: 0.18
L
R
L
R
Count
Original density
Normalized density


D2 | Mean: 0.15
L
R
L
R
1
0.75
0.5
0.25
0
Count
Original density
Normalized density
D1 | Mean: 0.17
H3 | Mean: 0.21


5HT2a | Mean: 0.72
L
R
L
R
1
0.75
0.5
0.25
0
Count
Original density
Normalized density
5HT1a | Mean: 0.52
L
R
L
R
1
0.75
0.5
0.25
0
Count
Original density
Normalized density
5HT6 | Mean: 0.34
L
R
L
R
1
0.75
0.5
0.25
0
Count
Original density
Normalized density

5) Abut the maps: D1 and D2 look almost identical. This is a bit surprising. Could the authors confirm that the correct maps are plotted?

Yes, we have confirmed this



6) Also, looking at the maps it is evident that all except the serotonin maps are dominated by subcortex. Those dominated by subcortex also have lower means. I think this is what the authors mean when they say "some maps with high mean values present a widespread anatomical distribution of receptors". MU and 5HT6 have a bit higher values in the cortex, and their performance is also in the middle. It seems that the comparison at the end of the day is between cortex-dominated and subcortex-dominated maps, which predicts the mean, which predicts performance. So the title should be about cortically-dominanted receptor maps in general, rather than serotonin.

The reviewer brings an interesting topic regarding a possible domination of subcortical regions. While this comment may be correct for the receptors included in the present analysis, our work does not include enough receptors to make an accurate general statement regarding subcortically dominated maps. Additionally, AAL90 only has four subcortical regions for each hemisphere so we would not like to speculate on cortical-subcortical dominance in this work. Instead, we believe this

should be addressed in future work using more detailed sub-cortical parcellations. We have however modified the discussion to highlight this insight:

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 (72). Although there are theoretical benefits for subcortical stimulation, the clinical results are not conclusive or still in early stages (60), which could be benefited from in silico experimentation.

7) The explanation: "although we can increase the global coupling in our model, biologically this is not possible in a real experimental setting. What we can do is stimulate different receptor systems to induce changes. This is the reason for choosing a biophysically realistic model for our simulations. Although these are two different mechanisms, we are observing the same effect induced through perturbations on the biologically interpretable parameter. Our result indicates that some receptors can easily replicate the global coupling increase while others cannot." This explanation is very useful, and I strongly recommend the authors to include it in the Discussion to clarify that the pharmacology is being used as a way of countering the lower global coupling observed in the DoC patients.

We thank for the reviewer's advice and have added the explanation in the Discussion:

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. Evenmore, although we cannot biologically increase global coupling in a patient we can modulate responsiveness through receptor stimulation.

I think if these clarifications are added, and the language is made to properly reflect what the results show, then this will be a valuable addition to the modelling literature.

Reviewer #3 (Remarks to the Author):

I accept the paper.

We thank the reviewer for their work.

REVIEWERS' COMMENTS:

Reviewer #2 (Remarks to the Author):

Thank you for being so responsive to my remaining concerns. I am very happy to say that I think they have been addressed. I recommend publication. The manuscript provides appropriate nuance and will be a great addition to this field, congratulations.