# Connectome-Based Attractor Dynamics Underlie Brain Activity in Rest, Task, and Disease

#### Robert Englert, Balint Kincses, Raviteja Kotikalapudi, Giuseppe Gallitto, Jialin Li, Kevin Hoffschlag, Choong-Wan Woo, Tor D. Wager, Dagmar Timmann, Ulrike Bingel, Tamas Spisak

## Introduction

Brain function is characterized by the continuous activation and deactivation of anatomically distributed neuronal
populations (Buzsaki, 2006).
Irrespective of the presence or absence of explicit stimuli, brain regions appear to work in concert, giving rise to a
rich and spatiotemporally complex fluctuation (Bassett & Sporns, 2017).
This fluctuation is neither random nor stationary over time (Liu & Duyn, 2013; Zalesky *et al.*, 2014).
It is organized around large-scale gradients (Margulies *et al.*, 2016; Huntenburg *et al.*, 2018) and exhibits quasi-periodic properties, with a limited number of recurring patterns known as “brain states” (Greene *et al.*, 2023; Vidaurre *et al.*, 2017; Liu & Duyn, 2013).

A wide variety of descriptive techniques have been previously employed to characterize whole-brain dynamics (Smith *et al.*, 2012; Vidaurre *et al.*, 2017; Liu & Duyn, 2013; Chen *et al.*, 2018).
These efforts have provided accumulating evidence not only for the existence of dynamic brain states but also for their clinical
significance (Hutchison *et al.*, 2013; Barttfeld *et al.*, 2015; Meer *et al.*, 2020).
However, the underlying driving forces remain elusive due to the descriptive nature of such studies.

Conventional computational approaches attempt to solve this puzzle by going all the way down to the biophysical properties of single neurons, and aim to construct a model of larger neural populations, or even the entire brain
(Breakspear, 2017).
These approaches have shown numerous successful applications (Murray *et al.*, 2018; Kriegeskorte & Douglas, 2018; Heinz *et al.*, 2019).
However, such models need to estimate a vast number of neurobiologically motivated free parameters to fit the data. This hampers their ability to effectively bridge the gap between explanations at the level of single neurons and the complexity of behavior (Breakspear, 2017).
Recent efforts using coarse-grained brain network models (Schirner *et al.*, 2022; Schiff *et al.*, 1994; Papadopoulos *et al.*, 2017; Seguin *et al.*, 2023) and linear network control theory (Chiêm *et al.*, 2021; Scheid *et al.*, 2021; Gu *et al.*, 2015) opted to trade biophysical fidelity to phenomenological validity.

Such models have provided insights into some of the inherent key characteristics of the brain as a dynamic system; for instance, the importance of stable patterns, so-called “attractor states”, in governing brain dynamics (Deco *et al.*, 2012; Golos *et al.*, 2015; Hansen *et al.*, 2015). While attractor networks have become established models of micro-scale canonical brain circuits in the last four decades (Khona & Fiete, 2022), these studies highlighted that attractor dynamics are essential characteristics of macro-scale brain dynamics as well. However, the standard practice among these studies is the use of models that capitalize on information about the structural wiring of the brain, leading to the grand challenge of modeling the relationship between the structural wiring of the brain and functional connectivity.

The “neuroconnectionist” approach (Doerig *et al.*, 2023) makes another step towards trading biophisical detail to “cognitive/behavioral fidelity” (Kriegeskorte & Douglas, 2018), by using artificial neural networks (ANNs) that are trained to perform various tasks, as brain models. However, the need to train ANNs for specific tasks inherently limits their ability to explain task-independent, spontaneous neural dynamics (Richards *et al.*, 2019).

Here we propose a minimal phenomenological model for large-scale brain dynamics that combines the advantages of large-scale attractor network models (Golos *et al.*, 2015), neuroconnectionism (Doerig *et al.*, 2023), and recent advances in undersanding the flow of brain activity across regions (Cole *et al.*, 2016), to investigate brain dynamics.
Similar to neuroconnectionism, we utilize an ANN as an abstract, high-level computational model of the brain.
However, our model is not explicitly trained for a specific task. Instead, we set its weights empirically.

Specifically, we employ a continuous-space Hopfield Neural Network (HNN) (Hopfield, 1982; Krotov, 2023), similar to the spin-glass and Hopfield-style attractor network models applied e.g. by Deco *et al.* (2012)Golos *et al.* (2015), where the nodes of the network model represent large-scale brain areas. However, in contrast to these previos efforts that starting from the structural wiring of the brain, we initialize the edge weights of the network based on direct estimates node-to-node information transfer. Our decision to employ a direct proxy of interregional communication, rather than biophisical wiring, capitalizes on the “activity flow” (Cole *et al.*, 2016; Ito *et al.*, 2017) principle, a toroughgly validated phenomenological model of the association between brain activity and functional connectivity, as measuured with functional magneic resonance imaging.
This allows us to circumvent the necessity of comprehensively understanding and accurately modeling structural-functional coupling in the brain. Instead, we can concentrate on the overarching dynamical properties of the system.

Based on the topology of the functional connectome, our model establishes an energy level for any arbitrary activation patterns and determines a “trajectory of least action” towards one of the finite number of stable patterns, known as *attractor states*, that minimize this energy. In the proposed framework, macro-scale brain dynamics can be conceptualized as an intricate, high-dimensional path on the energy landscape (Figure 1C), arising from the activity flow (Cole *et al.*, 2016) within the functional connectome and constrained by the “gravitational pull” of the attractor states of the system.
The generative nature of the proposed framework offers testable predictions for the effect of various perturbations and alterations of these dynamics, from task-induced activity to changes related to brain disorders.



**Figure 1:** **Connectome-based Hopfield networks as models of macro-scale brain dynamics.**   
**A** Hopfield artificial neural networks (HNNs) are a form of recurrent artificial neural networks that serve as content-addressable (“associative”) memory systems. Hopfield networks can be trained to store a finite number of patterns (e.g. via Hebbian learning a.k.a. “fire together - wire together”). During the training procedure, the weights of the HNN are trained so that the stored
patterns become stable attractor states of the network. Thus, when the trained network is presented partial, noisy or corrupted variations of the stored patterns, it can effectively reconstruct the original pattern via an iterative relaxation procedure that converges to the attractor states.
**B** We consider regions of the brain as nodes of a Hopfield network. Instead of initilaizing the network with the structural wiring of the brain or training it to solve specific tasks, we set its weights empirically, using information about the interregional “activity flow” across regions, as estimated via functional brain connectivity. Capitalizing on strong analogies between the relaxation rule of Hopfield networks and the activity flow principle that links activity to connectivity in brain networks, we propose the resulting
functional connectome-based Hopfield neural network (fcHNN) as a minimal phenomenological model for macro-scale brain dynamics. **C** The proposed computational framework assigns an energy level, an attractor state and a position in a
low-dimensional embedding to brain activation patterns. Additionally, it models how the entire state-space of viable activation patterns is restricted by the dynamics of the system and how alterations in activity and/or connectivity modify these dynamics.

In the present work, we investigate how well the functional connectome is suited to be an attractor network, map the corresponding attractoir states and model itinerant stochastic dynamics traversing the different basins of attraction of the system.
We use a diverse set of experimental, clinical and meta-analytic studies to evaluate our model’s ability to reconstruct various characteristics of resting state brain dynamics, as well as its capacity to detect and explain changes induced by experimental conditions or alterations in brain disorders.

## Results

### Connectome-based Hopfield network as a model of brain dynamics

First, we investigated the functional connectome as an attractor network in a sample of n=41 healthy young
participants (study 1, see Methods Table 1 for details). We estimated interregional activity flow (Cole *et al.*, 2016; Ito *et al.*, 2017) as the study-level average of regularized partial correlations among the resting state fMRI timeseries of m = 122
functionally defined brain regions (see Methods for details). We then used the standardized functional connectome as the weights of a fully connected recurrent ANN, specifically a continuous-state Hopfield network (Hopfield, 1982; Koiran, 1994), consisting of neural units, each having an activity . Hopfield networks can be initialized by an arbitrary activation pattern (consisting of activation values) and iteratively updated (i.e. “relaxed”) until their energy converges a local minimum, that is, to one of the finite number of attractor states (see Methods). The relaxation procedure is based on a simple rule; in each iteration, the activity of a region is constructed as the weighted average of the activities of all other regions, with weights defined by the connectivity between them. The average is then transformed by a sigmoidal activation function, to keep it in the desired [-1,1] interval.
This can be expressed by the following equation:

(1)

where is the activity of neural unit in the next iteration and is the sigmoidal activation
function ( in our implementation) and is the bias of unit and β is the so-called temperature parameter. For the sake of simplicity, we set in all our experiments. We refer to this architecture as a functional connectivity-based Hopfield Neural Network (fcHNN).
The relaxation of a fcHNN model can be conceptualized as the repeated application of the activity flow principle (Cole *et al.*, 2016; Ito *et al.*, 2017) , simultaneously for all regions: . The update rule also exhibits analogies with network control theory (Gu *et al.*, 2015) and the inner workings of neural mass models, as applied e.g. in dynamic causal modeling (Daunizeau *et al.*, 2012).

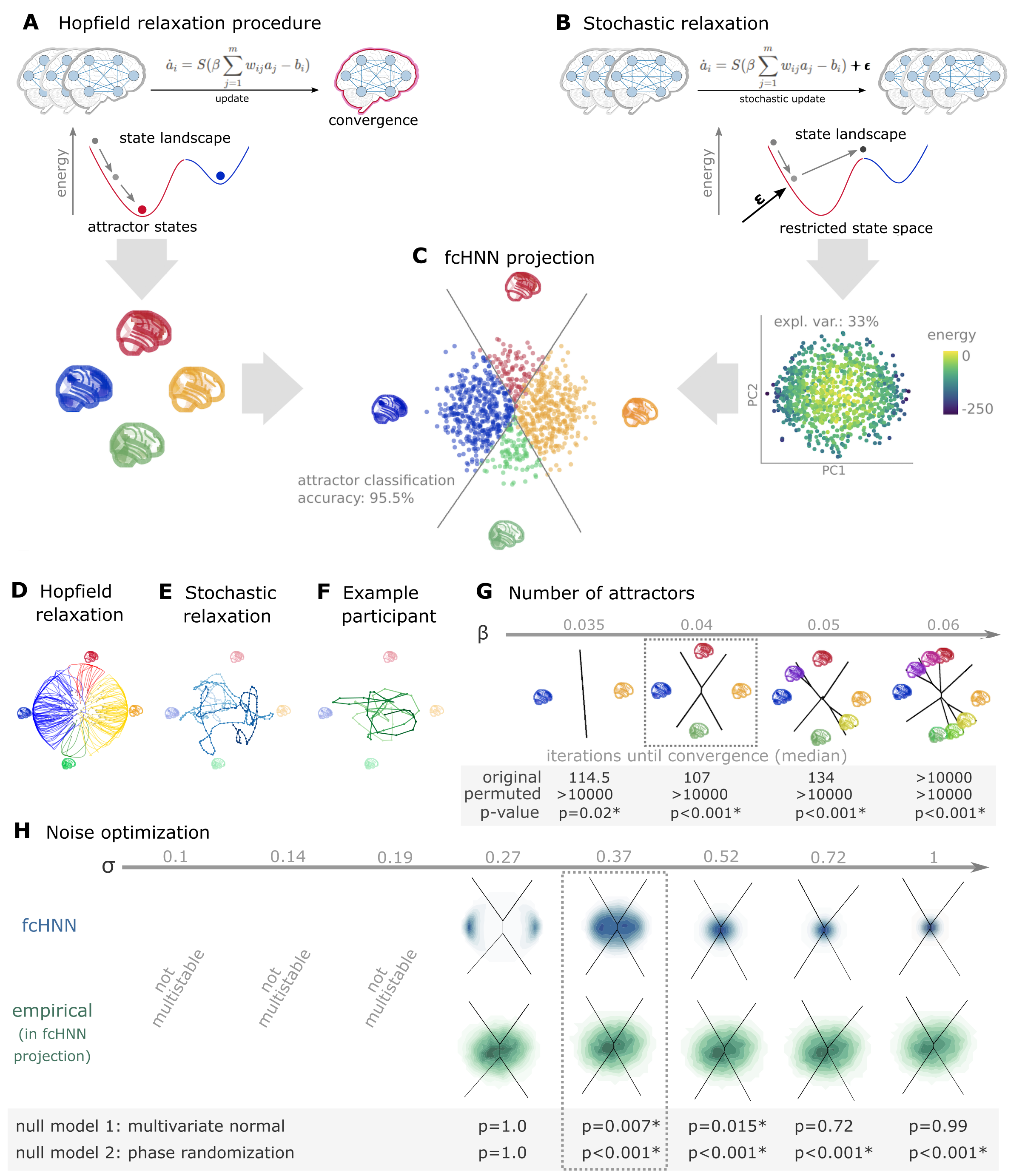
Hopfield networks assign an energy value to each possible activity configuration (Hopfield, 1982; Koiran, 1994), which decreases during the relaxation procedure until reaching an equilibrium state with minimal energy (Figure 2A, top panel).

We used a sufficiently large number of random initializations (n=100000) to obtain all possible attractor states of the connectome-based Hopfield network in study 1 (Figure 2A, bottom panel).
Consistent with theoretical expectations, we observed that increasing the temperature parameter β led to an
increasing number of attractor states (Figure 2E, left, Supplementary Figure 3), appearing in symmetric pairs (i.e. , see Figure 2G).

FcHNNs, without any modifications, always converge to an equilibrium state.
To incorporate stochastic fluctuations in neuronal activity (Robinson *et al.*, 2005), we introduced weak
Gaussian noise to the fcHNN relaxation procedure. This procedure, referred to as stochastic relaxation, prevents the system from reaching equilibrium and, somewhat similarly to stochastic DCM (Daunizeau *et al.*, 2012), induces complex, heteroclinic system dynamics (Figure 2B).

In order to enhance interpretability, we obtained the first two principal components (PCs) of the states sampled from the stochastic relaxation procedure. On the low-dimensional embedding, which we refer to as the *fcHNN* projection, we observed a clear separation of the attractor states (Figure 2C), with the two symmetric pairs of attractor states located at the extremes of the first and second PC.
To map the attractors’ basins on the space spanned by the first two PCs (Figure 2C), we obtained the attractor state of each point visited during the stochastic relaxation and fit a multinomial logistic regression model to predict the attractor state from the first two PCs.
The resulting model accurately predicted attractor states of arbitrary brain activity patterns, achieving a cross-validated accuracy of 96.5% (permutation-based p<0.001).
The attractor basins were visualized by using the decision boundaries obtained from this model. (Figure 2C). We propose the 2-dimensional fcHNN projection depicted on (Figure 2C) as a simplified representation of brain dynamics, and use it as a basis for all subsequent analyses in this work.

Panel D on Figure 2 uses the fcHNN projection to visualize the conventional Hopfield relaxation procedure. It depicts the trajectory of individual activation maps (sampled randomly from the timeseries data in Study 1) until converging to one of the four attractor states.
Panel E shows that the system does not converge to an attractor state anymore if weak noise is introduced to the system (stochastic relaxation), The resulting path is still influenced by the attractor states’ gravity, resulting in a heteroclinic dynamics that resembles the empirical timeseries data (example data on panel F).



**Figure 2:** **Attractor states and state-space dynamics of connectome-based Hopfield networks**   
**A** Top: During so-called relaxation procedure, activities in the nodes of an fcHNN model are iteratively updated based on the activity of all other regions and the connectivity between them. The energy of a
connectome-based Hopfield network decreases during the relaxation procedure until reaching an equilibrium state with
minimal energy, i.e. an attractor state. Bottom: Four attractor states of the fcHNN derived from the
group-level functional connectivity matrix from study 1 (n=44).
**B** Top: In presence of weak noise (stochastic update), the system
does not converge to equilibrium anymore. Instead, activity traverses on the state landscape in a way
restricted by the topology of the connectome and the “gravitational pull” of the attractor states. Bottom: We sample
the “state landscape” by running the stochastic relaxation procedure for an extended amount of time (e.g. 100.000 consecutive
stochastic updates), each point representing an activation configuration or state. To construct a
low-dimensional representation of the state space, we take the first two principal components of the simulated activity
patterns. The first two principal components explain approximately 58-85% of the variance of state energy (depending
on the noise parameter σ, see Supplementary Figure 1).
**C** We map all states of the state space sample to their corresponding attractor state, with the conventional
Hopfield relaxation procedure (A). The four attractor states are also visualized in their corresponding position on the
PCA-based projection. The first two principal components yield a clear separation of the attractive state basins
(cross-validated classification accuracy: 95.5%, Supplementary Figure 2). We refer to the resulting visualization
as the fcHNN projection and use it to visualize fcHNN-derived and empirical brain dynamics throughout the rest of
the manuscript.
**D** The fcHNN of study 1 seeded with real activation maps (gray dots) of an example participant. All activation maps converge to one of the four attractor states during the relaxation procedure (without noise) and the system reaches equilibrium. Trajectories are colored by attractor state.
**E** Illustration of the stochastic relaxation procedure in the same fcHNN model, seeded from a single starting point (activation pattern). The system does not converge to an attractor state but instead traverses the state space in a way restricted by the topology of the connectome and the “gravitational pull” of the attractor states. The shade of the trajectory changes with increasing number of iterations. The trajectory is smoothed with a moving average over 10 iterations for visualization purposes.
**F** Real resting state fMRI data of an example participant from study 1, plotted on the fcHNN projection. The shade of the trajectory changes with an increasing number of iterations. The trajectory is smoothed with a moving average over 10 iterations for visualization purposes.
**G** Consistent with theoretical expectations, we observed that increasing the temperature parameter β led to an
increasing number of attractor states, emerging in a nested fashion (i.e. the basin of a new attractor state is fully contained within the basin of a previous one). When contrasting the functional connectome-based HNN with a null model based on symmetry-retaining permuted variations of the connectome, we found that the topology of the original (unpermuted) functional brain connectome makes it significantly better suited to function as an attractor network; than the permuted null model. Table contains the meadian number of iterations until convergence for the original and permuted connectomes for different temperature parameters β and the corresponding p-value.
**H** We optimized the noise parameter σ of the stochastic relaxation procedure for 8 different σ values over a logarithmic range between and 1 so that the similarity (the timeframes distribution over the attractor basins) is maximized between the empirical data and the fcHNN-generated data. We used to null models to assess the significance of similarity: one based on multivariate normal data, with the covariance matrix set to the functional connectome’s covariance matrix, and one based on spatial phase-randomization. P-values are given in the table at the bottom of the panel. The fcHNN only reached multistability with , and it provided the most accurate reconstruction of the real data with (p=0.007, and p<0.001 for the two null models).

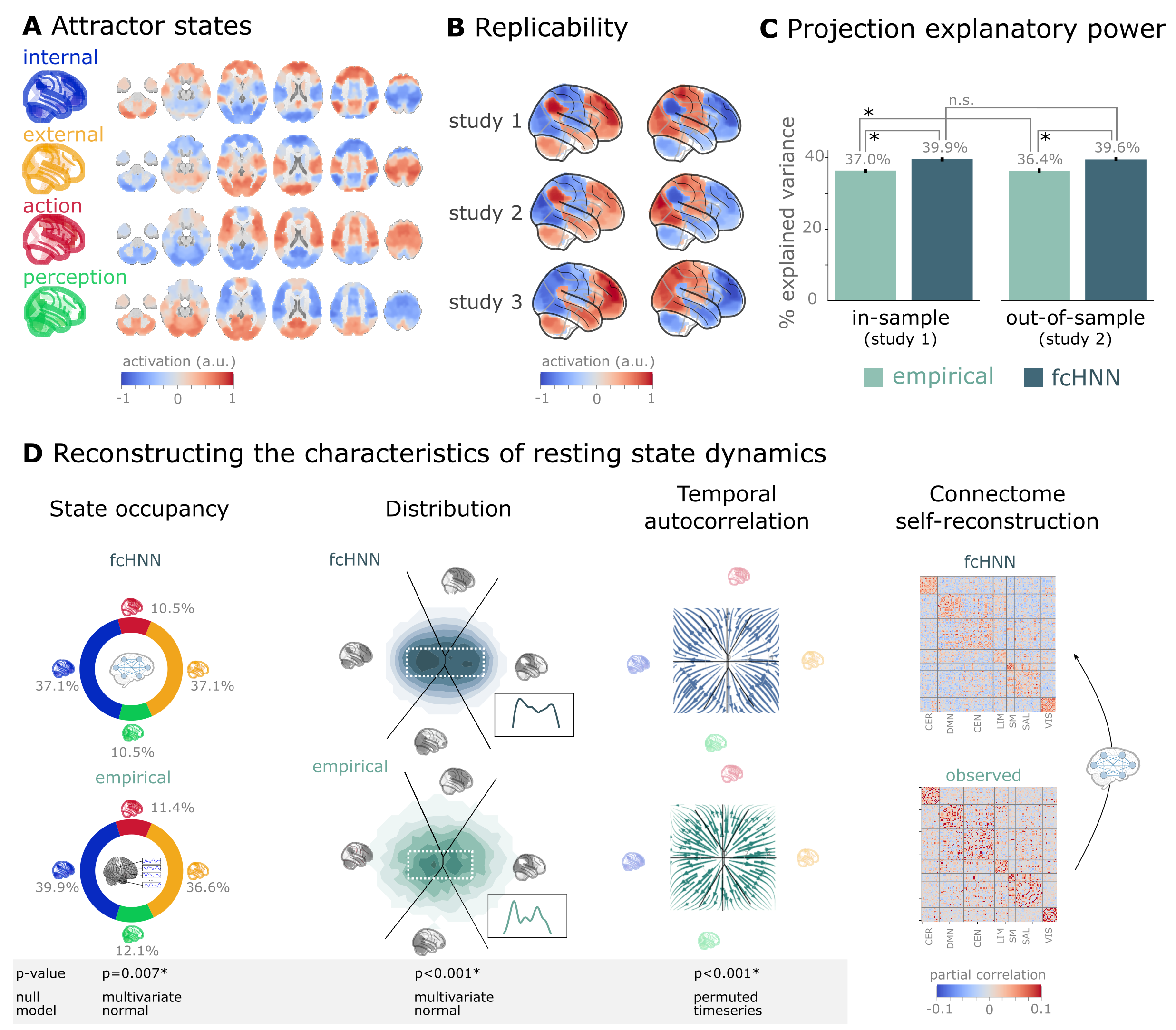
In study 1, we have investigated the convergence process of the functional connectivity-based HNN and contrasted it with a null model based on permuted variations of the connectome (while retaining the symmetry of the matrix). This null model preserves the sparseness and the degree distribution of the connectome, but destroys its topological structure (e.g. clusteredness). We found that the topology of the original (unpermuted) functional brain connectome makes it significantly better suited to function as an attractor network; than the permuted null model.
While the original connectome based HNN converged to an attractor state in less than 150 iterations in more than 50% of the cases, the null model did not reach convergence in more than 98% of the cases, even after 10000 iterations (Figure 2G, Supplementary Figure 4). This result was robustly observed, independent of the temperature parameter beta.
We set the temperature parameter for the rest of the paper to a value providing the fastest convergence (, median iterations: 107), resulting in 4 distinct attractor states. The primary motivation for selecting was to reduce computational burden for further analyses. However, as with increasing temperature, attractor states emerge in a nested fashion (i.e. the basin of a new attractor state is fully contained within the basin of a previous one), we expect that the results of the following analyses would be, although more detailed, but qualitatively similar with higher β values.

We optimized the noise parameter σ of the stochastic relaxation procedure for 8 different σ values over a logarithmic range between and 1 so that the similarity (the timeframes distribution over the attractor basins) is maximized between the empirical data and the fcHNN-generated data. We contrasted this similarity with two null-models (Figure 2H). First we generated null-data as random draws from a multivariate normal distribution with co-variance matrix set to the functional connectome’s covariance matrix (partial correlation-based connectivity estimates). This serves as a baseline for generating data that optimally matches the empirical data in terms of distribution and spatial autocorrelation, as based on information on the underlying co-variance structure (and given Gaussian assumptions), but without any mechanistic model of the generative process, e.g. without modelling any non-linear and non-Gaussian effects and temporal autocorrelations stemming from recurrent activity flow). We found that The fcHNN only reached multistability with , and it provided more accurate reconstruction of the real data than the null model for (p=0.007 and 0.015, 11.16, respectively).
With our second null model, we investigated whether the fcHNN-reconstructed data is more similar to the empirical data than synthetic data with identical distribution and spatial autocorrelation structure (generated by spatial phase randomization of the original volumes, see Methods).
We found that the fcHNNs significantly outperform this null model with all investigated σ values if
Based on this coarse optimization procedure, we set for all subsequent analyses.

### Reconstruction of resting state brain dynamics

The spatial patterns of the obtained attractor states exhibit high neuroscientific relevance and closely resemble previously described large-scale brain systems. (Figure 3A).

The first pair of attractors (mapped on PC1, horizontal axis) represent two complementary brain systems, that have been previously found in anatomical, functional, developmental, and evolutionary hierarchies, as well as gene expression, metabolism, and blood flow, (see Sydnor *et al.* (2021) for a review), an reported under various names, like intrinsic and extrinsic systems (Golland *et al.*, 2008), Visual-Sensorimotor-Auditory and Parieto-Temporo-Frontal “rings” (Cioli *et al.*, 2014), “primary” brain states (Chen *et al.*, 2018), unimodal-to-transmodal principal gradient (Margulies *et al.*, 2016; Huntenburg *et al.*, 2018) or sensorimotor-association axis (Sydnor *et al.*, 2021).
A common interpretation of these two patterns is that they represent (i) an “extrinsic” system linked to the immediate sensory environment and (ii) an “intrinsic” system for higher-level internal context, commonly referred to as the default mode network (Raichle *et al.*, 2001).
A common interpretation of these two patterns is that they represent (i) an “extrinsic” system linked to the immediate sensory environment and (ii) an “intrinsic” system for higher-level internal context, commonly observed during resting states.
The other pair of attractors spans an orthogonal axis, and resemble to patterns commonly associated with perception–action cycles (Fuster, 2004), and described as a gradient across sensory-motor modalities (Huntenburg *et al.*, 2018), recruiting regions associated with active inference (e.g. motor cortices) and perceptual inference (e.g visual areas).



**Figure 3:** **Connectome-based Hopfield networks reconstruct characteristics of real resting state brain activity.**  
**A** The four attractor states of the fcHNN model from study 1 reflect brain activation
patterns with high neuroscientific relevance, representing sub-systems previously associated with “internal context”
(blue), “external context” (yellow), “action” (red) and “perception” (green)
(Golland *et al.*, 2008; Cioli *et al.*, 2014; Chen *et al.*, 2018; Fuster, 2004; Margulies *et al.*, 2016).
**B** The attractor states show excellent replicability in two external datasets (study 2 and 3, mean correlation 0.93).
**C** The fcHNN projection (first two PCs of the fcHNN state space) explains significantly more variance (p<0.0001) in the real
resting state fMRI data than principal components derived from the real resting state data itself and generalizes
better (p<0.0001) to out-of-sample data (study 2). Error bars denote 99% bootstrapped confidence intervals.
**D** The fcHNN analysis reliably predicts various characteristics of real resting state fMRI data, such as the fraction of time spent on the basis of the four attractors (first column, p=0.007, contrasted to the multivariate normal null model), the distribution of the data on the fcHNN-projection (second column, p<0.001, contrasted to the multivariate normal null model) and the temporal autocorrelation structure of the real data (third column, p<0.001, contrasted to a null model based on temporally permuted data). This analysis was based on flow maps of the mean trajectories (i.e. the characteristic timeframe-to-timeframe transition direction) in fcHNN-generated data, as compared to a shuffled null model representing zero temporal autocorrelation. For more details, see Methods. Furthermore, (rightmost column), stochastic fcHNNs are capable of self-reconstruction: the timeseries resulting from the stochastic relaxation procedure mirror the co-variance structure of the functional connectome the fcHNN model was initialized with. While the self-reconstruction probability in itself does not strengthen the face-validity of the approach (no unknown information is reconstructed), it is a strong indicator of the model’s construct validity; i.e. that systems that behave like the proposed model inevitably “leak” their weights into the activity time series.

The discovered attractor states demonstrate remarkable replicability (mean Pearson’s correlation 0.93) across the discovery dataset (study 1) and two independent replication datasets (study 2 and 3, Figure 3C). Moreover, they were found to be significantly more robust to noise added to the connectome than nodal strengths scores (used as a reference, see Supplementary Figure 8 for details).

Further analysis in study 1 showed that connectome-based Hopfield models accurately reconstructed multiple characteristics of true resting-state data.
First, the two axes (first two PCs) of the fcHNN projection accounted for a substantial amount of variance in the real resting-state fMRI data in study 1 (mean ) and generalized well to out-of-sample data (study 2, mean ) (Figure 3E). The explained variance of the fcHNN projection significantly exceeded that of the first two PCs derived directly from the real resting-state fMRI data itself ( and 0.364 for in- and out-of-sample analyses).

Second, fcHNN analyses accurately reconstructed various aspects of true resting state brain dynamics.
During stochastic relaxation, the fcHNN model was found to spend approximately three-quarters of the time on the basis of the first two attractor states and one-quarter on the basis of the second pair of attractor states (approximately equally distributed between pairs). We observed similar temporal occupancies in the real data Figure 3D left column), statistically significant with two different null models (Supplementary Figure 5). Fine-grained details of the bimodal distribution observed in the real resting-state fMRI data were also convincingly reproduced by the fcHNN model (Figure 3F and Figure 2D, second column).
Not only spatial activity patterns but also timeseries generated by the fchNN are similar to empirical timeseries data. Next to the visual similarity shown on Figure 2E and F, we observed a statistically significant similarity between the average trajectories of fcHNN-generated and real timeseries “flow” (i.e. the characteristic timeframe-to-timeframe transition direction), as compared to null-models of zero temporal autocorrelation (randomized timeframe order, Figure 3D, third column Methods for analysis details).
Finally, fcHNNs were found to generate signal that preserves the covariance structure of the real functional connectome, indicating that dynamic systems of this type (including the brain) inevitably “leak” their underlying structure into the activity time series, strengthening the construct validity of our approach (Figure 3D).

### An explanatory framework for task-based brain activity

Next to reproducing various characteristics of spontaneous brain dynamics, fcHNNs can also be used to model responses to various perturbations. We obtained task-based fMRI data from a study by Woo *et al.* (2015) (study 4, n=33, see Figure 3), investigating the neural correlates of pain and its self-regulation.

We found that activity changes due to pain (taking into account hemodynamics, see Methods) were characterized on the fcHNN projection by a shift towards the attractor state of action/execution (permutation test for mean projection difference by randomly swapping conditions, p<0.001, Figure 4A, left). Energies, as defined by the fcHNN, were also significantly different between the two conditions (p<0.001), with higher energies during pain stimulation.

When participants were instructed to up- or downregulate their pain sensation (resulting in increased and decreased pain reports and differential brain activity in the nucleus accumbens, NAc (see Woo *et al.* (2015) for details), we observed further changes of the location of momentary brain activity patterns on the fcHNN projection (p<0.001, Figure 4A, right), with downregulation pulling brain dynamics towards the attractor state of internal context and perception. Interestingly, self-regulation did not trigger significant energy changes (p=0.36).



**Figure 4:** **Empirical Hopfield-networks reconstruct real task-based brain activity.**   
**A** Functional MRI time-frames during pain stimulation from study 4 (second fcHNN projection plot)
and self-regulation (third and fourth) are distributed differently on the fcHNN projection than brain states
during rest (first projection, permutation test, p<0.001 for all). Energies, as defined by the Hopfield model, are also
significantly different between rest and the pain conditions (permutation test, p<0.001), with higher energies during
pain stimulation. Triangles denote participant-level mean activations in the various blocks (corrected for
hemodynamics). Small circle plots show the directions of the change for each individual (points) as well as the mean direction
across participants (arrow), as compared to the reference state (downregulation for the last circle plot, rest for all
other circle plots).
**B** Flow-analysis (difference in the average timeframe-to-timeframe transition direction) reveals a non-linear difference in brain dynamics during pain and rest (left). When introducing weak pain-related signal in the fcHNN model during stochastic relaxation, it accurately reproduces these non-linear flow differences (right).
**C** Simulating activity in the Nucleus Accumbens (NAc) (the region showing significant activity differences in Woo *et al.* (2015)) reconstructs the observed non-linear flow difference between up- and downregulation (left).
**D** Schematic representation of brain dynamics during pain and its up- and downregulation, visualized on the fcHNN projection. In the proposed framework, pain does not simply elicit a direct response in certain regions, but instead, shifts spontaneous brain dynamics towards the “action” attractor, converging to a characteristic “ghost attractor” of pain. Down-regulation by NAc activation exerts force towards the attractor of internal context, leading to the brain less frequent “visiting” pain-associated states.
**E** Visualizing meta-analytic activation maps (see Supplementary Table 1 for details) on the fcHNN projection captures intimate relations between the corresponding tasks and **F** serves as a basis for a fcHNN-based theoretical interpretative framework for spontaneous and task-based brain dynamics. In the proposed framework, task-based activity is not a mere response to external stimuli in certain brain locations but a perturbation of the brain’s characteristic dynamic trajectories, constrained by the underlying functional connectivity. From this perspective, “activity maps” from conventional task-based fMRI analyses capture time-averaged differences in these whole brain dynamics.

Next, we conducted a “flow analysis” on the fcHNN projection, quantifying how the average timeframe-to-timeframe transition direction differs on the fcHNN projection between conditions (see Methods).
This analysis unveiled that during pain (Figure 4B, left side), brain activity tends to gravitate towards a distinct point on the projection on the boundary the basins of the internal and action attractors, which we term the “ghost attractor” of pain (similar to Vohryzek *et al.* (2020)). In case of downregulation (as compared to upregulation), brain activity is pulled away from the pain-related “ghost attractor” (Figure 4C, left side), towards the attractor of internal context.

Our fcHNN was able to accurately reconstruct these non-linear dynamics by adding a small amount of realistic “control signal” (similarly to network control theory Liu *et al.* (2011)Gu *et al.* (2015)). To simulate the alterations in brain dynamics during pain stimulation, we acquired a meta-analytic pain activation map (Zunhammer *et al.*, 2021) (n=603) and incorporated it as a control signal added to each iteration of the stochastic relaxation procedure. The ghost attractor found in the empirical data was present across a relatively wide range of signal-to-noise (SNR) values (Supplementary Figure 6). Results with SNR=0.005 are presented on Figure 4B, right side (Pearson’s r = 0.46, p=0.005 based on randomizing conditions on a per-participant basis).

The same model was also able to reconstruct the observed non-linear differences in brain dynamics between the up- and downregulation conditions (Pearson’s r = 0.62, p=0.023) without any further optimization (SNR=0.005,
Figure 4C, right side). The only change we made to the model was the addition (downregulation) or
subtraction (upregulation) of control signal in the NAc (the region in which (Woo *et al.*, 2015) observed significant changes between up- and downregulation), introducing a signal difference of ΔSNR=0.005 (the same value we found optimal in the pain-analysis). Results were reproducible with lower NAc SNRs, too (Supplementary Figure 7).

To provide a comprehensive picture on how tasks and stimuli other than pain map onto the fcHNN projection, we obtained various task-based meta-analytic activation maps from Neurosynth (see Methods) and plotted them on the fcHNN projection (Figure 4E). This analysis reinforced and extended our interpretation of the four investigated attractor states and shed more light on how various functions are mapped on the axes of internal vs. external context and perception vs. action.
In the coordinate system of the fcHNN projection, visual processing is labeled “external-perception”, sensory-motor processes fall into the “external-active” domain, language, verbal cognition and working memory belongs to the “internal-active” region and long-term memory as well as social and autobiographic schemata fall into the “internal-perception” regime (Figure 4F).

### Clinical relevance

We obtained data from n=172 autism spectrum disorder (ASD) and typically developing control (TDC) individuals, acquired at the New York University Langone Medical Center, New York, NY, USA (NYU) and generously shared in the Autism Brain Imaging Data Exchange dataset (study 7: ABIDE, (Di Martino *et al.*, 2014).
After excluding high-motion cases (with the same approach as in Study 1-4, see Methods), we visualized the distribution of time-frames on the fcHNN-projection separately for the ASD and TDC groups (Figure 5A).
First, we assigned all timeframes to one of the 4 attractor states with the fcHNN from study 1 and found several significant differences in the mean activity on the attractor basins (see Methods) of the ASD group as compared to the respective controls (Figure 5B).
Strongest differences were found on the “action-perception” axis (Table 1), with increased activity of the sensory-motor and middle cingular cortices during “action-execution” related states and increased visual and decreased sensory and auditory activity during “perception” states, likely reflecting the widely acknowledged, yet poorly understood, perceptual atypicalities in ASD (Hadad & Schwartz, 2019). ASD related changes in the internal-external axis were characterized by more involvement of the posterior cingulate, the precuneus, the nucleus accumbens, the dorsolateral prefrontal cortex (dlPFC), the cerebellum (Crus II, lobule VII) and inferior temporal regions during activity of the internalizing subsystem (Table 1). While similar, default mode network (DMN)-related changes have often been attributed to an atypical integration of information about the “self” and the “other” (Padmanabhan *et al.*, 2017), a more detailed fcHNN-analysis may help to further disentangle the specific nature of these changes.



**Figure 5:** **Connectome-based Hopfield analysis of autism spectrum disorder.**   
**A** The distribution of time-frames on the fcHNN-projection separately for ASD patients and typically developing control (TDC) participants.   
**B** We quantified attractor state activations in the Autism Brain Imaging Data Exchange datasets (study 7) as the
individual-level mean activation of all time-frames belonging to the same attractor state. This analysis captured alterations similar to those previously associated to ASD-related perceptual atypicalities (visual, auditory and somatosensory cortices) as well as atypical integration of information about the “self” and the “other” (default mode network regions). All results are corrected for multiple comparisons across brain regions and attractor states (122\*4 comparisons) with Bonferroni-correction. See Table 1 and Supplementary Figure 9 for detailed results.   
**C** The comparison of data generated by fcHNNs initialized with ASD and TDC connectomes, respectively, revealed a characteristic pattern of differences in the system’s dynamics, with increased pull towards (and potentially a higher separation between) the action and perception attractors and a lower tendency of trajectories going towards the internal and external attractors.   
***Abbreviations***: MCC: middle cingulate cortex, ACC: anterior cingulate cortex, pg: perigenual, PFC: prefrontal cortex, dm: dorsomedial, dl: dorsolateral, STG: superior temporal gyrus, ITG: inferior temporal gyrus, Caud/Acc: caudate-accumbens, SM: sensorimotor, V1: primary visual, A1: primary auditory, SMA: supplementary motor cortex, ASD: autism spectrum disorder, TDC: typically developing control.

**Table 1:** **The top ten largest changes in average attractor-state activity between autistic and control individuals.** Mean attractor-state activity changes are presented in the order of their absolute effect size. All p-values are based on permutation tests (shuffling the group assignment) and corrected for multiple comparisons (via Bonferroni’s correction). For a comprehensive list of significant findings, see {numref}`Supplementary Figure %s <si\_clinical\_results\_table>.

| region | attractor | effect size | p-value |
| --- | --- | --- | --- |
| primary auditory cortex | perception | -0.126 | <0.0001 |
| middle cingulate cortex | action | 0.109 | <0.0001 |
| cerebellum lobule VIIb (medial part ) | internal context | 0.104 | <0.0001 |
| mediolateral sensorimotor cortex | perception | -0.099 | 0.00976 |
| precuneus | action | 0.098 | <0.0001 |
| middle superior temporal gyrus | perception | -0.098 | <0.0001 |
| frontal eye field | perception | -0.095 | <0.0001 |
| dorsolateral sensorimotor cortex | perception | -0.094 | 0.00976 |
| posterior cingulate cortex | action | 0.092 | <0.0001 |
| dorsolateral prefrontal cortex | external context | -0.092 | <0.0001 |

Thus, we contrasted the characteristic trajectories derived from the fcHNN models of the two groups (initialized with the group-level functional connectomes). Our fcHNN-based flow analysis predicted that in ASD, there is an increased likelihood of states returning towards the middle (more noisy states) from the internal-external axis and an increased likelihood of states transitioning towards the extremes of the action-perception axis (Figure 5C). We observed a highly similar pattern in the real data (Pearson’s correlation: 0.66), statistically significant after permutation testing (shuffling the group assignment, p=0.009).

## Discussion

In this study, we have introduced and validated a simple and robust network-level generative computational framework that elucidates how activity propagation within the functional connectome orchestrates large-scale brain dynamics, leading to the spontaneous emergence of brain states, smooth gradients among them, and characteristic dynamic responses to perturbations.

The construct validity of our model is rooted in the activity flow principle, first introduced by Cole *et al.* (2016). The activity flow principle states that activity in a brain region can be predicted by a weighted combination of the activity of all other regions, where the weights are set to the functional connectivity of those regions to the held-out region. This principle has been shown to hold across a wide range of experimental and clinical conditions (Cole *et al.*, 2016; Ito *et al.*, 2017; Mill *et al.*, 2022; Hearne *et al.*, 2021; Chen *et al.*, 2018).
The proposed approach is based on the intuition that the repeated, iterative application of the activity flow equation exhibits close analogies with a type of recurrent artificial neural networks, known as Hopfield networks (Hopfield, 1982).

Hopfield networks have been widely acknowledged for their relevance for brain function, including the ability to store and recall memories (Hopfield, 1982), self-repair (Murre *et al.*, 2003),
a staggering robustness to noisy or corrupted inputs (Hertz *et al.*, 1991) (see also Supplementary Figure 8) and the ability to produce multistable dynamics organized by the “gravitational pull” of a finite number of attractor states (Khona & Fiete, 2022). While many of such properties of Hopfield networks have previously been proposed as a model for micro-scale neural systems (see Khona & Fiete (2022) for a review), the proposed link between macro-scale activity propagation and Hopfield networks allows transferring the vast body of knowledge on Hopfield networks to the study of large-scale brain dynamics.

Integrating Cole’s activity flow principle with the HNN architecture mandates the initiation of network weights with functional connectivity values, specifically partial correlations as suggested by Cole *et al.* (2016).
Considering the functional connectome as weights of an already trained neural network distinguishes our methodology from conventional biophysical and phenomenological computational modeling strategies, which usually rely on the structural connectome to model polysynaptic connectivity (Cabral *et al.*, 2017; Deco *et al.*, 2012; Golos *et al.*, 2015; Hansen *et al.*, 2015). Given the challenges of accurately modelling the structure-function coupling in the brain (Seguin *et al.*, 2023), such models are currently limited in terms of reconstruction accuracy, hindering translational applications.
Working with direct, functional MRI-based activity flow estimates, fcHNNs bypass the challenge of modelling the structural-functional coupling and provide a more direct and accurate representation of the brain’s dynamic activity propagation, although at the cost of losing the ability to provide biophysical details on the underlying mechanisms.
Another advantage of the proposed model is its simplicity. While many conventional computational models rely on the optimization of a high number of free parameters, the basic form of the fcHNN approach comprises solely two, easily interpretable “hyperparameters” (temperature and noise) and yields notably consistent outcomes across an extensive range of these parameters (Supplementary Figure 1, 3, 5, 6, 7). To underscore the potency of this simplicity and stability, in the present work, we avoided any unnecessary parameter optimization, leaving a negligible chance of overwriting. It is likely, however, that extensive parameter optimization could further improve the performance of the model.

Further, the fcHNN approach allows us to leverage on knowledge about the underlying ANN architecture. Specifically, Hopfield attractor dynamics provide a mechanistic account for the emergence of large-scale canonical brain networks (Zalesky et al., 2014) ), and shed light to the origin of characteristic task-responses that are accounted by “ghost attractors” in the system (Deco & Jirsa, 2012; Vohryzek *et al.*, 2020).
As fcHNNs do not need to be trained to solve any explicit tasks, they are well suited to examine spontaneous brain dynamics. However, it is worth mentioning that fcHNNs are compatible with the neuroconnectionist approach (Doerig et al., 2021), as well.
Like any other ANNs, fcHNNs can also be further trained via established ANN training techniques (e.g. via the Hebbian learning rule) to “solve” various tasks or to match developmental dynamics or pathological alterations. In this promising future direction, the training procedure itself becomes part of the model, providing testable hypotheses about the formation, and various malformations, of brain dynamics.

Given its simplicity, it is noteworthy, how well the fcHNN model is able to reconstruct and predict brain dynamics under a wide range of conditions.
Our finding that the two-dimensional fcHNN projection can explain more variance in real resting state fMRI data than the first two principal components derived from the data itself may indicate that through the known noise tolerance of the underlying ANN architecture, fcHNNs are able to capture essential principles of the underlying dynamic processes even if our empirical measurements are corrupted by noise and low sampling rate.
Indeed, fcHNN attractor states were found to be robust to noisy weights (Supplementary Figure 8) and highly replicable across datasets acquired at different sites, with different scanners and imaging sequences (study 2 and 3). The observed high level of replicability allowed us to re-use the fcHNN model constructed with the connectome of study 1 for all subsequent analyses, without any further fine-tuning or study-specific parameter optimization.

Conceptually, the notion of a global attractor model of the brain network is not new (Freeman, 1987; Deco & Jirsa, 2012; Vohryzek *et al.*, 2020; Deco *et al.*, 2012; Golos *et al.*, 2015; Hansen *et al.*, 2015).
The present work shows, however, that the brain as an attractor network necessarily ‘leaks its internal weights’ in form of the partial correlation across the regional timeseries. This indicates that, partial correlations across neural timeseries data from different regions (i.e. functional connectivity) may be a more straightforward entry point to investigating the brain’s attractor dynamics, than estimates of structural connectedness.
Thereby, the fcHNN approach provides a simple and interpretable way to infer and investigate the attractor states of the brain, without the need for additional assumptions about the underlying biophysical details. This is a significant advantage, as the functional connectome can be easily and non-invasively acquired in humans, while biophysical details required by other models are hard to measure or estimate accurately.
Furthermore, here we complement previous work on large-scale brain attractor dynamics, by demonstrating that the reconstructed attractor states are not solely local minima in the state-space but act as a driving force for the dynamic trajectories of brain activity. We argue that attractor-dynamics may be the main driving factor for the spatial and temporal autocorrelation structure of the brain, recently described to be predictive of network topology in relation to age, subclinical symptoms of dementia, and pharmacological manipulations with serotonergic drugs (Shinn *et al.*, 2023).
Nevertheless, attractor states should not be confused with the conventional notion of brain states (Chen *et al.*, 2015) and large-scale functional gradients (Margulies *et al.*, 2016). In the fcHNN framework, attractor states can rather be conceptualized as “Platonic idealizations” of brain activity, that are continuously approximated - but never reached - by the brain, resulting in re-occurring patterns (brain states) and smooth gradual transitions (large-scale gradients).

Relying on previous work, we can establish a relatively straightforward (although somewhat speculative) correspondence between attractor states and brain function, mapping brain activation on the axes of internal vs. external context (Golland *et al.*, 2008; Cioli *et al.*, 2014), as well as perception vs. action (Fuster, 2004).
This four-attractor architecture exhibits an appealing analogy with Friston’s free energy principle (Friston *et al.*, 2006) that postulates the necessary existence of brain subsystems for active and perceptual inference and proposes that the dynamical dependencies that drive the flow of information in the brain can be represented with a hierarchically nested structure (e.g. external and internal subsystem) that may be an essential ingredient of conscious (Ramstead *et al.*, 2023) and autonomous (Lee *et al.*, 2023) agents.

Both conceptually and in terms of analysis practices, resting and task states are often treated as separate phenomena. However, in the fcHNN framework, the differentiation between task and resting states is considered an artificial dichotomy.
Task-based brain activity in the fcHNN framework is not a mere response to external stimuli in certain brain locations but a perturbation of the brain’s characteristic dynamic trajectories, with increased preference for certain locations on the energy landscape (“ghost attractors”).
In our analyses, the fcHNN approach captured and predicted participant-level activity changes induced by pain and its self-regulation and gave a mechanistic account for how relatively small activity changes in a single region (NAcc) may result in a significantly altered pain experience.
Our control-signal analysis is different from, but compatible with, linear network control theory-based (Gu et al., 2015) approaches. Combining network control theory with the fcHNN approach could provide a powerful framework for understanding the effects of various tasks, conditions and interventions (e.g. brain stimulation) on brain dynamics.

Brain dynamics can not only be perturbed by task or other types of experimental or naturalistic interventions, but also by pathological alterations. Here we provide an initial demonstration (study 7) of how fcHNN-based analyses can characterize and predict altered brain dynamics in autism spectrum disorder (ASD). The observed ASD-associated changes in brain dynamics are indicative of a reduced ability to flexibly switch between perception and internal representations, corroborating previous findings that in ASD, sensory-driven connectivity transitions do not converge to transmodal areas (Hong *et al.*, 2019). Such findings are in line with previous reports of a reduced influence of context on the interpretation of incoming sensory information in ASD (e.g. the violation of Weber’s law) (Hadad & Schwartz, 2019).

Our findings open up a series of exciting opportunities for the better understanding of brain function in health and disease.
First, the 2-dimensional fcHNN projection offers a simple framework not only for the visualization, but also for the *interpretation*, of brain activity patterns, as it conceptualizes changes related to various behavioral or clinical states or traits as a shift in brain dynamics in relation to brain attractor states.
Second, fcHNN analyses may provide insights into the causes of changes in brain dynamics, by for instance, identifying the regions or connections that act as an “Achilles heel” in generating such changes. Such control analyses could, for instance, aid the differentiation of primary causes and secondary effects of activity or connectivity changes in various clinical conditions.
Third, the fcHNN approach can provide testable predictions about the effects of pharmacological interventions as well as non-invasive brain stimulation (e.g. transcranial magnetic or direct current stimulation, focused ultrasound, etc.) and neurofeedback. Obtaining the optimal stimulation or treatment target within the fcHNN framework (e.g. by means of network control theory (Liu *et al.*, 2011)) is one of the most promising future directions with the potential to significantly advance the development of novel, personalized treatment approaches.

The proposed approach is not without limitations. First, the fcHNN model is obviously a simplification of the brain’s dynamics, and it does not aim to explain the brain’s ability to perform certain computations, brain regions’ ability to perform certain functions or biophysical details underlying (altered) polysynaptic connections. Nevertheless, our approach showcases that many characteristics of brain dynamics, like multistability, temporal autocorrelations, states and gradients, can be explained, and predicted, by a very simple nonlinear phenomenological model.
Second, our model assumes a stationary connectome, which seems to contradict notions of dynamic connectivity. However, with realistically changing control signals, our model can easily reconstruct dynamic connectivity changes, which still stem from an underlying, stationary functional connectivity structure. This is in line with the notion of “latent functional connectivity”; and intrinsic brain network architecture built up from connectivity properties that are persistent across brain states McCormick *et al.* (2022).
In this initial work, we presented the simplest possible implementation of the fcHNN concept. It is clear that the presented analyses exploit only a small proportion of the richness of the full state-space dynamics reconstructed by the fcHNN model.
There are many potential ways to further improve the utility of the fcHNN approach. Increasing the number of reconstructed attractor states (by increasing the temperature parameter), investigating higher-dimensional dynamics, fine-tuning the hyperparameters, testing the effect of different initializations and perturbations are all important direction for future work, with the potential to further improve the model’s accuracy and usefulness.

## Conclusion

To conclude, here we have proposed a lightweight, high-level computational framework that accurately captures and predicts brain dynamics under a wide range of conditions, including resting states, task-induced activity changes and brain disorders. The framework models large-scale activity flow in the brain with a recurrent artificial neural network architecture that, instead of being trained to solve specific tasks or mimic certain dynamics, is simply initialized with the empirical functional connectome. The framework identifies neurobiologically meaningful attractor states and provides a model for how these restrict brain dynamics.
The proposed model establishes a conceptual link between connectivity and activity, provides a mechanistic account for the emergence of brain states, gradients and temporal autocorrelation structure and offers a simple, robust, and highly interpretable computational alternative to conventional descriptive approaches to investigating brain function. The generative nature of our proposed model opens up a wealth of opportunities for future research, including predicting the effect, and understanding the mechanistic bases, of various interventions; thereby paving the way for designing novel treatment approaches.

## References

Barttfeld, P., Uhrig, L., Sitt, J. D., Sigman, M., Jarraya, B., & Dehaene, S. (2015). Signature of consciousness in the dynamics of resting-state brain activity. *Proceedings of the National Academy of Sciences*, *112*(3), 887–892.

Bassett, D. S., & Sporns, O. (2017). Network neuroscience. *Nature Neuroscience*, *20*(3), 353–364.

Breakspear, M. (2017). Dynamic models of large-scale brain activity. *Nature Neuroscience*, *20*(3), 340–352.

Buzsaki, G. (2006). *Rhythms of the Brain*. Oxford university press.

Cabral, J., Kringelbach, M. L., & Deco, G. (2017). Functional connectivity dynamically evolves on multiple time-scales over a static structural connectome: Models and mechanisms. *NeuroImage*, *160*, 84–96.

Chen, J. E., Chang, C., Greicius, M. D., & Glover, G. H. (2015). Introducing co-activation pattern metrics to quantify spontaneous brain network dynamics. *Neuroimage*, *111*, 476–488.

Chen, R. H., Ito, T., Kulkarni, K. R., & Cole, M. W. (2018). The human brain traverses a common activation-pattern state space across task and rest. *Brain Connectivity*, *8*(7), 429–443.

Chiêm, B., Crevecoeur, F., & Delvenne, J.-C. (2021). Structure-informed functional connectivity driven by identifiable and state-specific control regions. *Network Neuroscience*, *5*(2), 591–613.

Cioli, C., Abdi, H., Beaton, D., Burnod, Y., & Mesmoudi, S. (2014). Differences in human cortical gene expression match the temporal properties of large-scale functional networks. *PloS One*, *9*(12), e115913.

Cole, M. W., Ito, T., Bassett, D. S., & Schultz, D. H. (2016). Activity flow over resting-state networks shapes cognitive task activations. *Nature Neuroscience*, *19*(12), 1718–1726.

Daunizeau, J., Stephan, K. E., & Friston, K. J. (2012). Stochastic dynamic causal modelling of fMRI data: should we care about neural noise? *Neuroimage*, *62*(1), 464–481.

Deco, G., & Jirsa, V. K. (2012). Ongoing cortical activity at rest: criticality, multistability, and ghost attractors. *Journal of Neuroscience*, *32*(10), 3366–3375.

Deco, G., Senden, M., & Jirsa, V. (2012). How anatomy shapes dynamics: a semi-analytical study of the brain at rest by a simple spin model. *Frontiers in Computational Neuroscience*, *6*, 68.

Di Martino, A., Yan, C.-G., Li, Q., Denio, E., Castellanos, F. X., Alaerts, K., Anderson, J. S., Assaf, M., Bookheimer, S. Y., Dapretto, M., & others. (2014). The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular Psychiatry*, *19*(6), 659–667.

Doerig, A., Sommers, R. P., Seeliger, K., Richards, B., Ismael, J., Lindsay, G. W., Kording, K. P., Konkle, T., Van Gerven, M. A., Kriegeskorte, N., & others. (2023). The neuroconnectionist research programme. *Nature Reviews Neuroscience*, 1–20.

Freeman, W. J. (1987). Simulation of chaotic EEG patterns with a dynamic model of the olfactory system. *Biological Cybernetics*, *56*(2–3), 139–150.

Friston, K., Kilner, J., & Harrison, L. (2006). A free energy principle for the brain. *Journal of Physiology-Paris*, *100*(1–3), 70–87.

Fuster, J. M. (2004). Upper processing stages of the perception–action cycle. *Trends in Cognitive Sciences*, *8*(4), 143–145.

Golland, Y., Golland, P., Bentin, S., & Malach, R. (2008). Data-driven clustering reveals a fundamental subdivision of the human cortex into two global systems. *Neuropsychologia*, *46*(2), 540–553.

Golos, M., Jirsa, V., & Daucé, E. (2015). Multistability in large scale models of brain activity. *PLoS Computational Biology*, *11*(12), e1004644.

Greene, A. S., Horien, C., Barson, D., Scheinost, D., & Constable, R. T. (2023). Why is everyone talking about brain state? *Trends in Neurosciences*.

Gu, S., Pasqualetti, F., Cieslak, M., Telesford, Q. K., Yu, A. B., Kahn, A. E., Medaglia, J. D., Vettel, J. M., Miller, M. B., Grafton, S. T., & others. (2015). Controllability of structural brain networks. *Nature Communications*, *6*(1), 8414.

Hadad, B.-S., & Schwartz, S. (2019). Perception in autism does not adhere to Weber’s law. *Elife*, *8*, e42223.

Hansen, E. C., Battaglia, D., Spiegler, A., Deco, G., & Jirsa, V. K. (2015). Functional connectivity dynamics: modeling the switching behavior of the resting state. *Neuroimage*, *105*, 525–535.

Hearne, L. J., Mill, R. D., Keane, B. P., Repovš, G., Anticevic, A., & Cole, M. W. (2021). Activity flow underlying abnormalities in brain activations and cognition in schizophrenia. *Science Advances*, *7*(29), eabf2513.

Heinz, A., Murray, G. K., Schlagenhauf, F., Sterzer, P., Grace, A. A., & Waltz, J. A. (2019). Towards a unifying cognitive, neurophysiological, and computational neuroscience account of schizophrenia. *Schizophrenia Bulletin*, *45*(5), 1092–1100.

Hertz, J., Krogh, A., & Palmer, R. G. (1991). Introduction to the Theory of Neural Computation, chapter 7. *Lecture Notes*, *1*.

Hong, S.-J., Vos de Wael, R., Bethlehem, R. A., Lariviere, S., Paquola, C., Valk, S. L., Milham, M. P., Di Martino, A., Margulies, D. S., Smallwood, J., & others. (2019). Atypical functional connectome hierarchy in autism. *Nature Communications*, *10*(1), 1022.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, *79*(8), 2554–2558.

Huntenburg, J. M., Bazin, P.-L., & Margulies, D. S. (2018). Large-scale gradients in human cortical organization. *Trends in Cognitive Sciences*, *22*(1), 21–31.

Hutchison, R. M., Womelsdorf, T., Allen, E. A., Bandettini, P. A., Calhoun, V. D., Corbetta, M., Della Penna, S., Duyn, J. H., Glover, G. H., Gonzalez-Castillo, J., & others. (2013). Dynamic functional connectivity: promise, issues, and interpretations. *Neuroimage*, *80*, 360–378.

Ito, T., Kulkarni, K. R., Schultz, D. H., Mill, R. D., Chen, R. H., Solomyak, L. I., & Cole, M. W. (2017). Cognitive task information is transferred between brain regions via resting-state network topology. *Nature Communications*, *8*(1), 1027.

Khona, M., & Fiete, I. R. (2022). Attractor and integrator networks in the brain. *Nature Reviews Neuroscience*, *23*(12), 744–766.

Koiran, P. (1994). Dynamics of discrete time, continuous state Hopfield networks. *Neural Computation*, *6*(3), 459–468.

Kriegeskorte, N., & Douglas, P. K. (2018). Cognitive computational neuroscience. *Nature Neuroscience*, *21*(9), 1148–1160.

Krotov, D. (2023). A new frontier for Hopfield networks. *Nature Reviews Physics*, 1–2.

Lee, S., Oh, Y., An, H., Yoon, H., Friston, K. J., Hong, S. J., & Woo, C.-W. (2023). Life-inspired Interoceptive Artificial Intelligence for Autonomous and Adaptive Agents. *arXiv Preprint arXiv:2309.05999*.

Liu, X., & Duyn, J. H. (2013). Time-varying functional network information extracted from brief instances of spontaneous brain activity. *Proceedings of the National Academy of Sciences*, *110*(11), 4392–4397.

Liu, Y.-Y., Slotine, J.-J., & Barabási, A.-L. (2011). Controllability of complex networks. *Nature*, *473*(7346), 167–173.

Margulies, D. S., Ghosh, S. S., Goulas, A., Falkiewicz, M., Huntenburg, J. M., Langs, G., Bezgin, G., Eickhoff, S. B., Castellanos, F. X., Petrides, M., & others. (2016). Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proceedings of the National Academy of Sciences*, *113*(44), 12574–12579.

McCormick, E. M., Arnemann, K. L., Ito, T., Hanson, S. J., & Cole, M. W. (2022). Latent functional connectivity underlying multiple brain states. *Network Neuroscience*, *6*(2), 570–590. [10.1162/netn\_a\_00234](https://doi.org/10.1162/netn_a_00234)

Meer, J. N. van der, Breakspear, M., Chang, L. J., Sonkusare, S., & Cocchi, L. (2020). Movie viewing elicits rich and reliable brain state dynamics. *Nature Communications*, *11*(1), 5004.

Mill, R. D., Hamilton, J. L., Winfield, E. C., Lalta, N., Chen, R. H., & Cole, M. W. (2022). Network modeling of dynamic brain interactions predicts emergence of neural information that supports human cognitive behavior. *PLoS Biology*, *20*(8), e3001686.

Murray, J. D., Demirtaş, M., & Anticevic, A. (2018). Biophysical modeling of large-scale brain dynamics and applications for computational psychiatry. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *3*(9), 777–787.

Murre, J. M., Griffioen, R., & Robertson, I. (2003). Selfreparing neural networks: a model for recovery from brain damage. *Knowledge-Based Intelligent Information and Engineering Systems: 7th International Conference, KES 2003, Oxford, UK, September 2003. Proceedings, Part II 7*, 1164–1171.

Padmanabhan, A., Lynch, C. J., Schaer, M., & Menon, V. (2017). The default mode network in autism. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *2*(6), 476–486.

Papadopoulos, L., Kim, J. Z., Kurths, J., & Bassett, D. S. (2017). Development of structural correlations and synchronization from adaptive rewiring in networks of Kuramoto oscillators. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *27*(7).

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences*, *98*(2), 676–682.

Ramstead, M. J., Albarracin, M., Kiefer, A., Klein, B., Fields, C., Friston, K., & Safron, A. (2023). The inner screen model of consciousness: applying the free energy principle directly to the study of conscious experience. *arXiv Preprint arXiv:2305.02205*.

Richards, B. A., Lillicrap, T. P., Beaudoin, P., Bengio, Y., Bogacz, R., Christensen, A., Clopath, C., Costa, R. P., de Berker, A., Ganguli, S., & others. (2019). A deep learning framework for neuroscience. *Nature Neuroscience*, *22*(11), 1761–1770.

Robinson, P. A., Rennie, C., Rowe, D. L., O’Connor, S., Gordon, & E. (2005). Multiscale brain modelling. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1457), 1043–1050.

Scheid, B. H., Ashourvan, A., Stiso, J., Davis, K. A., Mikhail, F., Pasqualetti, F., Litt, B., & Bassett, D. S. (2021). Time-evolving controllability of effective connectivity networks during seizure progression. *Proceedings of the National Academy of Sciences*, *118*(5), e2006436118.

Schiff, S. J., Jerger, K., Duong, D. H., Chang, T., Spano, M. L., & Ditto, W. L. (1994). Controlling chaos in the brain. *Nature*, *370*(6491), 615–620.

Schirner, M., Kong, X., Yeo, B. T., Deco, G., & Ritter, P. (2022). Dynamic primitives of brain network interaction. *NeuroImage*, *250*, 118928.

Seguin, C., Sporns, O., & Zalesky, A. (2023). Brain network communication: concepts, models and applications. *Nature Reviews Neuroscience*, *24*(9), 557–574.

Shinn, M., Hu, A., Turner, L., Noble, S., Preller, K. H., Ji, J. L., Moujaes, F., Achard, S., Scheinost, D., Constable, R. T., & others. (2023). Functional brain networks reflect spatial and temporal autocorrelation. *Nature Neuroscience*, 1–12.

Smith, S. M., Miller, K. L., Moeller, S., Xu, J., Auerbach, E. J., Woolrich, M. W., Beckmann, C. F., Jenkinson, M., Andersson, J., Glasser, M. F., & others. (2012). Temporally-independent functional modes of spontaneous brain activity. *Proceedings of the National Academy of Sciences*, *109*(8), 3131–3136.

Sydnor, V. J., Larsen, B., Bassett, D. S., Alexander-Bloch, A., Fair, D. A., Liston, C., Mackey, A. P., Milham, M. P., Pines, A., Roalf, D. R., & others. (2021). Neurodevelopment of the association cortices: Patterns, mechanisms, and implications for psychopathology. *Neuron*, *109*(18), 2820–2846.

Vidaurre, D., Smith, S. M., & Woolrich, M. W. (2017). Brain network dynamics are hierarchically organized in time. *Proceedings of the National Academy of Sciences*, *114*(48), 12827–12832.

Vohryzek, J., Deco, G., Cessac, B., Kringelbach, M. L., & Cabral, J. (2020). Ghost attractors in spontaneous brain activity: Recurrent excursions into functionally-relevant BOLD phase-locking states. *Frontiers in Systems Neuroscience*, *14*, 20.

Woo, C.-W., Roy, M., Buhle, J. T., & Wager, T. D. (2015). Distinct brain systems mediate the effects of nociceptive input and self-regulation on pain. *PLoS Biology*, *13*(1), e1002036.

Zalesky, A., Fornito, A., Cocchi, L., Gollo, L. L., & Breakspear, M. (2014). Time-resolved resting-state brain networks. *Proceedings of the National Academy of Sciences*, *111*(28), 10341–10346.

Zunhammer, M., Spisák, T., Wager, T. D., & Bingel, U. (2021). Meta-analysis of neural systems underlying placebo analgesia from individual participant fMRI data. *Nature Communications*, *12*(1), 1391.