Whole-brain modelling of low-dimensional manifold modes reveals organising principle of brain dynamics

Yonatan Sanz Perl[1,2,3*], Sebastian Geli[1], Eider Pérez-Ordoyo[1], Lou Zonca[1], Sebastian Idesis[1], Jakub Vohryzek[1], Viktor K. Jirsa[4], Morten L. Kringelbach[5,6,7,8], Enzo Tagliazucchi[2,9] & Gustavo Deco[1,10,11*]

1 Center for Brain and Cognition, Computational Neuroscience Group, Universitat Pompeu Fabra, Barcelona, Spain

2 National Scientific and Technical Research Council (CONICET), CABA, Buenos Aires, Argentina

3 Paris Brain Institute (ICM), Paris, France

4 Institut de Neurosciences des Systèmes, Aix Marseille Université, Marseille, France

5 Department of Psychiatry, University of Oxford, Oxford, United Kingdom

6 Center for Music in the Brain, Department of Clinical Medicine, Aarhus University, Århus, Denmark

7 Life and Health Sciences Research Institute (ICVS), School of Medicine, University of Minho, Braga Portugal

8 Centre for Eudaimonia and Human Flourishing, University of Oxford, Oxford, United Kingdom

9 Latin American Brain Health Institute (BrainLat), Universidad Adolfo Ibáñez, Santiago, Chile

10 Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, Spain

11 Institució Catalana de la Recerca i Estudis Avancats (ICREA), Barcelona, Spain

[*] *Corresponding authors: Yonatan Sanz Perl <yonisanz@gmail.com>, Gustavo Deco <gustavo.deco@upf.edu>*

Keywords: Modes, Manifold, whole-brain dynamics, computational modelling

Abstract

The revolutionary discovery of resting state networks radically shifted the focus from the role of local regions in cognitive tasks to the ongoing spontaneous dynamics in global networks. Yet, there is a growing realisation that these resting state networks could be a bit like the shadow tracings in Plato's famous cave, perhaps mere epiphenomena of an underlying hidden space from where these shadows emanate. Here we used deep variational auto-encoders to extract manifolds of low dimensionality from whole-brain dynamics measured with functional magnetic resonance imaging (fMRI). Crucially, we constructed the first dynamical model of the low dimensional manifold modes, i.e., networks of nodes using non-linear oscillators coupled with the effective functional connectivity, taking into account the level of non-equilibrium dynamics quantified by the non-reversibility of the signals. Irrespective of parcellation size, we found an optimal number of roughly ten manifold modes to best describe the whole-brain activity. Importantly, compared to traditional whole-brain modelling using all the nodes in a parcellation, we obtained better results for resting and task activity by modelling the dynamics of the coupled manifold modes. These findings show the key causal role of manifolds as a fundamental organising principle of brain function at the whole-brain scale, providing evidence that networks of brain regions rather than individual brain regions are the key computational engines of the brain.

Introduction

The high-dimensionality of brain activity recordings is one of the major obstacles hindering experimental and theoretical efforts in neuroscience. At the level of single cells, the human brain is constituted of billions of neurons and synapses where these neurons interact[1]; however, coarser scales can also be understood as complex interconnected systems where nodal dynamics represent the aggregate of a macroscopic population of neurons[2,3]. Even if exhaustive recordings to these systems could be obtained, it would remain difficult to reach encompassing theoretical principles due to the sheer complexity of the data. In particular, addressing the behaviour of individual cells does not suffice for this endeavour, since the high interconnectivity of the brain facilitates the emergence of coordinated activity[3,4], and thus of distributed representations where information is encoded in the global dynamics of large populations of neurons.

The discovery of resting state networks was a first step towards lowering the dimensionality of brain signals[5,6], which radically shifted the focus from the role of local regions in cognitive tasks to the ongoing spontaneous dynamics in global networks. Still, it is not clear how many resting states exist and indeed it is not clear how these emerge. Perhaps the resting state networks are a bit like the shadow tracings in Plato's famous cave; perhaps mere epiphenomena of an underlying hidden space from where these shadows emanate.

Here we provide a clear and mechanistic understanding of the hidden space in Plato's cave. This builds on the considerable efforts in recent years, which have been invested to tackle distributed representations by means of nonlinear methods capable of finding low dimensional representations of brain activity at multiple spatiotemporal scales, from cortical microcircuits[7,8] to whole-brain dynamics measured with functional magnetic resonance imaging (fMRI)[9–14], among other neuroimaging methods. The success of this approach does not only rely on methodological considerations, but also on the fundamental characteristics of brain activity. Despite the very large number of degrees of freedom of brain dynamics, coordinated cognition and behaviour cannot exist without the integration of this activity[15–17]. Neural information processing also exhibits redundancy[18–20], where the activity of several cells is constrained to develop within an abstract geometrical space of lower dimensionality, generally known as *manifold*. The organization of brain activity into manifolds with a reduced number of dimensions is a frequently replicated phenomenon, and has been proposed as a fundamental aspect of brain dynamics, fulfilling different scale-dependent computational roles[21]. The manifold spans a subspace of modes, in which the brain activity evolves following a flow, representing the evolution laws prescribing the behaviour of the state variables on the manifold[22].

The so-called structured flows on manifolds (SFMs) emerge from basic multi-scale processes such as symmetry breaking and near-criticality[23]. Indeed, a direct link has been proposed between the behavior and the flow of the low-dimensional manifold underlying the brain dynamics[24,25]. In particular, it has been shown that manifold coordinates parametrize motor control[26], perception[27,28], cognition and attention[29], navigation[30] and sleep[7], among other behaviours.

Here we address two predictions that emerge from assuming a key role of manifolds in brain activity supporting cognitive function. We first assume that brain activity is most adequately described in terms of a reduced number of abstract variables, here termed *modes*, each encoding the simultaneous dynamical behaviour of multiple macroscopic brain regions. These modes are neurobiologically meaningful as they emerge due to the large-scale anatomical and functional organization of the brain. Formally, modes are the coordinates spanning the subspace, in which the low-dimensional manifold is embedded. The *intrinsic dimensionality* of brain dynamics corresponds to the minimum number of such modes required for its description. If a manifold underlying whole-brain activity emerges as a consequence of anatomical interconnectivity, we can predict that the optimal number of dimensions will be independent of methodological choices such as the size of the original state space, which depends on the choice of parcellation. The second prediction is that computational models capturing the temporal evolution of the modes should supersede models optimized to reproduce dynamics within the original state space, which contains redundant information due to strong correlations between its variables. In particular, if the underlying manifold organization is relevant for the coordination of large-scale activity involved in cognition, we expect that modelling the dynamics of coupled modes more adequately tracks brain activity as the participants perform different tasks.

Building on previous work, we addressed these predictions using a deep learning architecture known as variational autoencoders (VAE) to obtain low-dimensional representations of whole-brain activity from a large-scale fMRI data set of more than 1000 participants (Human Connectome Project, HCP). First, we investigated whether the dimensionality of the low-dimensional manifold is independent of the dimension of empirical data by using three different grained brain parcellations. Then, we modelled the dynamics of each manifold coordinate with a non-linear oscillator operating close to the bifurcation point, which changes the dynamical behaviour from fixed point dynamics towards self-sustained oscillations, as successfully applied in several prior works modelling the dynamics of the high dimensional state space[31–33]. To optimize the models, we inferred the effective connectivity between the modes considering the level of non-equilibrium dynamics of brain activity quantified by the non-reversibility of the signals (termed GCAT, following the work of Kringelbach et al. 2023 [34]). We

demonstrate that by modelling brain dynamics on the manifold we achieve superior models compared to the traditional ones constructed in the source state space, more precisely tracking brain activity during seven cognitive tasks included in the HCP dataset. These results establish that the inferred connectivity in the manifold space provides more informative results compared to those obtained solely in the source space of empirical recordings.

Results

Methodological overview

Our methodological approach encompasses two complementary stages, addressing the two main questions we posed. The first analysis involves assessing the optimal dimension of the reduced space, while the second focuses on developing computational dynamical models within that manifold. Both analyses employ deep learning neural networks called autoencoders (AE) to obtain the low-dimensional manifold representation. In general, AEs consist of three key components: the encoder network, the middle layer (referred to as the bottleneck or latent space), and the decoder network. These components work together to capture and reconstruct the underlying structure of the data. A summary of the proposed framework is displayed in **Figure 1**.

In the first stage (**Figure 1A-C**), we employed the same architecture of the autoencoder (AE) multiple times to investigate the low-dimensional manifold obtained from the fMRI data of resting-state participants in the Human Connectome Project (HCP) for different cortical brain parcellations: the DK62 parcellation[35], which was constructed using the Mindboggle-modified Desikan–Killiany parcellation [36], with a total of 62 cortical region, and two Schaefer parcellations[37] with 500 and 1000 cortical regions (**Figure 1A)** . We used as input to the AE the time-by-time spatial patterns from each region empirical timeseries (**Figure 1B)**, i.e., the values of the *N* brain regions at each time, and we created a training set to optimize the parameters of the AE. Finally, we investigated the reconstruction error evaluated in a test set of participants as a function of the latent dimension for the three parcellations (**Figure 1C)**.

Next, we adopted the coarse-grained parcellation (DK62) and a we explored the manifold dimension around 10, based on the results obtained in the previous analysis (**Figure 1D**). We repeated the previously conducted procedure, while employing a variational autoencoder (VAE) instead. The VAE consists of the same three structures as the AE, but presents key differences that give VAEs generative properties. In brief, by forcing a regularization of the latent space through incorporation of a regularization term during the training process, the VAE ensures that the decoding step produces

outputs that are both relevant and meaningful[38] (see Methods). Importantly, since our aim was to construct generative models in the latent space with the intention of investigating the decoded modelled signals, regularization of the latent space provided by VAEs was essential. Subsequently, we constructed computational models in the latent space by describing the dynamics of each latent variable using a non-linear Stuart-Landau oscillator close to the bifurcation point, i.e. close to the critical regime (**Figure 1D**). To infer the connectivity between the latent variables, we implemented the GCAT framework developed by Kringelbach and colleagues, which involves a gradient descent procedure where the effective connectivity is updated at each step, considering the disparities between the empirical and simulated functional connectivity, as well as the forward shifted connectivity, to generate a good fit of both observables (i.e. zero and one-lag functional connectivity) at the same time. The outcome of this procedure is an optimized effective connectivity between the latent variables (LGEC), which captures the breakdown of detailed balance, reproducing the non-equilibrium dynamical behaviour of brain signals[34]. Finally, we were able to decode the modelled signals to further generate observables that we could compare with those computed in the original state space to quantify the performance of modelling the modes that constitute the low-dimensional manifold. We also created whole-brain models fitted to the same the empirical data in the source space, i.e., we replicated the GCAT framework in the original data consisting in the timeseries of each brain region (**Figure 1D,** lower row). We then compared the performance of both model in terms of the similarity between the empirical and modelled functional connectivity (FC). We found that the latent model more faithfully reproduces the empirical data compared with the source model (**Figure 1D,** rightmost panel).

We determined the brain regions contributed to each of the nine latent modes previously found in terms of brain functional networks. We created a set of surrogate signals in the latent space by introducing standard Gaussian noise into one latent dimension while keeping the other eight dimensions devoid of any signal. We repeated by changing the noise signal for each of the nine modes. We then decoded the surrogate signals in each case to obtain the corresponding spatial pattern in source space. We then associate these patterns with the activation (mode positive) or deactivation (mode negative) of the seven resting state networks from Yeo[39] (**Figure 1E**).

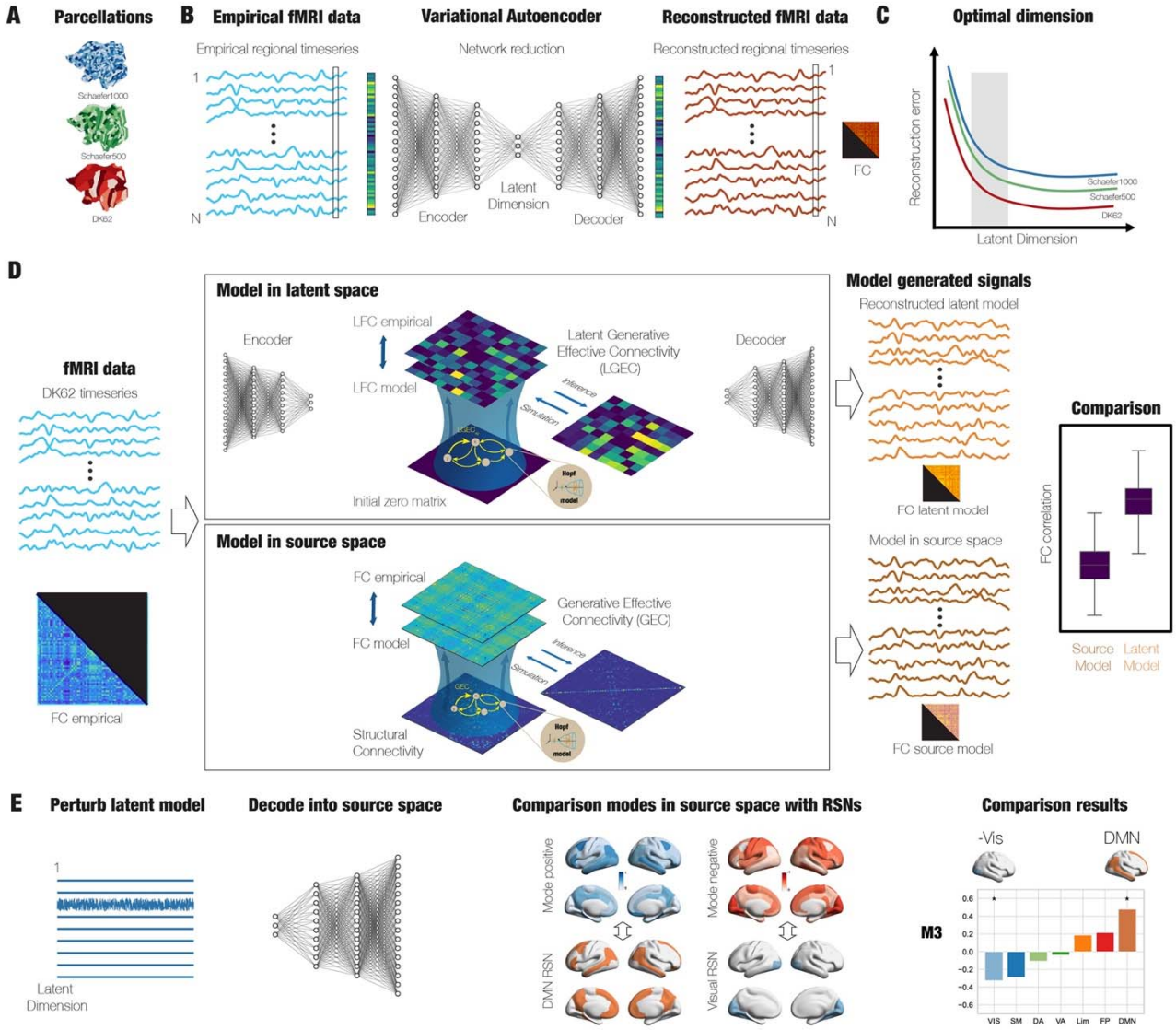

Figure 1. Overview of pipeline. A) *We considered three different brain parcellations with different amount of brain regions (N=62, 500 and 1000 brain regions).* ***B)*** *We extracted the empirical fMRI timeseries from a large-cohort of HCP participants in resting state in these three parcellations and we used these empirical regional timeseries to train an Autoencoder (AE). AEs are complex neural networks composed of three components: the encoder network, the middle layer (also known as the bottleneck or latent space), and the decoder network. In our implementation, the encoder network consisted of a deep neural network to reduce the dimensionality until reaching the latent space layer, where the dimensions were explored. Conversely, the decoder network replicates the architecture of the encoder, generating reconstructed matrices based on each encoded point in the latent space.* ***C)*** *We trained the same AE architecture using as an input the time-by-time spatial patters considering the three different parcellations (N=62, 500 and 1000 brain regions) and investigated the behaviour of the reconstruction error evaluated in the test set as a function of the latent dimension in the three cases.* ***D)*** *We trained a variational AE (VAE) using the parcellation with N=62 brain regions, which consists of the same three structures as the AE, but including a regularization loss function that gives VAEs generative properties. To*

model the dynamics of each neural mode, we utilized a non-linear Stuart-Landau oscillator near the bifurcation point, i.e. near the critical regime. In order to determine the connections between the latent variables, we employed the GCAT framework developed by Kringelbach and colleagues. The result of this procedure is an optimized effective connectivity between the latent variables, LGEC, which captures the departure from detailed balance and generates the non-equilibrium dynamics observed in brain signals. We also created whole-brain models in the source space to fit the same empirical data by repeating the same optimisation procedure for the latent space mode, i.e., the inference of the GEC by implementing the GCAT framework. We compared the performance of both models and we found that models in the latent space more faithfully reproduce the empirical data quantified by the similarity between the empirical and modelled FC. ***E)*** *We investigated which brain regions contributed to each of the nine latent modes previously found in terms of brain functional networks. To do so, we created a set of surrogate signals in the latent space by introducing standard Gaussian noise into one latent dimension while keeping the other eight dimensions devoid of any signal. We repeated by changing the noise signal for each of the nine modes. We then decoded the surrogate latent signals in each case obtaining the spatial pattern for each mode in source space. We then associate these patterns with the activation or deactivation of the seven resting state networks from Yeo.*

The ability of a low dimensional manifold to capture whole-brain dynamics

We trained the same autoencoder using three different time-by-time spatial patterns of dimension N=62, 500 or 1000 as the inputs (see Methods for more details of AE architecture). For each latent space dimension ranging from 2 to 20, we evaluated the performance of the AE in reconstructing each spatial pattern by measuring the mean squared error (MSE) between the input and the reconstructed pattern. We computed the average MSE across patterns and participants in each fold as well as in each parcellation. In **Figure 2A** we show the mean and standard deviation across 10-folds for each dimension, enabling us to compare the progression of the reconstruction error for each parcellation. We observed that the behaviour was consistent across all three cases, exhibiting a significant decrease in error in the initial dimensions until reaching an elbow point at approximately latent dimension equal to 10, beyond which the error remained relatively constant.

Gathering the reconstructed patterns, we were able to reconstruct the complete brain signals for each participant. We assessed the performance of the AE in terms of reconstructing the functional connectivity of each participant, i.e., the correlation between the signals of each pair of brain regions, which is one of the typical observables used in the literature [40]. To quantify the performance of the reconstruction, we computed the correlation between the empirical functional connectivity (FC_S^{Emp}), and the reconstructed (FC_S^{rec}) for each participant (the subindex S indicates that this matrix is computed in the original high dimensional state space). We explored the manifold dimension ranging from 2 to 20 for the three parcellations and we averaged across participants across the 10-folds. **Figure**

2B displays the average and standard deviation across folds for each latent dimension for the correlation between the FC_S^{Emp} and FC_S^{rec}. Aligned with the previous result, we observed that an elbow point exists approximately in latent dimension equal 10 for the three parcellations, also for the case of the correlation between the empirical and reconstructed functional connectivity.

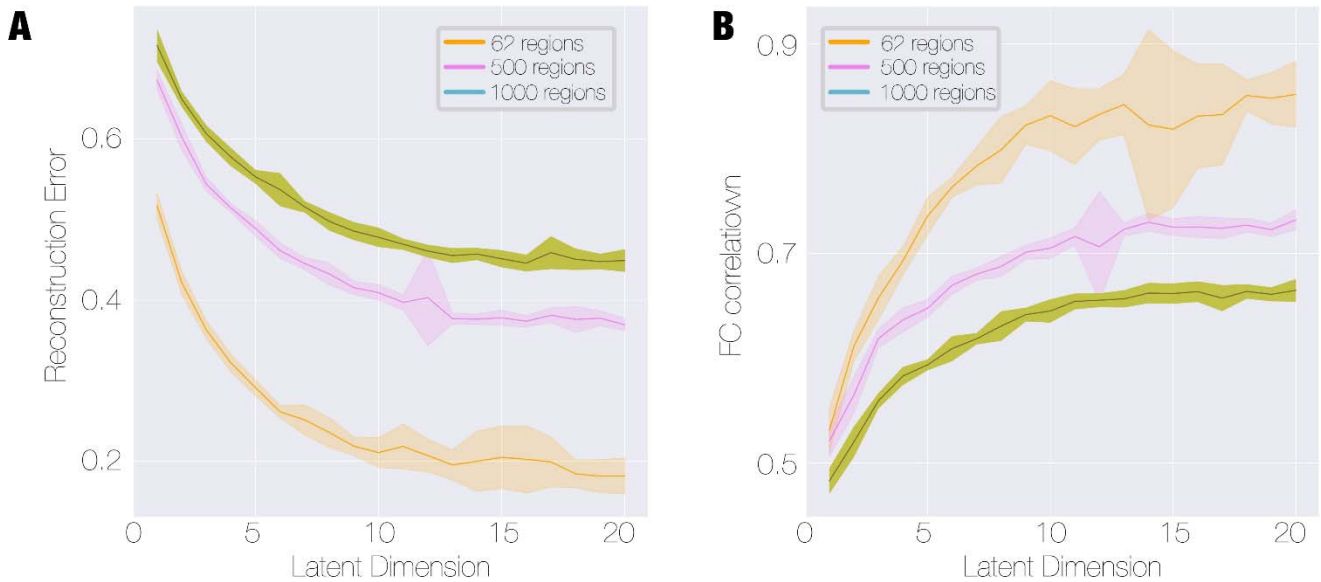

Figure 2. The optimal manifold dimension is invariant respect to the dimensionality of the original state space. A) Reconstruction error. *We trained the same AE using as input the time-by-time spatial patterns of 900 fMRI recordings of resting state participants obtained applying three different parcellations the HCP data (DK 62, Schaefer 500 and 1000). We trained and evaluated the AE with a 10-folds scheme randomly selecting the 90%/10% as training and testing sets. We repeated the procedure by changing the latent dimension from 2 to 20 and investigated the reconstruction error, computed as the MSE between each input and reconstructed pattern for 100 participants in the test set. We displayed the mean and standard deviations across folds, and we observed that around latent dimension equal to 10 the MSE presented an elbow for the three parcellations.* ***B) Functional Connectivity reconstruction.*** *We repeated the analysis but comparing the similarity between the empirical functional connectivity (FC_S^{Emp}), and the reconstructed (FC_S^{rec}) for each participant. We reconstructed the full signal of each participant and compute the pairwise correlation obtained the FC_S^{rec}, we then in the test set in each fold and dimension we compute the correlation between FC_S^{Emp} and FC_S^{rec} to quantify the level of reconstruction. We averaged across participants in each case. We display the mean and standard deviation across folds of the correlation between both FC for the test set. We found similar results in both the pattern reconstruction error and the FC correlation.*

Models of the coupled manifold modes outperform high dimensional models in the original state space

We employed a VAE to obtain a low-dimensional manifold representation of fMRI data obtained from resting-state healthy participants from the HCP dataset. The data was obtained by applying a coarse-grained parcellation scheme comprising 62 regions, following our previous findings.

Subsequently, we modelled the dynamics of each neural mode (i.e. manifold dimension) using non-linear Stuart-Landau oscillators, which exhibit dynamic characteristics determined by a Hopf bifurcation (see Methods for the mathematical expression of the model).

Typically, these models are fitted to reproduce empirical observables, such as the functional connectivity (FC)[41,42]. In our case we built a manifold model to fit the empirical functional connectivity between latent space modes (FC_L^{Emp} the subindex L stands for FC computed in the latent space and super index *emp* means that uses encoded empirical data), which is computed as the pairwise Pearson's correlation between the latent modes, and the empirical latent forward-shifted connectivity (FCf_L^{Emp}), which represents the pairwise Pearson's correlation between the signals in latent space and the same signals shifted forward in time (see Methods). To improve the model fitting capacities, we averaged the FC_L^{Emp} and FCf_L^{Emp} by subgroups of 10 participants and created 100 models that fit these average matrices.

To construct the model of the low-dimensional manifold, we fixed the bifurcation parameter for each mode to $a_j = -0.02$ and we built the whole manifold model by coupling the oscillators representing each latent neural mode. The connectivity between the modes was inferred using an iteratively pseudo-gradient algorithm, which was proposed by Kringelbach and colleagues for models in the source space[34]. This iterative process enabled us to accurately capture the underlying relationships between latent modes. By fitting the FC_L^{Emp} at the same time that the FCf_L^{Emp}, we could effectively capture the non-equilibrium dynamic patterns in latent signals[34]. The result of this process is an enhanced effective connectivity for each one of the 100 models between the latent modes (LGEC), which also captures the disruption of the detailed balance, leading to the generation of non-equilibrium dynamic patterns in latent signals. We show that this disrupted detailed balance is also captured by the decoded signals.

We proceeded to assess the overall performance of the complete framework by examining the similarity between the modelled and decoded FC (FC_{LS}^{Mod} computed as the Pearson's correlation of the signals modelled in the low-dimensional manifold and then decoded to the source space) and the

empirical FC in the source high dimensional space (FC_S^{Emp}) quantified by the correlation between both FC. We also computed the correlation between the forward matrices FCf_S^{Emp} and FCf_{LS}^{Mod} (**Figure 3A**). Both metrics exhibited similar behaviour, displaying an optimal point at latent dimension 9, with decreasing performance observed for higher and lower dimensions. To validate the reliability of our approach, we constructed the model directly in the source state space using the GCAT framework, following the exact same methodology as employed in the latent space. However, we observed significantly lower levels of fitting for both observables compared to the results obtained from the low-dimensional manifold modelling (blue boxplot in **Figure 3A**). This discrepancy was observed not only for the optimal dimension but also across all dimensions under evaluation, demonstrating that even outside of the optimal manifold representation, the low-dimension representation is still more accurate than the high-dimensional original space.

Figure 3B depicts hybrid matrices as a function of the latent dimension that compare the empirical functional connectivity FC_S^{Emp} (upper triangle) and one obtained as FC_{LS}^{Mod} (lower triangle).

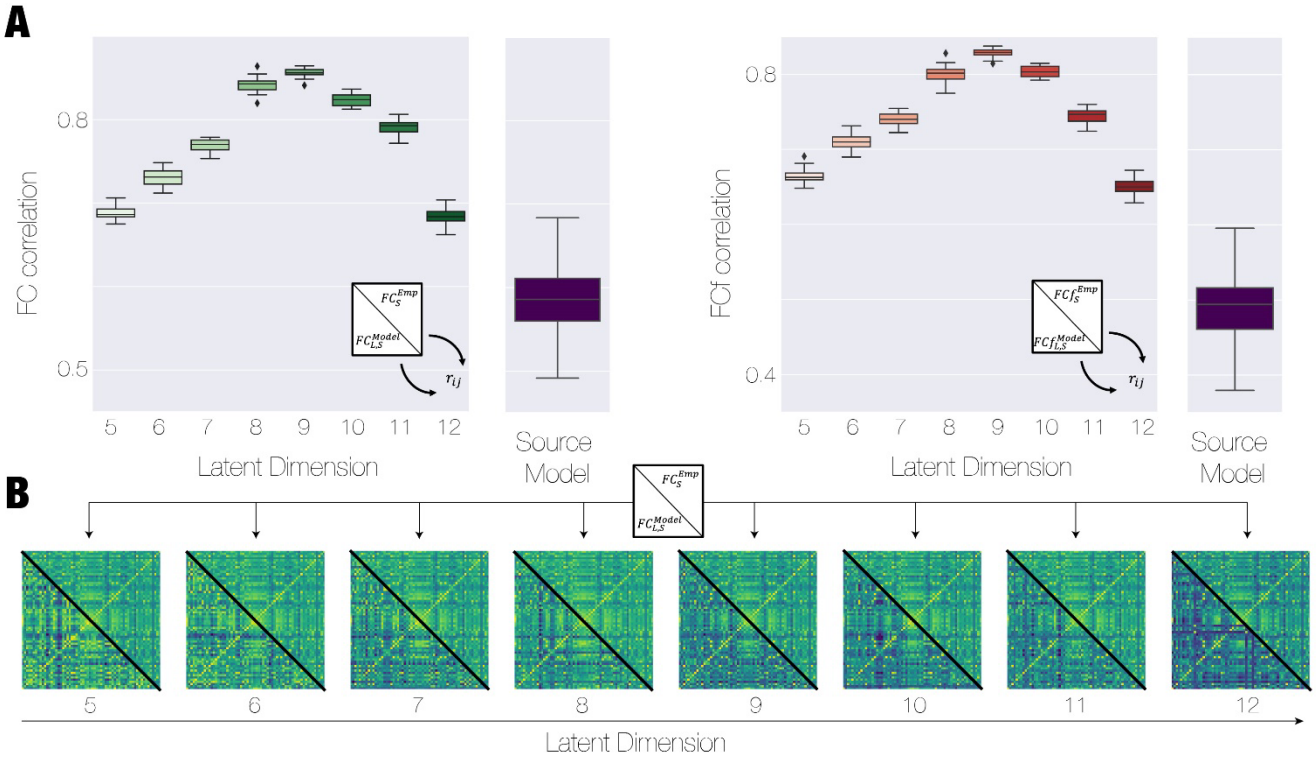

Figure 3. Modelling the low-dimensional brain manifold reconstructs empirical data, outperforming models in the original state space. A*) We generated with a VAE a low-dimensional manifold representation of fMRI data from resting-state healthy participants in the Human Connectome Project (HCP) using a coarse-grained parcellation scheme consisting of 62 regions. To model the dynamics of each latent dimension within the manifold, we employed non-linear Stuart-Landau*

oscillators. These oscillators exhibit dynamic characteristics that are governed by a Hopf bifurcation. We constructed a manifold modes model that accurately fit the FC_L^{Emp} *and* FCf_L^{Emp} *which represent the pairwise Pearson's correlation between the signals in the latent space and the same signals with the corresponding shifted forward in time, respectively. The results of the performance of the complete framework in terms of reconstruction of* FC_S^{Emp} *and* FCf_S^{Emp} *are quantified as the correlation between these matrices and the ones obtained through decoding the manifold models for each dimension. The maximum correlations are reach for latent dimension 9 and decreasing both for low and high dimension. Importantly, we demonstrated that these models overcome the performance of a model developed in the high-dimensional source space (blue boxplot), not only for the optimal dimensions but also for all latent dimension explored.* ***B)*** *The hybrid matrices display in the upper diagonal triangle the* FC_S^{Emp} *and in the lower diagonal triangle the* FC_{LS}^{mod} *to observe the similarity between both matrices when explore different latent dimension ranging from 5 to 12.*

Source network identification of latent modes.

Next, we investigated which brain regions contributed to each of the nine latent modes previously found in terms of brain functional networks. To assess this, we obtained one spatial pattern on the empirical original space associated to each latent mode (see Methods). We evaluated these patterns in terms of the reference functional brain networks estimated by Yeo and colleagues: Visual (VIS), Somatomotor (SM), Dorsal Attention (DA), Ventral Attention (VA), Limbic (Lim), Frontoparietal (FP) and Default mode (DMN), known as the 7 resting-state networks (RSNs) from Yeo[39]. We computed the percentage of participation of each brain region within the DK62 parcellation to each of the 7 RSNs. We then computed the Pearson's correlation between each of the nine patterns corresponding to each neural mode and the percentage of participation of each brain region to each RSN. In **Figure 4A** we show, as an example, brain renders corresponding to the spatial pattern of the mode 3 divided in positive and negative, standing for brain regions that present high correlation and anticorrelation to certain RSNs (DMN and Vis, respectively). **Figure 4B** shows the level of correlation between each latent mode (M"N", with "N" varying from 1 to 9) and each RSN indicating with a star the ones that are statistically significant after false discovery rate (FDR) correction. We found that some modes could be associated with the activation of the DMN at the same time with a deactivation of the SM network (latent modes 1) or deactivation of the VIS (M3, 7, 8 and 9), while other can be associated with the activation of the Limbic network (M2). Other latent modes could be described as the activation of Visual areas at the same time of other networks as DA (M4) and other with activation of the VA and DA (M5). Finally, others dimension could be characterised by the activation of SM at the same time with FP and VA with a deactivation of VIS (M9), the FP are also activated at the same time that the DMN with a deactivation of VIS networks in latent mode 8. Interestingly, our association of latent dimensions with activation and deactivation patterns in the brain, shows high similarity with

previous results based on non-supervised clustering on brain activity considering the well-known coactivation patterns (CAPs) as is presented by Huang and colleagues[43].

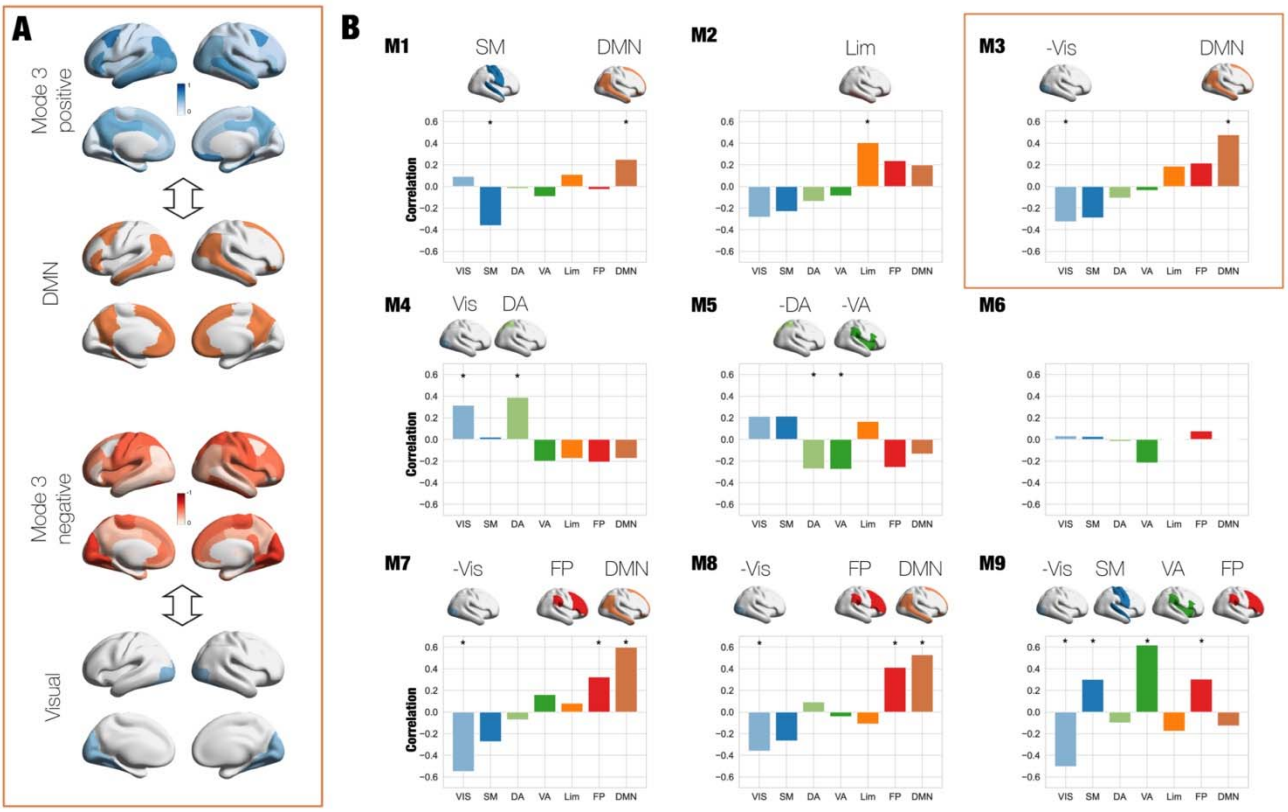

Figure 4. Resting state networks are formed from the modes revealed by VAE. A) *We determined the brain regions associated with each mode. We show as an example the brain renders corresponding to the spatial pattern of the mode 3 divided in mode 3 positive and mode 3 negative, standing for brain regions that present high correlation and anticorrelation to the DMN and Vis network, respectively.* ***B)*** *We identified each latent mode with a pattern in the source space of 62 brain regions. We associated each spatial patterns with the Yeo 7 resting state networks by computing the correlation of each pattern with the percentage of belonging of each region to each RSN (* indicates the correlation that are significant after false discovery rate correction). The reference functional brain networks estimated by Yeo and colleagues named: Visual (VIS), Somatomotor (SM), Dorsal Attention (DA), Ventral Attention (VA), Limbic (Lim), Frontoparietal (FP) and Default mode (DMN).*

Cognitive tasks modelled in the latent space

We investigated how cognitive processing is related to dynamics in the low-dimensional modes by analysing fMRI data from 7 cognitive tasks from the HCP data set (Social [SOC], Language [LAN], Working Memory [WM], Motor [MOT], Gambling [GAM], Emotion [EMO], Relational [REL]). Similar to the previous section, we trained the Variational Autoencoder (VAE) using the spatial patterns over time

obtained from the combined data of the rest condition and the seven tasks (see Methods). We then built a model for each task by inferring the effective connectivity between modes to obtain the best performance in reproduce FC_L^{Emp} and FCf_L^{Emp} (see Method). Thus, we built 100 models for each task and we assessed each model performance by measuring the correlation between the empirically functional connectivity (FC_S^{Emp}) and the latent model-decoded functional connectivity (FC_{LS}^{mod}) for each task. We observed that the goodness of fit was similar across tasks, but lower than what we obtained solely for the REST data (**Fig. 5A**). For each task, we obtained the LGEC for each one of the 100 models (**Fig. 5B)**. Finally, we trained a support vector machine (SVM) using the 100 LGEC matrices generated from modelling the manifold for each task. The objective was to determine if the information captured in each matrix could serve as a unique fingerprint for each task. We also trained the SVM classifier using the elements of the empirical functional connectivity (FC_S^{Emp}) and the LGEC matrices with randomized labels. Our results showed that the best classification accuracy (0.89±0.01) was achieved using the LGEC matrices outperforming the LGEC with randomized classes (0.13±0.03) and the FC_S^{Emp} (0.76±0.02) (**Fig 5C**). In **Fig 5C**, we also displayed in the rightmost column the confusion matrix obtained for the classification of the 8 labels, the 7 tasks and resting-state.

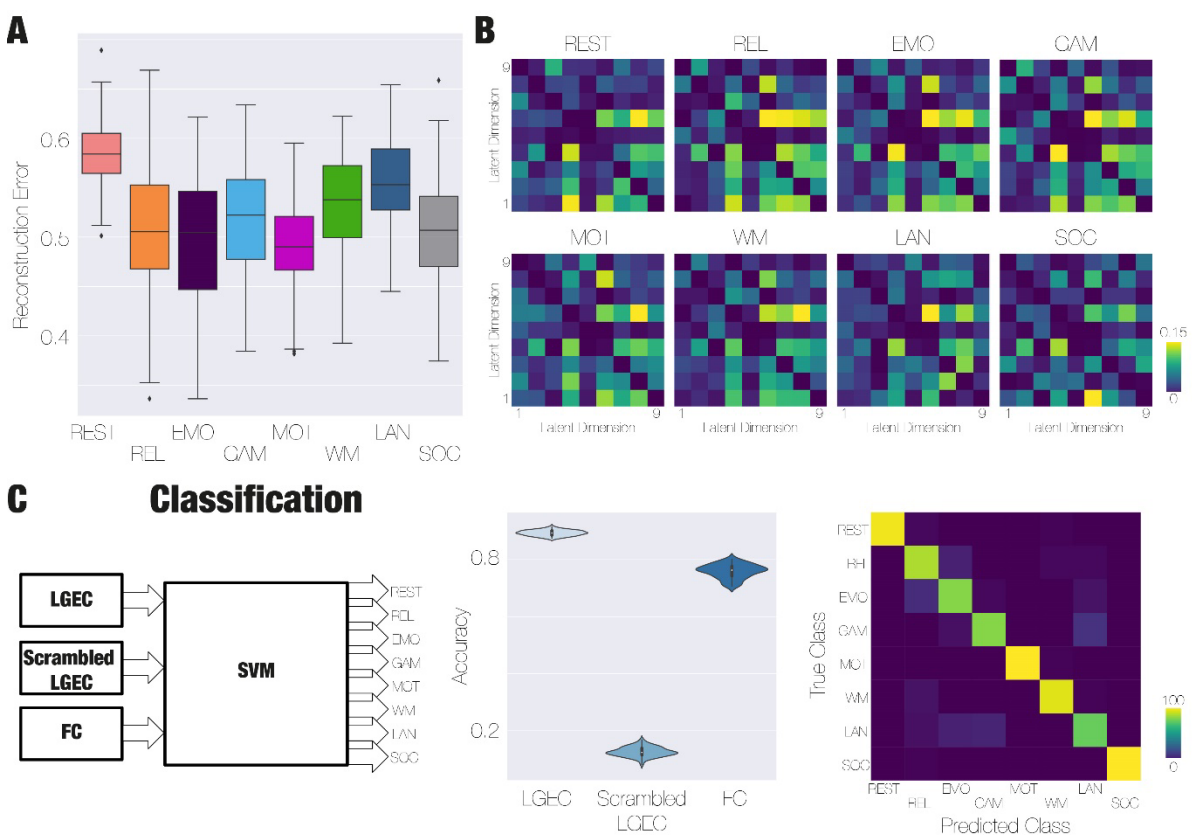

Figure 5. Modelling the low-dimensional brain manifold better distinguishes between cognitive tasks. A) *We constructed models for each task and evaluated the performance of each model by quantifying the correlation between the* FC_S^{Emp} *and* FC_{LS}^{mod} *in each case. We noticed that the level of fitting is similar across tasks, lower than the fitting obtained only for the REST condition* ***B)*** *The output of each model was the inferred connectivity between the latent variables, called the LGEC, for each task. We show the average LGEC across 100 models for each case.* ***C)*** *We trained a support vector machine (SVM) using the 100 LGEC generated by modelling the manifold in each task to evaluate whether the information condensed in each matrix is a fingerprint for each task. We also trained the same SVM classifier but using as an input the elements of the* FC_S^{Emp} *and also the LGEC with scramble labels. We found that the best classification is reached using the LGCAT (0.89±0.01) compared with the* FC_S^{Emp} *(0.76±0.02) and the scrambled (0.13±0.03).*

Discussion

We demonstrated that a low-dimensional manifold underlies whole-brain dynamics, and that the dimensions of this optimal manifold were preserved regardless the size of the original data. Moreover, we showed that generating dynamical models within the subspace in which the low-dimensional manifold is embedded could yield more accurate models in terms of reproducing empirical data compared with modelling the original high-dimensional whole-brain space. Both results are to be expected under the assumption that large-scale dynamics are orchestrated into a reduced number of independent degrees of freedom, which emerge as part of a broader neurobiological principle underlying brain activity. In the following section we discuss the implications of these results, both in terms of their contribution to systems neuroscience and their practical applications to neuroimaging.

Whole-brain fMRI dynamics are intrinsically low-dimensional

We investigated the existence and dimensionality of the manifold of fMRI resting-state recordings by implementing deep learning algorithms, called autoencoders, to whole-brain data in three different spatial resolution. We then showed that, regardless of the size of the original data, the optimal manifold to reproduce brain signals and their associated functional connectivity was approximately of dimension 10. Several past studies investigated the compressibility of resting state fMRI data using linear methods such as principal[44] and/or independent component analysis[5], revealing a set of canonical RSN overlapping with distinct neuroanatomical regions subserving specific cognitive functions[45]. Other methods, such as the clustering of co-activation patterns (CAPs)[46] results in similar information. These networks also show alterations in certain neuropsychiatric disorders[47], and contain information that can be used to predict and decode the contents of ongoing consciousness and cognition[48]; moreover, the

neurobiological underpinnings of the RSN are highlighted by multimodal imaging studies revealing that their activity correlates with different frequency bands measured using simultaneous EEG[49]. As shown **in Fig. 4**, the neural activity modes revealed using autoencoders are related to the RSN, although this relationship is not a simple one-to-one mapping. Instead, the modes are a combination of different RSN, indicating that whole-brain dynamics can be spanned by different sets of functional networks. In spite of the mapping between RSN and modes, the latter have distinct properties due to the nature of the generative autoencoder algorithm. Besides its capacity to capture nonlinear relationships between the modes and the high-dimensional data, autoencoders are generative, which is of key importance to translate the dynamics in the latent space towards that of resting-state fMRI data.

Large-scale manifolds and cognition

The use of dimensionality reduction techniques is a common strategy in statistics and machine learning, applied with the purpose of simplifying the description of the data by performing a reasonable grouping of the variables (e.g. clustering) or expressing them as a linear or nonlinear combination of a reduced set of components (e.g. ICA, PCA)[21]. The application of these methods to neural activity recordings is not only useful in terms of data processing, as expected, but also reveals important features of how the brain represents information relevant to cognition and behaviour[50]. Research conducted using a variety of methods and experimental organisms shows converging evidence of the manifold organization of neural activity, where the manifold dimensions can be parametrized by variables with correspondence to behavioural observations. Examples include the representation of movement parameters in the motor cortex[26], the mapping of spatio-temporal coordinates to the hippocampal cortex during navigation[30], and the encoding of sensory (e.g. visual[27], olfactory[28]) information. While these manifolds can change in response to learning[51], they are generally stable[52], and it is believed that they relate to the network of synaptic connections and their strengths; however, the relationship between manifolds and structural features remains difficult to demonstrate[53]. Our results suggest that large-scale brain activity also benefits from the organization into low-dimensional manifolds. Indeed, the corresponding dimensions (i.e. modes) may parametrize the engagement of different cognitive functions in spontaneous mentation, as shown by the classifier performances in **Fig. 5**. Thus, the organization of brain activity in the form of manifolds appears to occur across scales, where coarser scales index the participation of large assemblies of cells, whose dynamics may also be

understood in terms of a low-dimensional space. As in the case of neural circuits, future studies should attempt to relate the modes with the underlying anatomical connectivity of the human brain.

The origin of low dimensionality and anatomical maps

Our results also show that the original resolution of the data does not affect the optimal dimension of the underlying manifold revealed using autoencoders, which is close to 10, and similar to the number of RSNs reported in the literature[54]. Such relative low number of dimensions is likely related to two factors. First, even though fMRI is a method of high spatial resolution, the indirect nature of the hemodynamic response could obscure aspects of the neural dynamics required to perform a more fine-grained dissection of the data[55]. Second, both the RSN and the modes reported here relate to major neuroanatomical maps implicated in major sensory and cognitive capacities of the brain, which comprise a relatively short list[56]. Future studies should attempt to perform an in-depth characterization of the manifold dimensions in terms of variables of neurobiological relevance, such as receptor density, gene expression, myelinization, cortical thickness, among others, with the purpose of understanding how low-dimensional dynamics emerge and are shaped by human brain biology[57,58].

Possible disruptions in disease and unconsciousness

As in the case of RSN, we expect that the modes are sensitive to different neurological and psychiatric conditions. In particular, if these modes reflect the representation of ongoing cognition and consciousness in large-scale patterns of brain activity, we could expect the most salient alterations in disorders leading to severe impairments in these domains. An example is the case of brain injured patients who show a state of stable and persisting unresponsiveness, which can be interpreted as a lack of conscious thought[59]. Diminished or absent conscious content can also be present in certain transient states, either spontaneously occurring (such as deep sleep) or induced by pharmacological means[60]. Previous research shows that a low-dimensional representation of brain dynamics captures the differences between the aforementioned states and conscious wakefulness, the former showing a reduced repertoire of states visited during the recording, and less structured and complex transition between[11] these states[61]. Future studies should adopt the framework developed here to investigate different pathological and physiological brain states, with the purpose of obtaining complementary information on the landmark features of brain dynamics during health and disease.

Modelling the coupled modes dynamics

Our work provides novel methodological perspectives for the analysis of large-scale neuroimaging data. The majority of whole-brain modelling studies begins by adopting a certain parcellation of the human brain into regions of interest (ROI). Assuming the anatomical connectivity between ROI is known, it is then possible to couple the local equations and simulate whole-brain dynamics. Our work suggests that this methodology may be suboptimal since brain dynamics are better represented by modes which span multiple ROIs; instead, models should strive to capture the relationship between modes, as illustrated by **Fig. 5**. While the anatomical connectivity between the distributed sets of regions spanning the neural nodes is ill-defined, our result shows that equivalent results can be obtained using the effective connectivity. We also show that the classification of tasks from the HCP is optimal when based on the neural mode representation, suggesting that these manifold dimensions are not arbitrary but encode organizational principles of brain function instead. The generative capacity of autoencoders represents another attractive characteristic of our work, as it allows us to create new data following similar statistics as the empirical recordings, with potential applications to machine learning and data augmentation.

Limitations and future work

Our work presents some limitations that should be addressed in future studies. Perhaps the most salient one stems from our use of fMRI data, with known limitations to its biological interpretability. This should be assisted by other sources of information capable of complementing the recordings and shedding light on the origin of the neural models. Also, the use of improved MRI machines with higher fields could transcend the spatial organization of the cortex into maps, allowing researchers to focus on layer-specific contributions. The use of effective connectivity is a useful alternative to replace the anatomical couplings in case they are unavailable, but future studies should incorporate this data, investigating how departures from healthy organization affects the resulting modes. Finally, the set of tasks to be distinguished was relatively small. A larger repertoire of conditions should be included to tackle questions concerning the specificity of the underlying modes to classify pathological and physiological brain states.

Conclusions

We demonstrated that modes, understood as the independent dimensions of an underlying manifold, are relevant for the description of whole-brain dynamics. Regardless of the dimensionality of the original

data, a robust set of modes emerges after the application of deep generative autoencoders. The interactions between these modes could be modelled, leading to brain-state decoders superseding those constructed in the original high dimensionality space. We conclude that standard parcellations of brain activity are prone to overlook the underlying manifold organization of fMRI, and that future studies should attempt to characterize and model brain states adopting the perspective of coupled neural activity modes.

Acknowledgments

Y.S.P is supported by European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant 896354. M.L.K. is supported by the Center for Music in the Brain, funded by the Danish National Research Foundation (DNRF117), and Centre for Eudaimonia and Human Flourishing at Linacre College funded by the Pettit and Carlsberg Foundations.

ET is supported by grants FONCYT-PICT (2019-02294), CONICET-PIP (11220210100800CO), and ANID/FONDECYT Regular (1220995). G.D. was supported by a number of sources: NEMESIS project (ref. 101071900) funded by the EU ERC Synergy Horizon Europe; AGAUR research support grant (ref. 2021 SGR 00917) funded by the Department of Research and Universities of the Generalitat of Catalunya; project PID2022-136216NB-I00 financed by the MCIN /AEI /10.13039/501100011033 / FEDER, UE., the Ministry of Science and Innovation, the State Research Agency and the European Regional Development Fund.

Author contributions

Y.S.P, M.L.K, E.T. and GD designed the research. Y.S.P, E.T. and GD conducted the research. Y.S.P., M.L.K.,G.D and E.T. analyzed and interpreted the results. M.L.K curate the data. Y.S.P, M.L.K., E.T.end GD wrote the manuscript and made figures. Y.S.P. analyzed the data. E.P.B & Y.S.P, created and published the code. G.D and E.T, supervised the research. All authors provided analytic support. All authors edited the manuscripts.

Methods

Participants

The neuroimaging dataset used in this investigation was obtained from the HCP (Human Connectome Project) March 2017 public data release. From the total of 1,003 participants available in the release, a sample of 1000 to the first part when only considered rest and 990 participants was chosen for the second part due to not all participants performing all tasks.

The HCP task battery comprises seven tasks, namely working memory, motor, gambling, language, social, emotional, and relational tasks. Detailed descriptions of these tasks can be found on the HCP website[62]. The tasks were intentionally designed by HCP to cover seven major cognitive domains, which aim to capture the diversity of neural systems: 1) Visual, motion, somatosensory, and motor systems; 2) Working memory, decision-making, and cognitive control systems; 3) Category-specific representations;4) Language processing; 5) Relational processing;6) Social cognition; 7) Emotion processing.

Apart from the resting-state scans, all 1,003 participants completed all tasks in two separate sessions. The first session included working memory, gambling, and motor tasks, while the second session involved language processing, social cognition, relational processing, and emotion processing tasks.

Neuroimaging extraction of functional time series

The processing and extracting data from both resting-state and task-based fMRI datasets is detailed in the HCP (Human Connectome Project). The preprocessing steps, described in detail on the HCP website, utilized standardized methods from FSL (FMRIB Software Library), FreeSurfer, and the Connectome Workbench software. The process included correcting spatial and gradient distortions, addressing head motion, intensity normalization, bias field removal, registration to the T1-weighted structural image, and transformation to the 2-mm Montreal Neurological Institute (MNI) space. The FIX artefact removal procedure [63,64]and ICA+FIX processing[65,66] were employed to remove structured artefacts and head motion parameters, respectively.

The preprocessed time series data of all grayordinates were in HCP CIFTI grayordinates standard space and available in surface-based CIFTI files for each participant during resting-state and each of the seven tasks. A custom-made MATLAB script using the ft_read_cifti function from the FieldTrip toolbox was used to extract the average time series for all the grayordinates in each region of the each

parcellation used, as defined in the HCP CIFTI grayordinates standard space. The BOLD timeseries were filtered using a second-order Butterworth filter with a frequency range of 0.008–0.08Hz and extracted the first and last 5% volumes to avoid the borders effect of the filtering.

Parcellations

The neuroimaging data underwent processing using the three standard parcellations with different amount of brain cortical regions: i) the Mindboggle-modified Desikan-Killiany parcellation, called DK62 [36]; ii) the Schaefer considering 500 regions; and iii) the Schaefer parcellation considering 1000 regions This DK62 parcellation consists of a total of 62 cortical regions, with 31 regions in each hemisphere, as defined by Klein and Tourville in 2012. Consequently, the DK80 parcellation contains a total of 62 regions, precisely defined in the common HCP CIFTI grayordinates standard space. Then, the Schaefer 500 and 1000 parcellation was also used with 250 and 500 cortical regions per hemisphere respectively, and well defined in the HCP CIFTI. In this parcellation[37]

Autoencoders architecture and training

We trained the same autoencoder using three different time-by-time spatial patterns of dimension N=62, 500 or 1000 as the inputs. We created random splits of 90% for training (900 participants) and 10% for testing (100 participants) to obtain separate training and test sets. Then, we trained the AE using a 10-fold cross validation scheme to evaluate the performance of the network. The AE architecture is shown in **Figure 1A**, upper panel. Our encoder network was comprised of a deep neural network that employs rectified linear units as activation functions and consisted of 6 fully connected layers starting with N=1024. This network funnels down to the latent space layer, whose dimensions ranged from 2 to 20. On the other hand, the decoder network mimics the architecture of the encoder, generating reconstructed patters from each encoded point (see Methods for more details of AE architecture). For each latent dimension we evaluated the performance of the AE in reconstructing each spatial pattern by measuring the mean squared error (MSE) between the input and the reconstructed pattern. We computed the average MSE across patterns and participants in each fold as well as in each parcellation.

Variational Autoencoders architecture and training

The VAE consists of the same three structures as the AE, but it incorporates certain differences that give VAEs generative properties. To guarantee this feature, during training, errors were propagated through gradient descent to minimize a loss function consisting of two terms. The first term is a standard reconstruction error, computed from the units in the output layer of the decoder. The second term is a regularization term, calculated as the Kullback-Leibler divergence between the distribution in the latent space and a standard Gaussian distribution. This regularization term ensures continuity and completeness in the latent space, meaning that similar values are decoded into similar outputs and that these outputs represent meaningful combinations of the encoded inputs. Our encoder network is composed of a deep neural network with rectified linear units as activation functions and two dense layers. This network funnels down to a low-dimensional variational layer, spanning the latent space explored from dimensions 5 to 12. The encoder network applies a nonlinear transformation to map the values of each time pattern to Gaussian probability distributions within the latent space. Conversely, the decoder network replicates the architecture of the encoder, generating reconstructed temporal patterns by sampling from these distributions. We trained the VAE with time-by-time spatial patterns for the training set creating from the 90%/10% (training/testing) resting state HCP participants full dataset. The network training procedure consisted of batches with 128 samples and 50 training epochs using an Adam optimizer and the training consists of error backpropagation via gradient descent to minimize a loss function composed the two terms.

In the last section of this work, we extended the framework by including to resting state (REST) data the recordings from 7 cognitive tasks from HCP (Social [SOC], Language [LAN], Working Memory [WM], Motor [MOT], Gambling [GAM], Emotion [EMO], Relational [REL]) to find the low-dimensional manifold that represent the dynamics which supports cognitive processing. We repeated the procedure as explained above, but in this case, we train the VAE using the time-by-time spatial patterns from the full data by concatenating the rest condition and the 7 tasks. We fixed the latent dimensions to 9 and we used the DK62 parcellation based on our previous results to build models in the latent space.

Modes coupled model and LGEC inference

We modelled the local dynamics of each latent neural mode as a Stuart-Landau oscillator (a dynamical system that can be expressed as the normal form of a supercritical Hopf bifurcation), this model

presents the capability to shift from noisy to oscillatory dynamics. Importantly, this kind of applied to the whole-brain dynamics have been able to replicate key aspects of brain dynamics observed in electrophysiology, magnetoencephalography[67] and fMRI[31,32,41]. Specifically, given a parcellation that number M of modes, the whole-brain dynamics of the coupled modes can be expressed by the local dynamics of M Stuart-Landau oscillators connected via the connectivity matrix $\boldsymbol{C}$, which is defined by:

$$\frac{dz_j}{dt} = \left(a_j + \mathrm{i}\omega_j\right)z_j - \left|z_j\right|^2 z_j + \sum_{k=1}^{N} C_{jk}\left(z_k - z_j\right) + \eta_j, \tag{1}$$

where the complex variable z_j denotes the state ($z_j = x_j + \mathrm{i}y_j$) of mode j, η_j is additive uncorrelated Gaussian noise with variance σ^2 (for all j), ω_j is the intrinsic mode frequency. The intrinsic frequencies ω_j (which lie in the 0.008–0.08Hz band) were estimated from the empirical data as the averaged peak frequencies of the narrowband the signals of the different modes. Finally, the mode's bifurcation parameter is a_j which defines the dynamics of the mode with respect to the Hopf bifurcation. In particular, this bifurcation is characterised by bifurcation parameter (a) that induces qualitative changes in the dynamical behaviour when the system crosses the bifurcation point ($a = 0$). In the supercritical regime ($a > 0$,), the oscillator generates self-sustained oscillations, whereas in the subcritical regime ($a < 0$), the dynamics converge to a fixed point. Importantly, the addition of noise to the model causes the dynamics near the bifurcation point ($a \sim 0$) to stochastically transition between both regimes, resulting in oscillations with complex amplitude modulations. Previous studies have demonstrated that this regime represents the optimal point for fitting empirical data in whole-brain models [31,32,41]. We fix the bifurcation parameter of each mode in that regime ($a = -0.02$).

We fit $\boldsymbol{C}$ such that the model optimally reproduces the empirically correlation matrix in the latent space $\boldsymbol{FC}_L{}^{Emp}$ (i.e., the pairwise correlation between the latent modes) and the empirical forward time-shifted correlation between the latent modes $\boldsymbol{FCf}_L{}^{Emp}(\tau)$. We selected the parameter $\tau = 3$ which led to a decrease in the averaged autocorrelation also based on previous works[68]. We note that fitting the time-shifted correlations can lead to asymmetries in the connectivity $\boldsymbol{C}$, which, in turn, can produce non-equilibrium dynamics. Using a heuristic pseudo-gradient algorithm, we proceeded to update the $\boldsymbol{C}$ until the fit is fully optimised. More specifically, the updating uses the following form:

$$C_{ij} = C_{ij} + \varsigma\left(FC_{L_{ij}}^{Emp} - FC_{L_{ij}}^{model}\right)$$

$$+\varsigma\left(FCf_{L_{ij}}^{empirical}(\tau) - FCf_{L_{ij}}^{model}(\tau)\right), \quad (2)$$

The model was run repeatedly with the updated $\boldsymbol{C}$ until the fit converges towards a stable value. We initialized $\boldsymbol{C}$ as a zero matrix of dimension MxM (the number of latent modes) and we update all the connections from this matrix (in either hemisphere). For the Stuart-Landau model, we used $\varsigma = 0.00001$ and continue until the algorithm converges. We generate 100 model by repeating this procedure to fit the average empirical matrices, which was obtained by averaging set of 10 (9) participants for the resting state (resting state plus tasks). In this way we manage to improve the fitting of the models. In each model, for each iteration we average over 10 (9) simulations as we made for the empirical data. Overall, we use the term Latent Generative Effective Connectivity (LGEC) for the optimised $\boldsymbol{C}$, which reflects the connectivity between the modes.

Obtaining sources networks by decoding the latent modes.

We created a set of surrogate signals by introducing standard Gaussian noise into one latent dimension while keeping the other eight dimensions devoid of any signal. We repeated this process, varying the dimension with noise and the dimensions without any signal, and we called M "N" with N from 1 to 9 to each one. Subsequently, we decoded each of these nine sets of surrogate latent signals, mapping them from the latent dimensions to the 62 brain regions of the original empirical data for each case. Next, we examined the regions in the source space that were activated in response to the presence of noise in each latent dimension. We quantified the variance of the decoded time signals to evaluate the statistical response of the activity associated with each latent dimension. This analysis allowed us to obtain one spatial pattern in the empirical original space associated to each latent dimension.

Support vector machine classifier

We use a SVM with Gaussian kernels as implemented in the MATLAB function fitcecoc. The function returns a full, trained, multiclass, error-correcting output codes (ECOC) model. This is achieved using the predictors in the input with class labels. The function uses K(K − 1)/2 binary SVM models using the one-versus-one coding design, where we used K = 8 as the number of unique class labels. In other words, the SVM had as inputs the elements of the LGEC obtained in each of the 100 models for each one of the 7 tasks and the resting-state condition. The output was eight classes corresponding to the

conditions (rest and seven tasks). We subdivided into 90% training and 10% validation, repeated and shuffled 100 times. For the classification using the functional connectivity in the original source space, we average the FC across 9 participants, as we did with the LGEC computation, and use as an input the upper diagonal elements of the matrices.

Statistical analyses

We applied the Wilcoxon rank-sum method to test the significance and we applied the False Discovery Rate (FDR) at the 0.05 level of significance to correct multiple comparisons[69].

Reference

1. Sporns, O., Tononi, G. & Kötter, R. The Human Connectome: A Structural Description of the Human Brain. *PLOS Comput. Biol.* **1**, e42 (2005).
2. Deco, G., Jirsa, V. K. & McIntosh, A. R. Emerging concepts for the dynamical organization of resting-state activity in the brain. *Nat. Rev. Neurosci.* **12**, 43–56 (2011).
3. Deco, G., Jirsa, V. K., Robinson, P. A., Breakspear, M. & Friston, K. The Dynamic Brain: From Spiking Neurons to Neural Masses and Cortical Fields. *PLOS Comput. Biol.* **4**, e1000092 (2008).
4. Biswal, B., Zerrin Yetkin, F., Haughton, V. M. & Hyde, J. S. Functional connectivity in the motor cortex of resting human brain using echo planar mri. *Magn. Reson. Med.* **34**, 537–541 (1995).
5. Damoiseaux, J. S. *et al.* Consistent resting-state networks across healthy subjects. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 13848–13853 (2006).
6. Beckmann, C. F., DeLuca, M., Devlin, J. T. & Smith, S. M. Investigations into resting-state connectivity using independent component analysis. *Philos. Trans. R. Soc. B Biol. Sci.* **360**, 1001–1013 (2005).
7. Chaudhuri, R., Gerçek, B., Pandey, B., Peyrache, A. & Fiete, I. The intrinsic attractor manifold and population dynamics of a canonical cognitive circuit across waking and sleep. *Nat. Neurosci. 2019 229* **22**, 1512–1520 (2019).
8. Mitchell-Heggs, R., Prado, S., Gava, G. P., Go, M. A. & Schultz, S. R. Neural manifold analysis of brain circuit dynamics in health and disease. *J. Comput. Neurosci. 2022 511* **51**, 1–21 (2022).
9. Perl, Y. S. *et al.* Low-dimensional organization of global brain states of reduced consciousness. *Cell Rep.* **42**, 112491 (2023).

10. Luppi, A. I. *et al.* Distributed harmonic patterns of structure-function dependence orchestrate human consciousness. *Commun. Biol. 2023 61* **6**, 1–19 (2023).
11. Rué-Queralt, J. *et al.* Decoding brain states on the intrinsic manifold of human brain dynamics across wakefulness and sleep. *Commun. Biol. 2021 41* **4**, 1–11 (2021).
12. Gao, S., Mishne, G. & Scheinost, D. Nonlinear manifold learning in functional magnetic resonance imaging uncovers a low-dimensional space of brain dynamics. *Hum. Brain Mapp.* **42**, 4510–4524 (2021).
13. Vohryzek, J. *et al.* Harmonic decomposition of spacetime (HADES) framework characterises the spacetime hierarchy of the DMT brain state. *bioRxiv* 2023.08.20.554019 (2023) doi:10.1101/2023.08.20.554019.
14. Glomb, K. *et al.* Functional harmonics reveal multi-dimensional basis functions underlying cortical organization. *Cell Rep.* **36**, 109554 (2021).
15. Shine, J. M. *et al.* The Dynamics of Functional Brain Networks: Integrated Network States during Cognitive Task Performance. *Neuron* **92**, 544–554 (2016).
16. Tononi, G., Boly, M., Massimini, M. & Koch, C. Integrated information theory: from consciousness to its physical substrate. *Nat. Rev. Neurosci. 2016 177* **17**, 450–461 (2016).
17. Tononi, G., Edelman, G. M. & Sporns, O. Complexity and coherency: integrating information in the brain. *Trends Cogn. Sci.* **2**, 474–484 (1998).
18. Hennig, J. A. *et al.* Constraints on neural redundancy. *Elife* **7**, (2018).
19. Shine, J. M. *et al.* Human cognition involves the dynamic integration of neural activity and neuromodulatory systems. *Nat. Neurosci. 2019 222* **22**, 289–296 (2019).
20. Tononi, G., Sporns, O. & Edelman, G. M. Measures of degeneracy and redundancy in biological networks. *Proc. Natl. Acad. Sci.* **96**, 3257–3262 (1999).
21. Pang, R., Lansdell, B. J. & Fairhall, A. L. Dimensionality reduction in neuroscience. *Curr. Biol.* **26**, R656--R660 (2016).
22. Pillai, A. S. & Jirsa, V. K. Symmetry Breaking in Space-Time Hierarchies Shapes Brain Dynamics and Behavior. *Neuron* **94**, 1010–1026 (2017).
23. Jirsa, V. & Sheheitli, H. Entropy, free energy, symmetry and dynamics in the brain. *J. Phys. Complex.* **3**, 015007 (2022).
24. Jirsa, V. Structured Flows on Manifolds as guiding concepts in brain science. *Selbstorganisation – ein Paradig. für die Humanwissenschaften* 89–102 (2020) doi:10.1007/978-3-658-29906-4_6.
25. Fousek, J. *et al.* Symmetry breaking organizes the brain's resting state manifold. *Cold Spring*

Harb. Lab. Press (2023) doi:10.1101/2022.01.03.474841.

26. Gallego, J. A., Perich, M. G., Miller, L. E. & Solla, S. A. Neural manifolds for the control of movement. *Neuron* **94**, 978–984 (2017).
27. Stringer, C., Pachitariu, M., Steinmetz, N., Carandini, M. & Harris, K. D. High-dimensional geometry of population responses in visual cortex. *Nature* **571**, 361–365 (2019).
28. Chandak, R. & Raman, B. Neural manifolds for odor-driven innate and acquired appetitive preferences. *Nat. Commun.* **14**, 4719 (2023).
29. Song, H., Shim, W. M. & Rosenberg, M. D. Large-scale neural dynamics in a shared low-dimensionalstate space reflect cognitive and attentional dynamics. *Elife* **12**, e85487 (2023).
30. Derdikman, D. & Moser, E. I. A manifold of spatial maps in the brain. *Trends Cogn. Sci*. **14**, 561–569 (2010).
31. Perl, Y. S. *et al.* Strength-dependent perturbation of whole-brain model working in different regimes reveals the role of fluctuations in brain dynamics. *PLOS Comput. Biol*. **18**, e1010662 (2022).
32. Deco, G., Kringelbach, M. L., Jirsa, V. K. & Ritter, P. The dynamics of resting fluctuations in the brain: metastability and its dynamical cortical core. *Sci. Rep.* **7**, 3095 (2017).
33. Jobst, B. M. *et al.* Increased Stability and Breakdown of Brain Effective Connectivity During Slow-Wave Sleep: Mechanistic Insights from Whole-Brain Computational Modelling. *Sci. Reports 2017 71* **7**, 1–16 (2017).
34. Kringelbach, M. L., Perl, Y. S., Tagliazucchi, E. & Deco, G. Toward naturalistic neuroscience: Mechanisms underlying the flattening of brain hierarchy in movie-watching compared to rest and task. *Sci. Adv.* **9**, (2023).
35. Deco, G. *et al.* Dynamical consequences of regional heterogeneity in the brain's transcriptional landscape. *Sci. Adv.* **7**, eabf4752 (2021).
36. Desikan, R. S. *et al.* An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* **31**, 968–980 (2006).
37. Schaefer, A. *et al.* Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cereb. Cortex* **28**, 3095–3114 (2018).
38. Kingma, D. P. & Welling, M. Auto-Encoding Variational Bayes. *2nd Int. Conf. Learn. Represent. ICLR 2014 - Conf. Track Proc.* (2013).
39. Thomas Yeo, B. T. *et al.* The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, 1125–1165 (2011).

40. Fox, M. D. & Greicius, M. Clinical applications of resting state functional connectivity. *Front. Syst. Neurosci.* **4**, 1443 (2010).
41. Ipiña, I. P. *et al.* Modeling regional changes in dynamic stability during sleep and wakefulness. *Neuroimage* **215**, 116833 (2020).
42. Perl, Y. S. *et al.* Perturbations in dynamical models of whole-brain activity dissociate between the level and stability of consciousness. *PLOS Comput. Biol.* **17**, e1009139 (2021).
43. Huang, Z., Zhang, J., Wu, J., Mashour, G. A. & Hudetz, A. G. Temporal circuit of macroscale dynamic brain activity supports human consciousness. *Sci. Adv.* **6**, 87–98 (2020).
44. Carbonell, F., Bellec, P. & Shmuel, A. Global and system-specific resting-state fMRI fluctuations are uncorrelated: principal component analysis reveals anti-correlated networks. *Brain Connect.* **1**, 496–510 (2011).
45. Smith, S. M. *et al.* Correspondence of the brain's functional architecture during activation and rest. *Proc. Natl. Acad. Sci.* **106**, 13040–13045 (2009).
46. Liu, X. & Duyn, J. H. Time-varying functional network information extracted from brief instances of spontaneous brain activity. *Proc. Natl. Acad. Sci.* **110**, 4392–4397 (2013).
47. Greicius, M. Resting-state functional connectivity in neuropsychiatric disorders. *Curr. Opin. Neurol.* **21**, 424–430 (2008).
48. Cole, M. W., Ito, T., Bassett, D. S. & Schultz, D. H. Activity flow over resting-state networks shapes cognitive task activations. *Nat. Neurosci.* **19**, 1718–1726 (2016).
49. Mantini, D., Perrucci, M., Gratta, C. Del, Romani, G. & Corbetta, M. Electrophysiological signatures of resting state networks in the human brain. *Proc Natl Acad Sci USA* **104**, 13170–13175 (2007).
50. Chung, S. & Abbott, L. F. Neural population geometry: An approach for understanding biological and artificial neural networks. *Curr. Opin. Neurobiol.* **70**, 137–144 (2021).
51. Nieh, E. H. *et al.* Geometry of abstract learned knowledge in the hippocampus. *Nature* **595**, 80–84 (2021).
52. Gallego, J. A., Perich, M. G., Chowdhury, R. H., Solla, S. A. & Miller, L. E. Long-term stability of cortical population dynamics underlying consistent behavior. *Nat. Neurosci.* **23**, 260–270 (2020).
53. Langdon, C., Genkin, M. & Engel, T. A. A unifying perspective on neural manifolds and circuits for cognition. *Nat. Rev. Neurosci.* 1–15 (2023).
54. Damoiseaux, J. S. *et al.* Consistent resting-state networks across healthy subjects. *Proc. Natl.*

Acad. Sci. **103**, 13848–13853 (2006).

55. Logothetis, N. K. What we can do and what we cannot do with fMRI. *Nature* **453**, 869–878 (2008).
56. Poldrack, R. A. & Yarkoni, T. From brain maps to cognitive ontologies: informatics and the search for mental structure. *Annu. Rev. Psychol.* **67**, 587–612 (2016).
57. Markello, R. D. *et al.* neuromaps: structural and functional interpretation of brain maps. *Nat. Methods 2022* 1–8 (2022) doi:10.1038/s41592-022-01625-w.
58. Hansen, J. Y. *et al.* Mapping neurotransmitter systems to the structural and functional organization of the human neocortex. *Nat. Neurosci. 2022 2511* **25**, 1569–1581 (2022).
59. Schiff, N. D., Nauvel, T. & Victor, J. D. Large-scale brain dynamics in disorders of consciousness. *Curr. Opin. Neurobiol.* **25**, 7–14 (2014).
60. Brown, E. N., Lydic, R. & Schiff, N. D. General Anesthesia, Sleep, and Coma. *N. Engl. J. Med.* **363**, 2638–2650 (2010).
61. Varley, T. F., Denny, V., Sporns, O. & Patania, A. Topological analysis of differential effects of ketamine and propofol anaesthesia on brain dynamics. *R. Soc. open Sci.* **8**, (2021).
62. Barch, D. M. *et al.* Function in the human connectome: Task-fMRI and individual differences in behavior. *Neuroimage* **80**, 169–189 (2013).
63. Schröder, T. N., Haak, K. V., Jimenez, N. I. Z., Beckmann, C. F. & Doeller, C. F. Functional topography of the human entorhinal cortex. *Elife* **4**, 1–17 (2015).
64. Smith, S. M. *et al.* Resting-state fMRI in the Human Connectome Project. *Neuroimage* **80**, 144–168 (2013).
65. Griffanti, L. *et al.* Hand classification of fMRI ICA noise components. *Neuroimage* **154**, 188–205 (2017).
66. Arslan, S. *et al.* Human brain mapping: A systematic comparison of parcellation methods for the human cerebral cortex. *Neuroimage* **170**, 5–30 (2018).
67. Cabral, J., Kringelbach, M. L. & Deco, G. Exploring the network dynamics underlying brain activity during rest. *Prog. Neurobiol.* **114**, 102–131 (2014).
68. Deco, G., Sanz Perl, Y., Bocaccio, H., Tagliazucchi, E. & Kringelbach, M. L. The INSIDEOUT framework provides precise signatures of the balance of intrinsic and extrinsic dynamics in brain states. *Commun. Biol. 2022 51* **5**, 1–13 (2022).
69. Hochberg, Y. & Benjamini, Y. More powerful procedures for multiple significance testing. *Stat. Med.* **9**, 811–818 (1990).