Dynamical models reveal anatomically reliable attractor landscapes embedded in resting state brain networks

Ruiqi Chen[a,1], **Matthew Singh**[b], **Todd S. Braver**[c], **and ShiNung Ching**[b]

[a]Division of Biology and Biomedical Sciences, Washington University in St. Louis, St. Louis, MO 63108; [b]Department of Electrical and Systems Engineering, Washington University in St. Louis, St. Louis, MO 63108; [c]Department of Psychological & Brain Sciences, Washington University in St. Louis, St. Louis, MO 63108

This manuscript was compiled on January 15, 2024

Analyses of functional connectivity (FC) in resting-state brain networks (RSNs) have generated many insights into cognition. However, the mechanistic underpinnings of FC and RSNs are still not well-understood. It remains debated whether resting state activity is best characterized as noise-driven fluctuations around a single stable state, or instead, as a nonlinear dynamical system with nontrivial attractors embedded in the RSNs. Here, we provide evidence for the latter, by constructing whole-brain dynamical systems models from individual resting-state fMRI (rfMRI) recordings, using the Mesoscale Individualized NeuroDynamic (MINDy) platform. The MINDy models consist of hundreds of neural masses representing brain parcels, connected by fully trainable, individualized weights. We found that our models manifested a diverse taxonomy of nontrivial attractor landscapes including multiple equilibria and limit cycles. However, when projected into anatomical space, these attractors mapped onto a limited set of canonical RSNs, including the default mode network (DMN) and frontoparietal control network (FPN), which were reliable at the individual level. Further, by creating convex combinations of models, bifurcations were induced that recapitulated the full spectrum of dynamics found via fitting. These findings suggest that the resting brain traverses a diverse set of dynamics, which generates several distinct but anatomically overlapping attractor landscapes. Treating rfMRI as a unimodal stationary process (i.e., conventional FC) may miss critical attractor properties and structure within the resting brain. Instead, these may be better captured through neural dynamical modeling and analytic approaches. The results provide new insights into the generative mechanisms and intrinsic spatiotemporal organization of brain networks.

Resting state fMRI | Dynamical systems modeling | Individual differences | Resting state networks | Bifurcations

Resting state fMRI (rfMRI) has become an important tool to probe the link between ongoing brain dynamics and cognition. The most common analytic approach utilized in rfMRI studies is to characterize brain-wide statistical associations (known as functional connectivity or FC) and relate them to cognitive and behavioral indices. However, the dynamical processes that generate the observed rfMRI fluctuations and statistics (e.g., FC) remain elusive. In particular, it is unknown whether resting state dynamics can be best described as a unimodal stationary process featuring statistical fluctuations around the mean, or instead, as a nonlinear dynamic system with nontrivial fluctuations associated with stable patterns different from the mean.

The prior FC literature has provided mixed support for both hypotheses. Traditionally, FC is considered stationary over the scanning session (1). Correspondingly, the underlying dynamics are found to contain a stable equilibrium (point attractor) at the global mean, and the noisy fluctuations around this stable mean produce the observed FC pattern (2). However, recent years have witnessed the rapid development of an analysis technique called time-varying functional connectivity (tvFC), also known as dynamic FC (3). The tvFC method identifies recurring short-time-windowed FC patterns that differ from the mean FC. These transient patterns are reliable within individuals and across populations (4, 5). More interestingly, transient FC patterns but not the mean FC were found to predict psychopathology (6). These findings seem to suggest the existence of nontrivial, functionally salient fluctuation in resting state dynamics.

However, the nature and validity of tvFC characterization is itself still under debate. For example, it is known that head motion and physiological noise generate confounds in FC (7), and even more so for tvFC, which relies on data from shorter duration timeseries (i.e., windowed epochs). More fundamentally, even if tvFC faithfully captures the temporal evolution of neural covariation patterns, it is still unclear whether tvFC states are merely snapshots of the noisy fluctuations around a stable mean, or if they can be associated with nontrivial dynamics. Indeed, tvFC states might be generated from various kinds of nontrivial dynamics (8). However, an influential paper (9) showed that tvFC clustering would produce very similar results when applied to either real data or stationary noise with matched mean FC and power spectral density. Therefore, to understand the substrate of brain-wide associations,

Significance Statement

Our brain remains active even when not engaged in cognitively demanding tasks. However, the processes that determine such ‘resting state’ activity are still not well-understood. Using a large ($n > 1000$) functional neuroimaging dataset and new techniques for computationally modeling brain activity, we found that the resting brain possesses several distinct mechanisms by which activity can be generated. These mechanisms, or dynamics, vary moment to moment, but result in the activation of similar anatomical regions across different individuals. Our results suggest that the resting brain is neither idle, nor monolithic in its governing mechanisms, but rather possesses a diverse but consistent taxonomy of ways in which it can activate and hence transition to cognitive tasks.

R.C., T.S.B, and S.C. designed research; R.C., M.S., T.S.B., and S.C. developed methods; R.C. analyzed data; and R.C., T.S.B, and S.C. wrote the paper.

The authors declare no competing interests.

---

[1] Address correspondence to chen.ruiqi@wustl.edu

their temporal fluctuations, and ultimately the resting state dynamics that produce such associations, it is necessary to go beyond descriptive methods like tvFC clustering and adopt a more mechanistic framework.

Dynamical systems modeling and analysis can provide unique insights into the problem of nontrivial fluctuations in resting state dynamics. Dynamic models of brain activity predict the evolution of activation timeseries given an initial estimate of hidden states. They thus provide a generative mechanism for resting state dynamics and associated statistics, such as FC. To date, the evidence for nontrivial fluctuations from dynamic modeling is also mixed. There are two relevant types of neural models utilized to characterize resting state brain dynamics as measured by rfMRI: structural-connectome-informed models, and directly-parameterized models. The first type of models usually contain hundreds of sub-components representing brain parcels, connected by weights derived from the brain's structural connectome, e.g., through diffusion tensor imaging. Early models typically included only a few free parameters, such as the global scaling factor for connectivity, that are directly fit to the fMRI data. It was consistently found that the emergent dynamics involved multiple nontrivial attractors (2, 10), and the distribution of nontrivial attractors might reflect the organization of resting state functional networks (11). However, a recent study (12) that replaced the neural mass approximation of regional dynamics with a more powerful approximation scheme (i.e., an artificial neural network) reported the opposite, with a single globally stable attractor located at the mean. The second type of models do not assume that the structural connectome is a good surrogate for functional coupling, but rather directly optimize the effective connectivity between regions by predicting empirical fMRI time series. Most of the existing works adopting this approach has utilized the framework of Dynamical Causal Modeling (DCM) (13). Although nonlinear DCM has been proposed, it is computationally too expensive for more than ten nodes (14). Therefore, most rfMRI DCMs have used a linear approximation (15), which by definition cannot express nontrivial fluctuations. It has been argued that such stationary linear models have even lower mean estimation error than common nonlinear models for rfMRI (16). However, a rigorous Bayesian model comparison found that a time-varying ('dynamic', short-time-windowed) linear DCM clearly outperformed a stationary linear DCM (17). In short, previous studies have associated rfMRI with a variety of dynamics ranging from a monostable linear or weakly nonlinear system to a multistable strongly nonlinear system, with diverging evidence for nontrivial fluctuations. What might be the explanation for such inconsistent results?

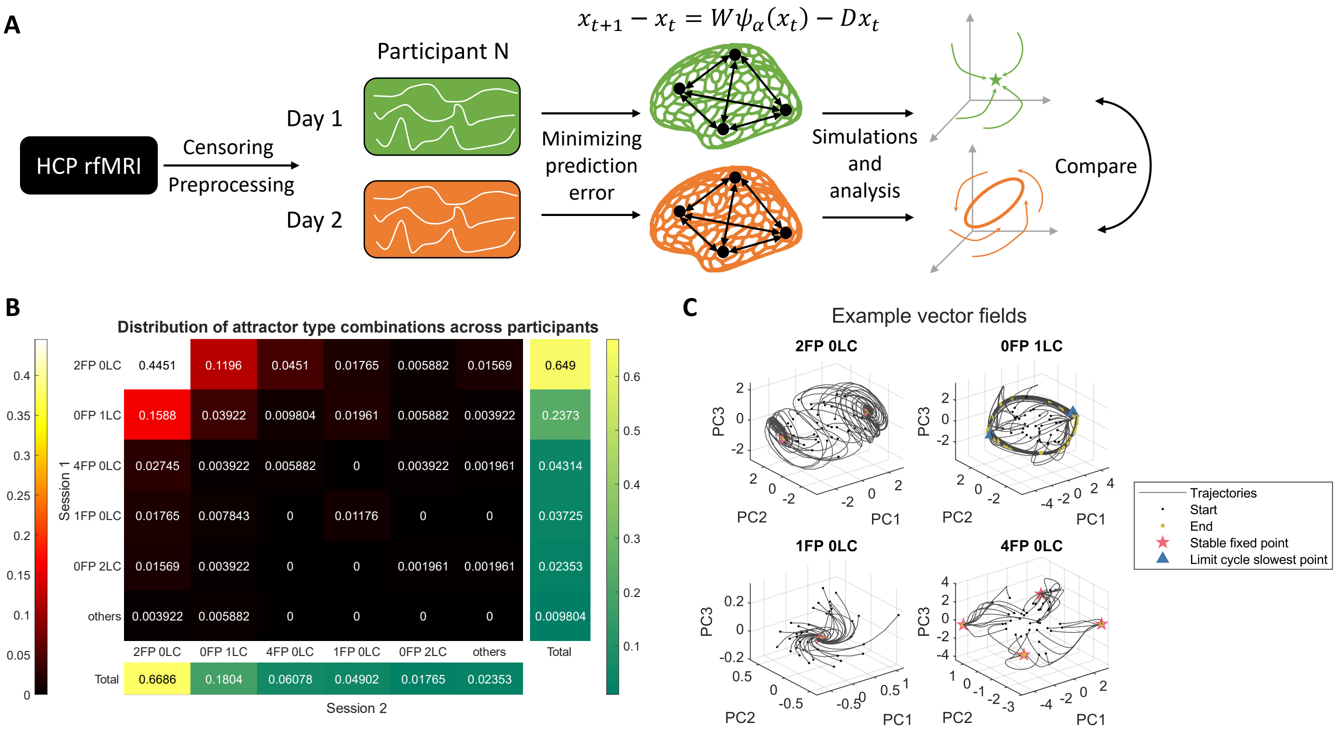
Here, we suggest that prior approaches have captured some, but not all of the critical aspects of resting state brain dynamics. We hypothesize that the resting brain is particularly sensitive to modulation, and as such, can manifest a spectrum of different dynamics that systematically vary across individuals and time. It has been suggested that the resting brain is close to bifurcation, such that a small change in control parameters will alter the stability of the trivial attractor located at the mean (2). However, it remains unknown whether both sides of the bifurcation can be observed in a same fMRI dataset, and whether such a bifurcation best characterizes differences between individuals, or state changes within individuals across different time periods. Previous models were either too constrained to express diverse sets of dynamics, or lack the specificity to describe individual differences and session-to-session variations. In this project, we overcome these prior limitations by adopting the Mesoscale Individualized NeuroDynamics (MINDy) framework (18). A key advantage of MINDy models is that they combine the expressiveness of nonlinear neural mass models with the flexibility and individuality of directly parameterized effective connectivity. Our prior work validated that MINDy models generate individualized, robust, and reliable fits of rfMRI data, with nontrivial dynamics observed (18). Here, we used the MINDy framework to analyze the taxonomy of resting state brain dynamics, and to more comprehensively characterize how they change across individuals and time. We fit MINDy models of rfMRI data from over five hundred participants and each of two scanning sessions in the Human Connectome Project (HCP) (19) to elucidate the dynamic profiles that best explained rfMRI signals. We then analyzed the existence of anatomically reliable attractors and ghost attractors, which are the signatures of a class of bifurcations, showing that the latter frequently occurs. Finally, we show that such attractors were consistent across the population and represent differential activation of well-known functional brain networks, such as the default mode network (DMN) and frontoparietal control network (FPN).

Results

Model parameters captured reliable individual differences. We obtained 1020 MINDy models, one for each of two rfMRI scanning session associated with 510 HCP participants (Figure 1A). The consistency of each parameter set (connectivity, curvature, decay) within individuals and across sessions was around 0.7-0.8. This quantity dropped to around 0.5-0.7 between individual, indicating that the obtained models were reliable and individualized (Supplementary Figure S3).

To perform an initial validity check that our models captured meaningful individual differences in the dynamics, we attempted to assess whether our obtained parameters can be connected to individual variation in cognitive measures. We performed Canonical Correlation Analysis (CCA) between the connectivity matrix of the models and the phenotypic measures available in the HCP dataset (20, 21). Interestingly, we obtained very similar results to (20), who used a connectivity matrix obtained via independent components analysis. We identified a single 'positive-negative mode' that was significantly correlated between the MINDy connectivity matrix and behavioral measures, and explained a significant proportion of variance for both. Post-hoc correlation found that this mode is most positively related to fluid intelligence, and most negatively related to substance use (Supplementary Figure S4). Therefore, the obtained MINDy models indeed capture reliable individual behavioral differences.

rfMRI embeds diverse nonlinear dynamics with nontrivial attractors. We next analyzed the asymptotic behavior of the neural trajectories. For this, we randomly initialized and forward simulated the models to obtain their limit sets and, specifically, reveal any asymptotically stable fixed points. As initially suggested in (18), the dynamics captured in our obtained MINDy models were highly nonlinear, with nontrivial



Session 1 \ Session 2	2FP 0LC	0FP 1LC	4FP 0LC	1FP 0LC	0FP 2LC	others	Total
2FP 0LC	0.4451	0.1196	0.0451	0.01765	0.005882	0.01569	0.649
0FP 1LC	0.1588	0.03922	0.009804	0.01961	0.005882	0.003922	0.2373
4FP 0LC	0.02745	0.003922	0.005882	0	0.003922	0.001961	0.04314
1FP 0LC	0.01765	0.007843	0	0.01176	0	0	0.03725
0FP 2LC	0.01569	0.003922	0	0	0.001961	0.001961	0.02353
others	0.003922	0.005882	0	0	0	0	0.009804
Total	0.6686	0.1804	0.06078	0.04902	0.01765	0.02353	

Fig. 1. Nonlinear dynamical landscapes underlying individual rfMRI data. (A) Diagram of the analysis pipeline. **(B)** Distribution of dynamical landscapes across the participants. Numbers indicate the proportion of participants showing certain type of dynamics in session 1 (column index) and session 2 (row index). The types were defined using the number and types of attractors. 'FP' indicates stable equilibrium (fixed point) and 'LC' indicates stable limit cycle. Types with frequency less than 0.5% were grouped into 'others' (see Supplementary Figure S5). Marginal distribution for session 1 and session 2 are shown in the panel on the right and at the bottom respectively. **(C)** Simulated trajectories and obtained attractors from example models, projected onto the first three principal components (PCs) of the trajectories. Black and yellow dots mark the initial and final state of each trajectory. Numerically identified stable equilibria are denoted by red stars. The slowest points on the limit cycles were denoted by blue triangles (see Methods). Note that the distribution of final states (yellow dots) on the limit cycle reflects the relative size of the speed. The denser the distribution, the slower the dynamics. We used 120 simulations per model to identify the attractors but only showed 40 here for visualization.

attractors and limit cycles found in the large majority of cases (Figure 1C). Less than 5% of models exhibited a single stable equilibrium at the origin (Figure 1B). It is worth noting that the observed nontrivial dynamics can not be simply attributed to any bias of our method, because MINDy models will actually recover trivial dynamics when fit on the noisy simulations of a stable linear system (Supplementary Figure S2). Most importantly, MINDy also correctly produced a globally attractive origin, rather than nontrivial attractors, when fit on noise with covariance and mean spectral power density that matches real data (Supplementary Figure S2). On the contrary, it is known that standard dFC methods cannot disambiguate such noise and actual data (9). Our findings thus support an interpretation of time-varying rfMRI activity as being most appropriately described emanating from a nonlinear dynamical system.

Induced bifurcations explained the heterogeneity of observed dynamics. Interestingly, although the obtained MINDy parameters were very reliable within each individual across sessions (Supplementary Figure S3), the number or type (fixed point or limit cycle) of attractors were still different in about half of all participants (Figure 1B). This finding indicates the existence of bifurcations, in which a small change in the model parameters results in a topological discontinuity of the dynamical landscape. Therefore, we hypothesize that resting state dynamics can be better described by a spectrum of possible dynamics that are sampled at each session, rather than a single monolithic brain state. To test the hypothesis, we constructed a parameterized 'interpolated' model as a convex combination of obtained models (i.e., $(\gamma)\text{Model}_1 + (1-\gamma)\text{Model}_2$), either within each participant (between the two models) or across all participants. If the hypothesis is correct, by varying γ, it should be possible to induce bifurcations. We analyzed the dynamics of the interpolated models and compared the distribution of the number and types of attractors with those of the original fitted models (Figure 2). The taxonomy of dynamics was very similar for fitted models and interpolated models, consistent with the idea that the dynamics we obtained are a reflection of whole brain dynamics that are undergoing bifurcations between topologically distinct vector fields.

Anatomically reliable individualized attractors marked signatures of bifurcations. Having established that the observed dynamics can be understood as bifurcating across a continuous spectrum, it is natural to ask how the brain can maintain stable or consistent function if the underlying dynamics are constantly changing. We thus hypothesized that there must be some anatomical commonality between the different obtained dynamics, leading to consistency in whole-brain activation patterns.

To probe this question, we looked for aspects of the dynamics that were relatively invariant to the bifurcations we observed. Most notably, we found that when the dynamics

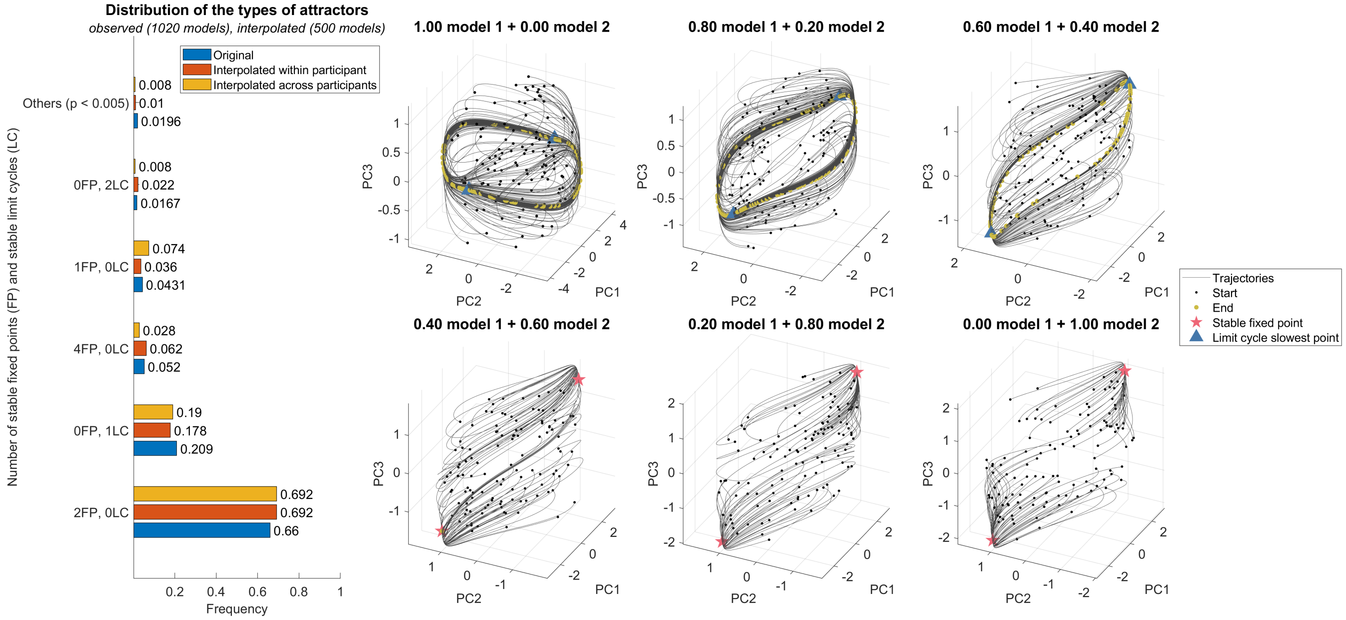


Fig. 2. Observed taxonomy of dynamics were consistent with bifurcation-induced continuous spectrum. Left: Distribution of the type of dynamics in linearly interpolated models and empirically fitted models. The interpolated models were obtained by convex combinations of the dynamics from two models from the same participant in different sessions ('within-participant') or randomly sampled from all models ('across-participant'). Right: Bifurcation process between two models of a same HCP participant. Trajectories and attractors were visualized in the same way as Figure 1. Note that the distribution of speed on the limit cycle (indicated by the distribution of final states) became more and more non-uniform as the first model bifurcated towards the second model.

differed across sessions for a same person, they most commonly switched between having two equilibria or having one limit cycle (Figure 1B). This is consistent with the well-known infinite-period bifurcation (22). An infinite-period bifurcation begins with a limit cycle that contains a ghost attractor. As the bifurcation parameter changes, the ghost attractor becomes 'infinitely slow' (i.e., neural activity lingers near it for long periods of time) until eventually the bifurcation occurs and a stable node (along with an saddle node) gets created out of it (see Supplementary Figure S6 for an analytic toy model). In our data, when limit cycles were observed, the distribution of speed along them was highly non-uniform, sometimes varying by orders of magnitude (Figure 3, left). In our interpolated models, this form of bifurcation indeed occurs (Figure 2, right panel). Therefore, we hypothesize that the ghost attractors and point attractors provide a set of stable 'operating points' for the changing dynamics during rest.

If the session-to-session variability in fitted dynamics can be explained by such a bifurcation, we would expect that ghost attractors should be close to the stable equilibria in the other session, i.e., they should represent anatomically similar activation patterns. We thus defined ghost attractors as the slowest point on each limit cycle, and calculated the anatomical similarity between all (true and ghost) attractors. Because the number of attractors might differ across sessions, we defined the *dominant attractor similarity (DAS)* to be the maximum correlation over all pairs of attractors from the two models under consideration (see Methods). The DAS was higher within subject versus across subject (Figure 3, right), even when restricting the former to models showing different types of dynamics (e.g., two equilibria or one limit cycle) and the latter to models showing same type of dynamics (the second and the third boxes in the middle in Figure 3, right). Therefore, our results support the hypothesis that the resting brain is bifurcating between different dynamics, while maintaining a set of reliable attractors as operating points.

Attractors aligned with canonical resting state networks across population. Next, we examined whether these attractors could be interpreted from a functional standpoint. Interestingly, the DAS was high (around 0.5) even between different participants (Figure 3), indicating the existence of consistent patterns across the whole population. Therefore, we clustered the locations of the attractors across all participants and sessions, where the number of clusters K were selected according to cluster instability index (Methods, (23)). We obtained near perfect cluster stability only for $K = 4$, apart from the less interesting solution $K = 2$. Note that the attractors always exist in pairs (see Methods) so $K = 2$ represents only one pattern and its opposite (Supplementary Figure S7, S8).

The individual attractors within each cluster aligned very well with the cluster centers (Figure 4). Even more interestingly, the activation patterns were highly modular, respecting the functional network organization defined in the (24) atlas, even though the model fitting process was completely agnostic to parcel labeling. In particular, one large cluster was dominated by the activation of DMN and FPN. Another cluster was dominated by the FPN and dorsal/ventral attention networks. The other two clusters showed the opposite activation profiles. The clusters and network organization in combination accounted for more than 40% of the total variation across all participants, sessions, and parcels (Supplementary Figure S11). The activation pattern was better explained by functional network organization rather than the spatial proximity of parcels on the cortical surface (Supplementary Figure S12). Furthermore, the attractor clusters emerged regardless of whether we only included the equilibria, the limit cycle ghost attractors,

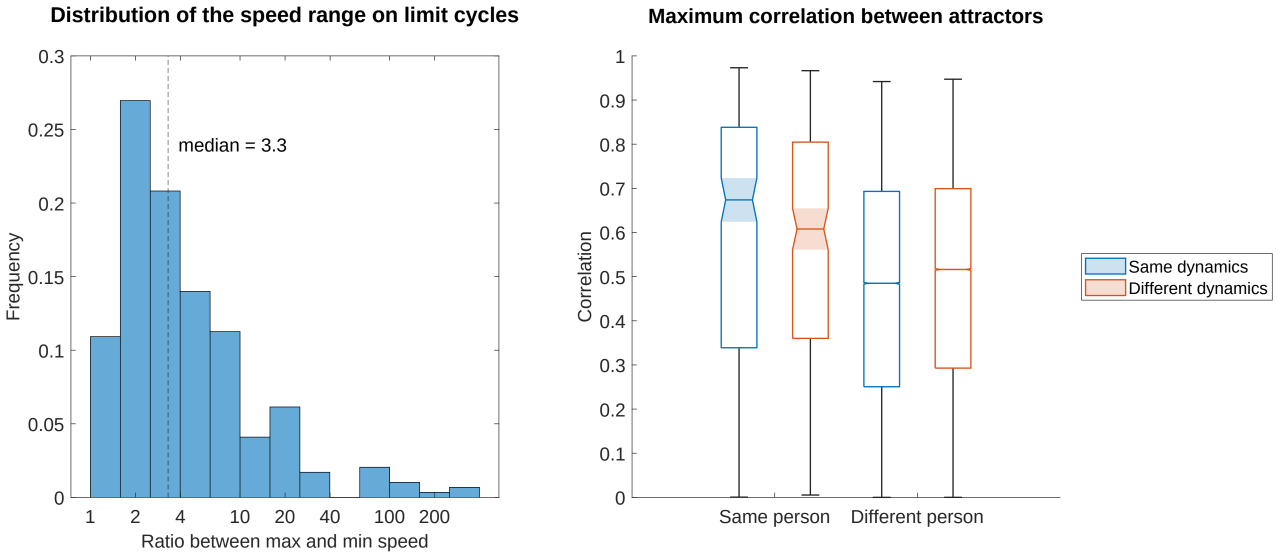


Fig. 3. Attractors were more similar within than across individual regardless of changes in dynamical landscapes. Left: Distribution of the speed variability on limit cycles. For each limit cycle in all fitted models, the ratio between the maximum and minimum speed on the limit cycles was calculated. Right: box plot for the distribution of maximum attractor correlation between all pairs of models, separated by whether the two models come from a same person and whether they have the same type of dynamics (see Methods). Lines on the boxes indicate the maximum, first quartile, median, last quartile, and minimum of each distribution. Tapered, shaded region around the median indicates 95% confidence interval of the median.

or both (Supplementary Figure S9, S10), supporting the hypothesis that they represent the reliable activation dynamics of resting state networks.

Discussion

In this study, we adopted a nonlinear dynamical systems modeling framework to analyze how resting state brain dynamics varied across individuals and time. We found that resting state brain activity embedded diverse nonlinear dynamics that included nontrivial attractors rather than a single globally stable equilibrium at the mean. Interestingly, the dynamics reliably varied between individuals. Moreover, instead of being stationary, the dynamics underwent bifurcations across different scanning sessions even within the same individual. Furthermore, the observed spectrum of dynamics could be fully recovered through induced bifurcations between fitted models. Consistent with such bifurcations, the attractors and ghost attractors were anatomically reliable within and between individuals. These attractors were organized into distinct clusters that reflected the activation of different functional brain networks, particularly the DMN, the FPN, and the dorsal/ventral attention networks. Using the formal language of dynamical systems and bifurcation theory, our results shed light on the connection between nontrivial fluctuations in resting state activity, the organization of functional brain networks, and stable individual differences. We provide a modeling and analysis framework that describes the individualized nontrivial dynamics in rfMRI, while maintaining interpretability and tractability. A variety of dynamical features can be derived from the models for future brain-phenotype association studies. Our results also enable model-based interventions into brain dynamics, which might hold great potential for individualized treatments of brain disorders and even cognitive enhancement (25).

Nontrivial attractors and resting state networks. Resting state brain activity is traditionally treated as a unimodal stationary process fluctuating around a stable mean. However, recent studies have begun to explore nontrivial recurring patterns in resting state dynamics (26, 27). Mostly notably, time-varying functional connectivity (tvFC, also called ‘dynamic’ FC) studies have suggested that the resting brain is traversing multiple states associated with distinct brain-wide association patterns (3). However, it is debated whether these recurring patterns indeed represent nontrivial fluctuations, or merely capture the snapshots of the trivial fluctuations around the mean (9). Due to these challenges, it has been questioned whether the new insights we gain from these techniques are relatively marginal compared to the elevated difficulties in analysis and interpretation (16).

We argue that the controversies over non-trivial fluctuations reflect the lack of mechanistic interpretations that can be derived from short-time-windowed methods typically utilized to estimate tvFC. Dynamic-systems-based modeling provides a more generative explanation to the observed dynamics and statistics. Phase portraits of the fitted models reveal insights about the resting brain dynamics without sacrificing the ease and interpretability of analysis. Here, by using a nonlinear, individualized, and fully trainable dynamical systems modeling framework, we provide evidence for nontrivial fluctuations in rfMRI dynamics. Most importantly, when fitted on stationary noise with FC and a mean power spectrum that match real data, our model correctly recreated a monostable dynamic system, while tvFC has been demonstrated to produce spurious nontrivial states in this situation (9). Therefore, the nontrivial attractors consistently observed in our fitted models are less likely to be explained by a bias inherent in the method. Compared to other modeling studies, our model relies on fewer assumptions about connectivity structure. Instead of assuming that white matter density is a good surrogate of functional

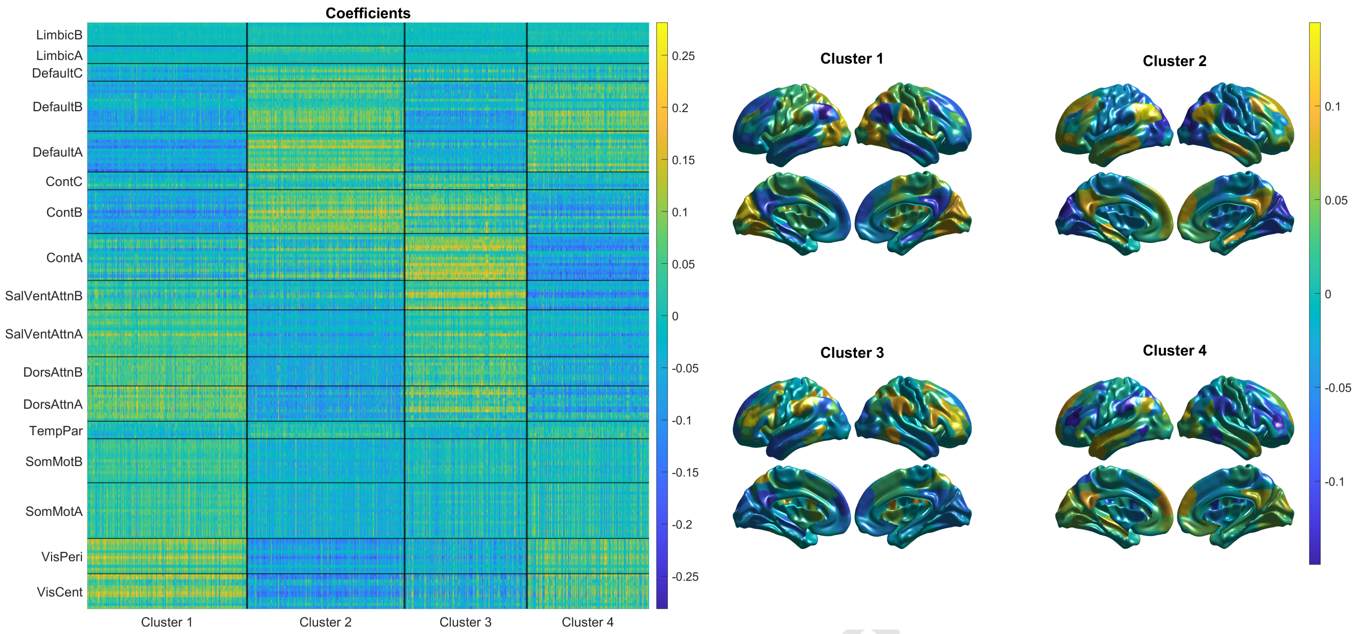


Fig. 4. Attractors aligned with canonical resting state networks across the population. Left: locations (coefficients) of all attractors. Each row corresponds to one parcel and each column corresponds to one attractor. Rows are sorted according to the (24) atlas and columns are sorted according to cluster assignments. Thick horizontal and vertical lines separates the functional networks and clusters, respectively. Attractors are scaled to unit norm for visualization purpose. Right: cluster centroids visualized as the activation patterns over the cortex.

coupling at the mesoscale level, which has been challenged by some studies (28), we posit a 'sparse plus low rank' structure with all the weights directly optimized towards empirical fMRI timeseries. Very interestingly, even without prior constraint on the functional network structure, the attractors that emerged from the models still exhibited highly modular activation patterns that respect the organization of functional networks. Such convergence of evidence is further supported by the fact that, the MINDy effective connectivity between localized parcels and the functional connectivity between distributed ICA-defined networks, despite using very different node types and connectivity measures, encoded highly similar information about cognitive individual differences as revealed by CCA (Supplementary Figure S4). Furthermore, unlike most previous work that has relied upon population-level models, we were able to show that the nontrivial attractors are not only consistent across the population, but also test-retest reliable within each individual.

It has long been hypothesized that nontrivial attractors represent 'functional' states that can be spontaneously traversed during rest, as if exploring the repertoire of operating points (29). Although there has been some theoretical work focused on the functional relevance of resting state attractors (30), empirical evidence has been scarce. Our work supports this hypothesis by showing that nontrivial resting state attractors reflect selective activation of functional brain networks, and contain reliable individual differences. It will thus be very interesting to extend the MINDy framework to task states, in order to analyze how such nontrivial attractors might be engaged in cognitive computation. Similar to DCM methods (14), we can couple the MINDy recurrent dynamics with an input term representing the task control signal. The interaction between task demand and inherent dynamics and attractors can thus be potentially characterized through control-theory-based analysis. Such analysis will shed light on how the resting brain 'prepares' stable motifs for cognitive computation, advancing of our understanding of the mechanistic link between the resting state and cognition.

Bifurcations and the critical brain. We presented three lines of evidence that resting state brain dynamics are not only nontrivial, but also bifurcating. First, the topology of the dynamic landscapes changed across sessions even when the controlling parameters remained highly reliable, consistent with the definition of bifurcations. Second, we induced bifurcations between fitted models and recovered the full spectrum of the dynamics observed, confirming that the fitted models can be understood as samples from such a continuous spectrum encompassing several bifurcations. Third, we identified anatomically reliable attractors or ghost attractors on a cycle, consistent with the prediction of an infinite-period bifurcation. Our results thus provide a richer description of both the invariants and changes in resting state dynamics, showing that even though the statistical outputs (FC) of two datasets might seem similar, they could be supported by distinct but still intimately related dynamics.

The most interesting conclusion from our analysis is that the resting state brain is highly sensitive to modulation, in which a small perturbation can bifurcate it towards various different dynamics. To illustrate this point, we first showed that even though the correlation between the parameters of any two models was high (around 0.6), we still observed at least eight different kinds of dynamics, in the sense of different numbers or types of attractors. Moreover, we found that although the parameter correlation within each participant was even higher (around 0.75), the dynamics were still different between sessions for almost half of all participants.

We suggest that the sensitivity of resting brain dynamics might provide one mechanism for cognitive flexibility, as the brain can easily bifurcate from the resting state towards various different dynamics that might be advantageous for different cognitive computations. If that is the case, the task state brain dynamics should be more rigid than resting state, and should vary across tasks according to the computation required. In line with this idea, it was found that task-driven input reduced the variability of whole-brain dynamics compared to rest (31, 32). Further, the global connectivity of the frontoparietal network was found to systematically vary across 64 tasks, with more similar connectivity for tasks that require more similar computation (33). Most interestingly, it has been found that the similarity between rsFC and task-specific FC positively correlated with task performance across multiple tasks, and the similarity between rsFC and task-general FC can be related to fluid intelligence (34), indicating that cognitive ability is related to how efficiently resting state dynamics can transform into a variety of different task dynamics. Our study provides a more mechanistic framework that can capture the changes in the spectrum of generative dynamics, not just the statistics of such dynamics (i.e., FC). Extending the MINDy framework to task contexts will thus provide a strong test for such hypothesis.

Our results also relate to the idea of criticality in the brain. Note that the word ‘criticality’ has been used in at least two different ways in neuroscience. In the broad sense, criticality refers to the emergence of slow, large-amplitude (scale-free) fluctuations in a dynamical system that is close to losing its stability, without specifying the kind of instability to which it is transformed (35). In a narrow sense, such transformation is restricted to that between an ordered state (stability) and an unordered state (chaos) (36). Our results support criticality of rfMRI dynamics in the broad sense rather than the narrower sense. We found that our models show either a stable low-activity state with no nontrivial attractors, or an unstable low-activity state with nontrivial attractors, suggesting the existence of supercritical bifurcations where a high-activity attractor emerged ‘above’ the low-activity attractor as the latter loses its stability. Such bifurcations have been shown to give rise to slow, scale-free fluctuations (35). In previous models of whole-brain dynamics, it has been found that the criticality associated with such bifurcations improved the response sensitivity to external stimuli (29). However, critical dynamics are not always optimal for all tasks. For example, the sensitivity to inputs also reduces the reliability of the response (37). Therefore, it is proposed that instead of staying critical, brain dynamics should reverberate between multiple regimes near criticality (38). Our results support this hypothesis with novel evidence that the brain traverses near criticality, potentially balancing the computational advantages of each regime across different computations.

If the resting brain resides in this critical regime that can be easily modulated into different dynamics, one interesting question is how the brain might implement such modulation to utilize different dynamics. On a longer timescale, such as the dynamics we obtained here across a 30-minute resting state scan, neuromodulator systems might be the best candidate. It is well established that neuromodulators can affect brain connectivity and whole-brain dynamics (39, 40). The arousal system might play a particularly important role in the fluctuations of resting state dynamics (9). On a shorter timescale, such as during the execution of cognitive tasks, goal-directed top-down modulation might also bifurcate the dynamics. Theoretically, such bifurcations might relate to the proactive control mode in the dual mechanism framework for cognitive control (41). In contrast to reactive control, proactive control refers to the active maintenance of goal-related information and biasing cognitive computations even before cognitively demanding events occur. In parallel, goal-directed bifurcation might transform the dynamical landscape to bias the neural processing even before receiving cognitive inputs. Therefore, combining MINDy modeling with cognitive tasks as well as neuromodulatory manipulations holds great potential to further our understanding about the dynamical flexibility of the brain.

Limitations and future directions. In this study, we characterized the variable nature of resting state dynamics by comparing across participants and scanning sessions. However, such variation inevitably interweaves with measurement and modeling error. A stronger test for the non-stationarity of resting state dynamics requires comparing the dynamics across different periods within a single scanning run. Such analysis is difficult to applied to the HCP which only has 15-minute long scans, roughly as much as the data needed to obtain a reliable MINDy model (18). However, it can be done on datasets with as long as 30 contiguous minutes of rfMRI data, such as the Midnight Scanning Club (42). Another approach is to use EEG or MEG data which have much lower dimension and much higher sampling rate. It might be even possible to update the model parameters in real time for EEG/MEG data (43).

There are also limitations associated with the MINDy model used in the analysis. The models did not include an intercept/bias term, and the dynamics are anti-symmetric with respect to the neural hidden states (see Methods). Such assumption was made according to the excitation-inhibition-balance principle (44) and has been adopted in resting state DCM studies too (15). In such models, the zero vector (corresponding to the mean of the data) is always an equilibrium (though not necessarily stable) and nontrivial attractors (if any) must exist in pairs. Extending MINDy to task contexts by introducing a task-related input can break such symmetry. Another limitation of the current method is the ability to account for the variations in haemodynamics. Here, we estimated the hidden neural states by a noise-aware deconvolution of BOLD signal with the canonical haemodynamic response function (HRF) (45). However, it has been suggested that HRF varies significantly across brain regions and individuals (46). Although MINDy parameters have been demonstrated to be robust against HRF variations (18), it is unclear how much the emergent dynamics will be influenced, given the observation that the dynamics were sensitive to parameters. Extending our analysis with region-specific HRF (47) is a natural step to follow. Another direction is to use a biologically more detailed regional dynamics model (14). Currently, the recurrent dynamics within each brain region is modeled as a simple self-excitation (or inhibition) with an exponential decay. Despite being computationally more tractable, such a model might not capture the full dynamics within a region, especially the interactions between sub-populations (43). Multi-scale modeling and EEG-fMRI data fusion might be a possible way to improve biological specificity while maintaining computational efficiency.

From a cognitive neuroscience perspective, the current study demonstrates how the novel lens of dynamical systems and attractor landscapes can be utilized for theory and analysis regarding the relationship between intrinsic dynamics (i.e., present during resting states) and task-related cognitive computations. Future studies can try to associate individualized resting state dynamical motifs with cognitive traits and task performance using either traditional correlational analysis, or more interestingly, by adapting the MINDy framework to analyze how dynamics and attractors change between resting and task states. As a mechanistic alternative to FC and dFC analysis, our framework can also generate new insights into the dynamical changes associated with different dimensions of cognitive variation, including states of consciousness (e.g., sleep, meditation, psychedelics), developmental stages, or dysfunction associated with psychiatric and neurological disorders.

Materials and Methods

Data preprocessing. We obtained rfMRI data from the HCP Young Adult dataset (19). Data was originally collected on a 3T scanner with a TR of 720ms and 2mm isotropic voxels. Participants underwent two scanning sessions on separate days. Each session included two scanning runs of 1200 TRs (around 15 minutes), one using right-to-left phase encoding direction and the other left-to-right. Participants were instructed to stay awake with eyes open and relax fixation on a bright cross hair on a dark background, presented in a darkened room.

We adopted the preprocessing pipeline suggested in (48), which was shown to effectively suppress the influence of head motion in rsFC-behavior associations. Because we are particularly interested in session-to-session variations in dynamics, we used a relatively strict inclusion criteria to make sure the data in all runs were sufficiently clean. Specifically, we only included participants with no missing runs or runs that had more than 1/3 (400 out of 1200) frames with high head motion (see below), resulting in a total number of 510 participants.

We began with the rfMRI data provided by HCP that has been minimally preprocessed, motion-corrected, and denoised with FIX-ICA (19). Following (48), data was first detrended and then motion scrubbed with framewise displacement (FD) and temporal derivative of variation (DVARS) (49). FD and DVARS were filtered for respiratory artifact with a 40-th order 0.06-0.14Hz band stop filter. Frames with FD above 0.2mm or DVARS above 1.05 times of the median were linearly interpolated. We then regressed out from the data the top five principal components of the white matter and the cerebrospinal fluid signals (CompCor), and the mean signal from the gray matter.

After preprocessing, the data was averaged within each parcel according to the 200-parcel atlas from (24). Data points exceeding 5 standard deviations in each time series were linearly interpolated. To obtain the underlying neural activity, we deconvolved the data with the canonical haemodynamic response function (HRF) from (45) using Wiener deconvolution (50), a deconvolution technique that minimizes the influence of noise. We used a 30-point HRF kernel and a noise-power to signal-power ratio of 0.02. Finally, the data were z-scored within each timeseries.

Model architecture and fitting. We adopted the Mesoscale Individualized NeuroDynamics (MINDy) framework from (18). A MINDy model contains interconnected neural masses representing brain parcels, with trainable and individualized connection weights. Each neural mass is assumed to follow an S-shape input-output transfer function with a trainable region-specific curvature. Activity decays with a trainable region-specific rate. The dynamics of the model is thus:

$$d\mathbf{x}(t) = W\psi_\alpha(\mathbf{x}(t)) - D \odot \mathbf{x}(t) \qquad [1]$$

$$\psi_\alpha(x) = \sqrt{\alpha^2 + (bx+0.5)^2} - \sqrt{\alpha^2 + (bx-0.5)^2} \qquad [2]$$

$\mathbf{x}(t) \in \mathbb{R}^N$ is the neural activity hidden state at time t, where N is the number of parcels. $W \in \mathbb{R}^{N\times N}$ is the connectivity matrix. The transfer function ψ_α is applied element-wise with each region's respective curvature parameter α. $D \in \mathbb{R}^N$ is the decay and $\odot$ indicates element-wise multiplication. b is another parameter controlling the shape of the transfer function, currently fixed as $20/3$. To prevent overfitting and improve interpretability, we required the connectivity matrix W to be the sum of a sparse matrix W_S and a low-rank matrix $W_L = W_1 W_2^T$, where $W_1, W_2 \in \mathbb{R}^{N\times k}$ and $k < N$. Here we chose $N = 200$ ($k = 72$) as it achieves a balance between granularity and computational efficiency. However, we also repeated the analysis with $N = 100$ and $N = 400$ and obtained very similar attractor motifs.

We obtained $\mathbf{x}(t)$ and ground-truth $d\mathbf{x}(t)$ from the preprocessed rfMRI data for each participant and session, with two runs combined. The derivative was computed using forward differentiation. Since HCP data utilizes an exceptionally fast TR, we performed two-point moving average smoothing on the derivative to reduce noise, i.e., $d\mathbf{x}(t) = [\mathbf{x}(t+2) - \mathbf{x}(t)]/2$. Removing the smoothing led to very similar results. We optimized W, α and D to minimize the model's prediction error while enforcing sparsity of the connectivity using L1 regularization (51). The loss function is thus defined as:

$$\begin{aligned} J = &\frac{1}{2}||d\mathbf{x}(t) - \left[(W_S + W_1 W_2^T)\psi_\alpha(\mathbf{x}(t)) - D \odot \mathbf{x}(t)\right]||_2^2 \\ &+ \lambda_1||W_S||_1 + \lambda_2 Tr(|W_S|) + \lambda_3(||W_1||_1 + ||W_2||_1) \end{aligned}$$

where $\lambda_1 = 0.075$, $\lambda_2 = 0.2$, $\lambda_3 = 0.05$ are the regularization hyperparameters, $Tr(|W_S|)$ is the absolute sum of the diagonal elements of W_S. Optimization was performed using gradient descent with Nesterov Accelerated Adaptive Moment Estimation (NADAM, (52)). Parameters were updated after each minibatch of 300 random samples. We stopped the training at 5000 minibatches when the test-retest reliability of the parameters started to drop. To prevent the weights from being unnecessarily small due to regularization penalty, we performed an additional global rescaling of the parameters by fitting $d\mathbf{x} = p_W W\psi_\alpha(\mathbf{x}) - p_D D \odot \mathbf{x}(t)$ with two scalar parameters $p_W, p_D \in \mathbb{R}$, and factored them into W and D. It's worth noting that a 200-parcel MINDy model can be fit on a standard laptop within 15 seconds, enabling the analysis of the whole HCP dataset within a reasonable amount of time.

Model simulation and numerical analysis of attractors. For each model, we randomly sampled 120 initial conditions from the standard normal distribution. We also tried to sample from the data $\mathbf{x}(t)$, but the identified attractors were the same as long as the sample size is large enough. The dynamics (equation 1) was integrated using Euler's method for 1600 TRs with step size equal to one TR. For some models with limit cycles that contained extremely slow ghost attractors, we prolonged the simulation until the state recurred after a full cycle. The stable equilibria were identified as follows: First, we defined a trajectory (simulation) as already converged to a stable equilibrium if $|x(t+1) - x(t)| < 10^{-6}$ for every parcel and every time point $T - 10 \leq t \leq T - 1$, where T is the simulation time. The terminal state $x(T)$ from all converged trajectories were than clustered together based on a simple Euclidean distance threshold of 0.1. The cluster centroids were extracted as the stable equilibria. Similarly, we defined a trajectory as already converged to a stable limit cycle if it approached and then left a small neighbourhood of the terminal state $\{x \mid ||x - x(T)||_2 < 0.5\}$ at least once. The interval during the last recurrence and the end of the simulation was considered as the period of the limit cycle, and the samples during this period was extracted to represent the limit cycle. We confirmed the validity of the method by visually inspecting the trajectories and identified attractors after dimensionality reduction using Principal Component Analysis (PCA), as depicted in Figure 1 and 2.

We observed that the distribution of speed on the limit cycles was very non-uniform, sometimes varying by orders of magnitude. Therefore, we selected the slowest point on each limit cycle

$\arg\min_{\mathbf{x}(t)} ||\mathbf{x}(t+1)-\mathbf{x}(t)||_2$ as the ’ghost’ attractors. Besides, for models with one limit cycle, due to the symmetry of the dynamics $f(-\mathbf{x})=-f(\mathbf{x})$, we added a pair of symmetric slowest points rather than a single one.

Bifurcation analysis. We induced a bifurcation between two models by creating convex combinations of the dynamics. Denoting the two model’s dynamics as $\mathbf{x}(t+1)-\mathbf{x}(t)=f_1(\mathbf{x})$ and $\mathbf{x}(t+1)-\mathbf{x}(t)=f_2(\mathbf{x})$ respectively, we construct a new model as $\mathbf{x}(t+1)-\mathbf{x}(t)=\alpha f_1(\mathbf{x})+(1-\alpha)f_2(\mathbf{x})$, where $0\leq\alpha\leq 1$ is the bifurcation parameter. We considered two ways to combine the observed models. In the within-individual case, f_1 and f_2 were the two models from the two sessions of a same participant. In the across-individuals cases, f_1 and f_2 were sampled from all models. We randomly sampled α from the uniform distribution in $[0,1]$ as well as a pair of models for 500 times, and extracted the attractors of these bifurcated models using the same numerical method described above. We then characterized the vector field by the number and types (equilibria or limit cycles) of attractors and compared their distribution across the bifurcated models and the fitted models in Figure 2.

Reliability analysis. We quantified the anatomical similarity of two attractors (and ghost attractors, same below) by the Pearson correlation between their coefficients (i.e., anatomical projections). As the number of attractors might differ across models, we defined the *dominant attractor similarity (DAS)* between two models as the maximum similarity between their attractors. The distribution of DAS between all pairs of models from different sessions was shown in Figure 3, separated by whether the two models come from a same participant and whether they have same type of dynamics (i.e., same number of stable equilibria and limit cycles).

Clustering analysis. We used K-means algorithm to cluster the anatomical location of all attractors and ghost attractors across all participants and sessions. We used cosine distance $d(x,y)=1-\cos\angle(x,y)$ for clustering and scaled the samples to unit norm for visualization. Using Euclidean distance produced almost identical clustering results. The number of clusters K is determined by the cluster instability index (23) with a candidate list of K ranging from 2 to 10. For each K, the data was randomly divided into two subsets for 30 times. We ran K-means algorithm on the first partition to obtained the cluster centroids, and used these centroids to classify the samples in the second partition. The classification result was compared to the results of directly running K-means on the second partition. A misclassification cost was computed after matching the labels using the Hungarian algorithm. This cost is then averaged across the 30 partitions and normalized by the null cost computed in the same way but with random labels, resulting in an instability index for each K. The local minimum of instability was selected as the number of clusters for the final clustering. We also repeated our analysis using either only the stable equilibria or only the ghost attractors on the limit cycles (Supplementary Figure S9, S10), and in each case both the clustering instability index and cluster centroids were very similar to the results in the main text.

Data and code availability. The rfMRI data is available at Human Connectome Project’s website. The MINDy modeling toolbox is available at https://github.com/singhmf/MINDy. The preprocessing and analysis scripts will be available after peer review.

ACKNOWLEDGMENTS. Portions of this work were funded by NIMH 1R21MH132240-01 and NINDS R01NS130693.

1. B Biswal, F Zerrin Yetkin, VM Haughton, JS Hyde, Functional connectivity in the motor cortex of resting human brain using echo-planar mri. *Magn. Reson. Medicine* **34**, 537–541 (1995).
2. G Deco, et al., Resting-State Functional Connectivity Emerges from Structurally and Dynamically Shaped Slow Linear Fluctuations. *J. Neurosci.* **33**, 11239–11252 (2013).
3. DJ Lurie, et al., Questions and controversies in the study of time-varying functional connectivity in resting fMRI. *Netw. Neurosci.* **4**, 30–69 (2020).
4. AS Choe, et al., Comparing test-retest reliability of dynamic functional connectivity methods. *NeuroImage* **158**, 155–175 (2017).
5. A Abrol, et al., Replicability of time-varying connectivity patterns in large resting state fMRI samples. *NeuroImage* **163**, 160–176 (2017).
6. JM Reinen, et al., The human cortex possesses a reconfigurable dynamic network architecture that is disrupted in psychosis. *Nat. Commun.* **9**, 1157 (2018).
7. JD Power, KA Barnes, AZ Snyder, BL Schlaggar, SE Petersen, Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage* **59**, 2142–2154 (2012).
8. S Heitmann, M Breakspear, Putting the “dynamic” back into dynamic functional connectivity. *Netw. Neurosci.* **02**, 150–174 (2018).
9. TO Laumann, et al., On the Stability of BOLD fMRI Correlations. *Cereb. Cortex* **27**, 4719–4732 (2017).
10. G Deco, VK Jirsa, Ongoing Cortical Activity at Rest: Criticality, Multistability, and Ghost Attractors. *J. Neurosci.* **32**, 3366–3375 (2012).
11. M Golos, V Jirsa, E Daucé, Multistability in Large Scale Models of Brain Activity. *PLOS Comput. Biol.* **11**, e1004644 (2015).
12. V Sip, et al., Characterization of regional differences in resting-state fMRI with a data-driven network model of brain dynamics. *Sci. Adv.* **9**, eabq7547 (2023).
13. KJ Friston, L Harrison, W Penny, Dynamic causal modelling. *NeuroImage* **19**, 1273–1302 (2003).
14. KJ Friston, et al., Dynamic causal modelling revisited. *NeuroImage* **199**, 730–744 (2019).
15. A Razi, et al., Large-scale DCMs for resting-state fMRI. *Netw. Neurosci.* **1**, 222–241 (2017).
16. E Nozari, et al., Macroscopic resting-state brain dynamics are best described by linear models. *Nat. Biomed. Eng.* pp. 1–17 (2023).
17. HJ Park, KJ Friston, C Pae, B Park, A Razi, Dynamic effective connectivity in resting state fMRI. *NeuroImage* **180**, 594–608 (2018).
18. MF Singh, TS Braver, MW Cole, S Ching, Estimation and validation of individualized dynamic brain models with resting state fMRI. *NeuroImage* **221**, 117046 (2020).
19. SM Smith, et al., Resting-state fMRI in the Human Connectome Project. *NeuroImage* **80**, 144–168 (2013).
20. SM Smith, et al., A positive-negative mode of population covariation links brain connectivity, demographics and behavior. *Nat. Neurosci.* **18**, 1565–1567 (2015).
21. N Goyal, D Moraczewski, P Bandettini, ES Finn, A Thomas, Computationally replicating the Smith et al. (2015) positive-negative mode linking functional connectivity and subject measures (2020).
22. JP Keener, Infinite Period Bifurcation and Global Bifurcation Branches. *SIAM J. on Appl. Math.* **41**, 127–144 (1981).
23. T Lange, V Roth, ML Braun, JM Buhmann, Stability-Based Validation of Clustering Solutions. *Neural Comput.* **16**, 1299–1323 (2004).
24. A Schaefer, et al., Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cereb. Cortex* **28**, 3095–3114 (2018).
25. MF Singh, MW Cole, TS Braver, S Ching, Developing control-theoretic objectives for large-scale brain dynamics and cognitive enhancement. *Annu. Rev. Control.* **54**, 363–376 (2022).
26. X Liu, JH Duyn, Time-varying functional network information extracted from brief instances of spontaneous brain activity. *Proc. Natl. Acad. Sci.* **110**, 4392–4397 (2013).
27. R Abreu, J Jorge, A Leal, T Koenig, P Figueiredo, EEG Microstates Predict Concurrent fMRI Dynamic Functional Connectivity States. *Brain Topogr.* **34**, 41–55 (2020).
28. P Barttfeld, et al., Signature of consciousness in the dynamics of resting-state brain activity. *Proc. Natl. Acad. Sci.* **112**, 887–892 (2015).
29. G Deco, VK Jirsa, AR McIntosh, Resting brains never rest: Computational insights into potential cognitive architectures. *Trends Neurosci.* **36**, 268–274 (2013).
30. T Kurikawa, K Kaneko, Embedding Responses in Spontaneous Neural Activity Shaped through Sequential Learning. *PLOS Comput. Biol.* **9**, e1002943 (2013).
31. A Ponce-Alvarez, BJ He, P Hagmann, G Deco, Task-Driven Activity Reduces the Cortical Activity Space of the Brain: Experiment and Whole-Brain Modeling. *PLOS Comput. Biol.* **11**, e1004445 (2015).
32. T Ito, et al., Task-evoked activity quenches neural correlations and variability across cortical areas. *PLOS Comput. Biol.* **16**, e1007983 (2020).
33. MW Cole, et al., Multi-task connectivity reveals flexible hubs for adaptive task control. *Nat. Neurosci.* **16**, 1348–1355 (2013).
34. DH Schultz, MW Cole, Higher Intelligence Is Associated with Less Task-Related Brain Network Reconfiguration. *J. Neurosci.* **36**, 8551–8561 (2016).
35. L Cocchi, LL Gollo, A Zalesky, M Breakspear, Criticality in the brain: A synthesis of neurobiology, models and cognition. *Prog. Neurobiol.* **158**, 132–152 (2017).
36. J O’Byrne, K Jerbi, How critical is brain criticality? *Trends Neurosci.* **45**, 820–837 (2022).
37. ED Fagerholm, et al., Cascades and Cognitive State: Focused Attention Incurs Subcritical Dynamics. *J. Neurosci.* **35**, 4626–4634 (2015).
38. J Wilting, et al., Operating in a Reverberating Regime Enables Rapid Tuning of Network States to Task Requirements. *Front. Syst. Neurosci.* **12** (2018).
39. G Shafiei, et al., Dopamine Signaling Modulates the Stability and Integration of Intrinsic Brain Networks. *Cereb. Cortex* **29**, 397–409 (2019).
40. JM Shine, RL van den Brink, D Hernaus, S Nieuwenhuis, RA Poldrack, Catecholaminergic manipulation alters dynamic network topology across cognitive states. *Netw. Neurosci.* **2**, 381–396 (2018).
41. TS Braver, The variable nature of cognitive control: A dual mechanisms framework. *Trends Cogn. Sci.* **16**, 106–113 (2012).
42. EM Gordon, et al., Precision Functional Mapping of Individual Human Brains. *Neuron* **95**, 791–807.e7 (2017).
43. MF Singh, TS Braver, M Cole, S Ching, Precision data-driven modeling of cortical dynamics reveals idiosyncratic mechanisms of canonical oscillations (2023).
44. C van Vreeswijk, H Sompolinsky, Chaos in Neuronal Networks with Balanced Excitatory and Inhibitory Activity. *Science* **274**, 1724–1726 (1996).
45. KJ Friston, et al., Event-Related fMRI: Characterizing Differential Responses. *NeuroImage* **7**, 30–40 (1998).
46. GK Aguirre, E Zarahn, M D’Esposito, The Variability of Human, BOLD Hemodynamic Responses. *NeuroImage* **8**, 360–369 (1998).
47. MF Singh, A Wang, TS Braver, S Ching, Scalable surrogate deconvolution for identification of partially-observable systems and brain modeling. *J. Neural Eng.* **17**, 046025 (2020).
48. JS Siegel, et al., Data Quality Influences Observed Links Between Functional Connectivity

and Behavior. *Cereb. Cortex* **27**, 4492–4502 (2017).

49. S Afyouni, TE Nichols, Insight and inference for DVARS. *NeuroImage* **172**, 291–312 (2018).
50. N Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series: With Engineering Applications.* (The MIT Press), (1949).
51. DL Donoho, For most large underdetermined systems of linear equations the minimal ℓ1-norm solution is also the sparsest solution. *Commun. on Pure Appl. Math.* **59**, 797–829 (2006).
52. T Dozat, Incorporating Nesterov Momentum into Adam in *International Conference on Learning Representations.* (2016).

Supporting Information for

Dynamical models reveal anatomically reliable attractor landscapes embedded in resting state brain networks

Ruiqi Chen, Matthew Singh, Todd S. Braver, ShiNung Ching

Ruiqi Chen.
E-mail: chen.ruiqi@wustl.edu

This PDF file includes:

Supporting text
Figs. S1 to S12
SI References

Supporting Information Text

Validation of MINDy modeling

Goodness of fit. We tested the prediction accuracy of the fitted models in three cases: the training data, the data from the same participant but in the other session, and the data from a random participant in the same session. The R squared was around 0.4 for all three cases, with the training accuracy being the highest, the within-person transfer being the second and across-person transfer being the third (Figure S1). Therefore, our models indeed captured the individuality of the dynamics.

Surrogate data simulations. To make sure that the nontrivial dynamics observed in fitted models were not simply due to methodological bias, we investigated whether MINDy can correctly capture the trivial dynamics in closely matched surrogate datasets. We tried two different surrogate simulation schemes and fit MINDy on these simulations using the same hyperparameters as in the main text. The first method follows (1), generating a dataset from noise (without any dynamics) but preserves the covariance (thus FC) and mean power spectrum of the rfMRI data. In (1), it was found that the tvFC method generated indistinguishable results for real data and such surrogate data. However, our model correctly produced a monostable dynamic system without nontrivial fluctuations (Figure S2). The second method fit MINDy on the noisy simulations of a closely-matched linear system. This linear system, referred to as 'linear MINDy', replaces the nonlinear activation function of MINDy with its best linear approximation. The 'linear MINDy' model was fit to rfMRI data using the same hyperparameters and loss function as actual MINDy, thus capturing the statistics of the data and also maintaining the 'sparse plus low-rank' connectivity structure. In fact, we found that the connectivity and decay parameters of 'linear MINDy' is highly correlated with the actual MINDy model fitted on the same data. After fitting the linear model, we simulated the model with additive noise. The magnitude of the noise was set to the root of mean squared error during the fitting of the linear model. The noisy simulations thus represent a close match of the true dataset but generated from inherently linear (and monostable) dynamics. We then fit (nonlinear) MINDy models on these simulations, and the models correctly reproduced a monostable dynamic system with no nontrivial attractors (Figure S2).

Reliability and behavioral correlation for MINDy parameters

MINDy parameters were individualized and consistent across the population. We characterized the consistency of model parameters across all participants, as well as within each participant. We computed the correlation of each set of parameters (connectivity, curvature or decay) between each pair of models. For each participant, we quantified the similarity between their two models as well as the mean similarity between their models and all other models. Results indicate that the parameters were highly consistent across the population, and even more within each participant (Figure S3, top-left). Next, for each parameter (e.g., one entry in the connectivity matrix), we calculated its intraclass correlation coefficient (ICC), which is the correlation of its value between the two sessions across all participants. ICC characterized the reliability of the individual differences in each parameter. We observed a large set of connectivity parameters with high ICC (Figure S3, top-right; also note that the connectivity is sparse so a lot of entries have low ICC). The ICC for curvature and decay parameters were acceptable, around 0.5 (Figure S3, bottom).

MINDy parameters encode cognitive differences. Here, we conducted a canonical correlation analysis (CCA) between the connectivity matrices of fitted models and the phenotypic measures in the HCP. The connectivity matrices from the two sessions were averaged within each participant before entering the analysis. We used the scripts provided by (2) which extends (3) to the whole HCP dataset. The connectivity matrices and subject measures were first projected to their first 100 principal components to reduce dimensionality. Then, CCA was carried out to identify the directions to which the projection of the connectivity data and phenotypic data are maximally correlated across the population. We identified a unique pair of such directions (modes) with statistically significant correlation (1000 permutations, Figure S4, top right). Further, this connectivity mode and phenotypic mode explained a significant amount of variance in their data respectively (Figure S4, middle and bottom right). Post-hoc correlation between the phenotypic mode and all phenotypic measures revealed that this mode is more related to fluid intelligence and substance use (Figure S4, left). Interestingly, our results are very similar to the original finding in (3) even though we are using very different node types (parcels vs. ICA networks) and connectivity measures (effective vs. correlational).

More examples of model dynamics

Here we show more examples of the dynamics observed in the fitted models (Figure S5). The three except the top-left one were grouped into 'others' type in the main text because of their rare occurrence.

Toy model for infinite-period bifurcation

Here we show a toy model for infinite period bifurcation mentioned in the main text. The model dynamics are written in polar coordinates (but shown in Cartesian coordinates in Figure S6) as:

$$\begin{cases} \dot{r} = r(1 - r^2) \\ \dot{\theta} = \mu - |\sin\theta| \end{cases}$$

where μ is the bifurcation parameter. The infinite-period bifurcation happens when μ equals one. When $\mu > 1$, the system shows a stable limit cycle and an unstable equilibrium at the origin. As μ approaches 1, the speed distribution on the limit cycle becomes more and more extreme and a ghost attractor emerges at $(0, 1)$ on the limit cycle. When μ equals 1, the ghost attractor dissolves into a pair of equilibria, one stable and one unstable.

Clustering of attractor patterns

Selection of the number of clusters. We used K-means algorithm to cluster the attractors and select the number of clusters K using the cluster instability index (see Methods for more details). We found that regardless of whether we use only the stable equilibria, only the ghost attractors, or both, we got near perfect clustering stability only for K equals to two or four (Figure S7).

Two cluster solution. We show the two-cluster solution in Figure S8. In this case the two clusters are the reflections of each other. One of them showed strong activation for the DMN and FPN, while the other show strong activation for the visual and the dorsal/ventral attention networks.

Clustering results with only stable equilibria or only ghosts. In the main text we clustered all stable equilibria and ghost attractors together. Here we show the results using only stable equilibria (Figure S9) and only ghost attractors (Figure S10) respectively.

Parcel activation in attractors

Parcel activation follows network and cluster structure. We analyzed the activation of parcels across all attractors with a mixed-effect model: *activation* $\sim$ *1 + cluster + network + cluster:network + (cluster*$|$*network:parcel)*. We computed the hierarchical (type I) sum of squares explained by each term (Figure S11). The main effects of cluster and network were small while their interactions explained over 40% of total variation, indicating that (1) the same brain network showed very different activation in different clusters (or equivalently, that each cluster is associated with the strong activation of different sets of networks); and (2) parcel activation was mostly determined by the combination of attractor clusters and functional brain network structure. The random effect of parcels explained about 15% of variation, indicating that heterogeneity still exist across the parcels within each network. The error sum of squares was smaller than the total variation explained by the modeling, suggesting high consistency across participants and sessions.

Functional rather than spatial organization explains the data better. To show that the activation was driven by functional network segmentation rather than the spatial proximity between parcels, we calculated the Pearson correlation between the activation patterns of each pair of parcels across all attractors. The correlation coefficient was Fisher-transformed into a Zr statistic (Figure S12, top-left). We modeled this similarity matrix by the combination of the spatial proximity (negative cortical distance) between parcels and their functional network assignment. The distance between all cortical vertices along the surface were calculated using the surface geometry file from HCP and MATLAB's graph distance function, and then averaged within the two parcels under consideration. All inter-hemisphere entries were excluded since the distances were undefined. The functional network assignment similarity was set to one if two parcels belong to the same network as defined by the 17-network atlas in (4), and zero otherwise. We then predicted the activation similarity using negative cortical distance, functional network assignment and their interactions. A hierarchical sum of squares analysis showed that network organization explained 15% of total variation even after excluding the effect of spatial proximity (which explained must less variance, Figure S12, bottom-right, first column). Therefore, the attractors indeed reflected the organization of functional brain networks over and above the spatial configuration of cortical regions.

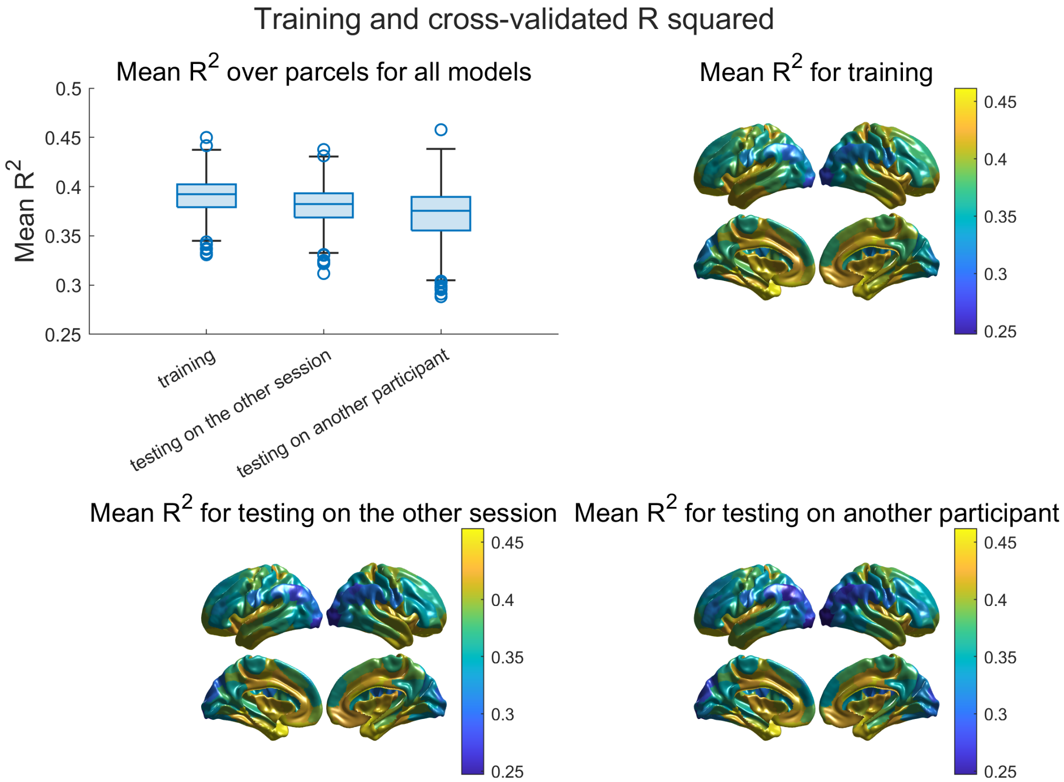


Fig. S1. Goodness-of-fit and cross validation accuracy. Top-left: distribution of the mean R squared for all models when predicting the training data, the data from the same participant in the other session, or the data from another randomly selected participant. Top-right: Distribution of mean R squared over all models when testing on training data. Bottom-left and bottom-right: similar plots for testing on the other session within participant, or testing on another participant's data.

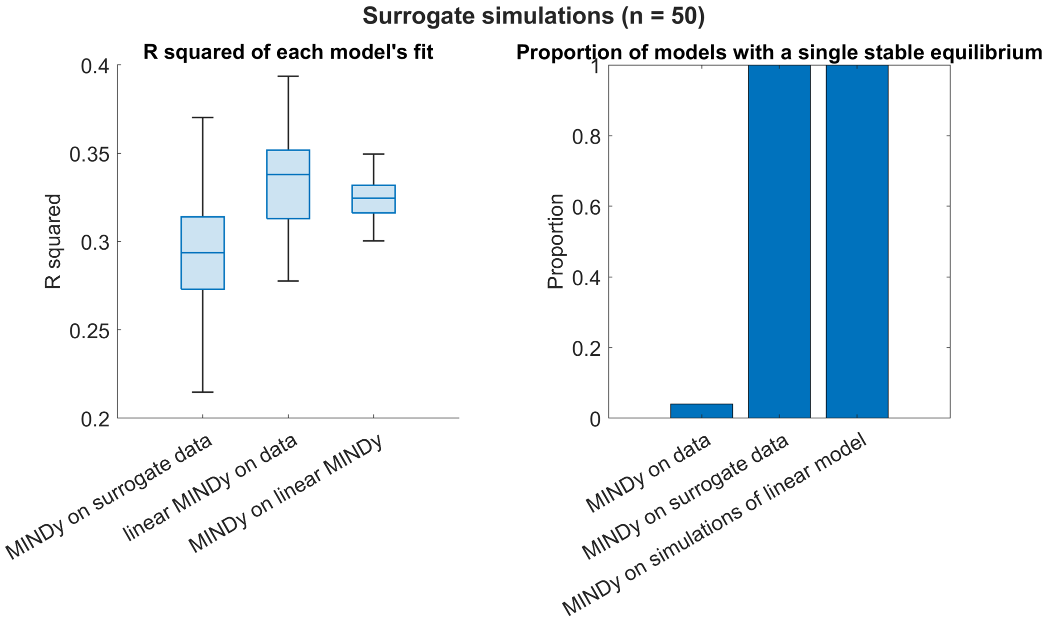


Fig. S2. Fitting MINDy on surrogate data revealed trivial dynamics. Left: R squared for fitting MINDy on surrogate data; fitting 'linear MINDy' on rfMRI data; and fitting MINDy on the noisy simulations of 'linear MINDy' models. Right: the proportion of models showing a single stable equilibrium (at the origin), for MINDy models fitted on rfMRI data, on surrogate data, and on the noisy simulations of 'linear MINDy' models.

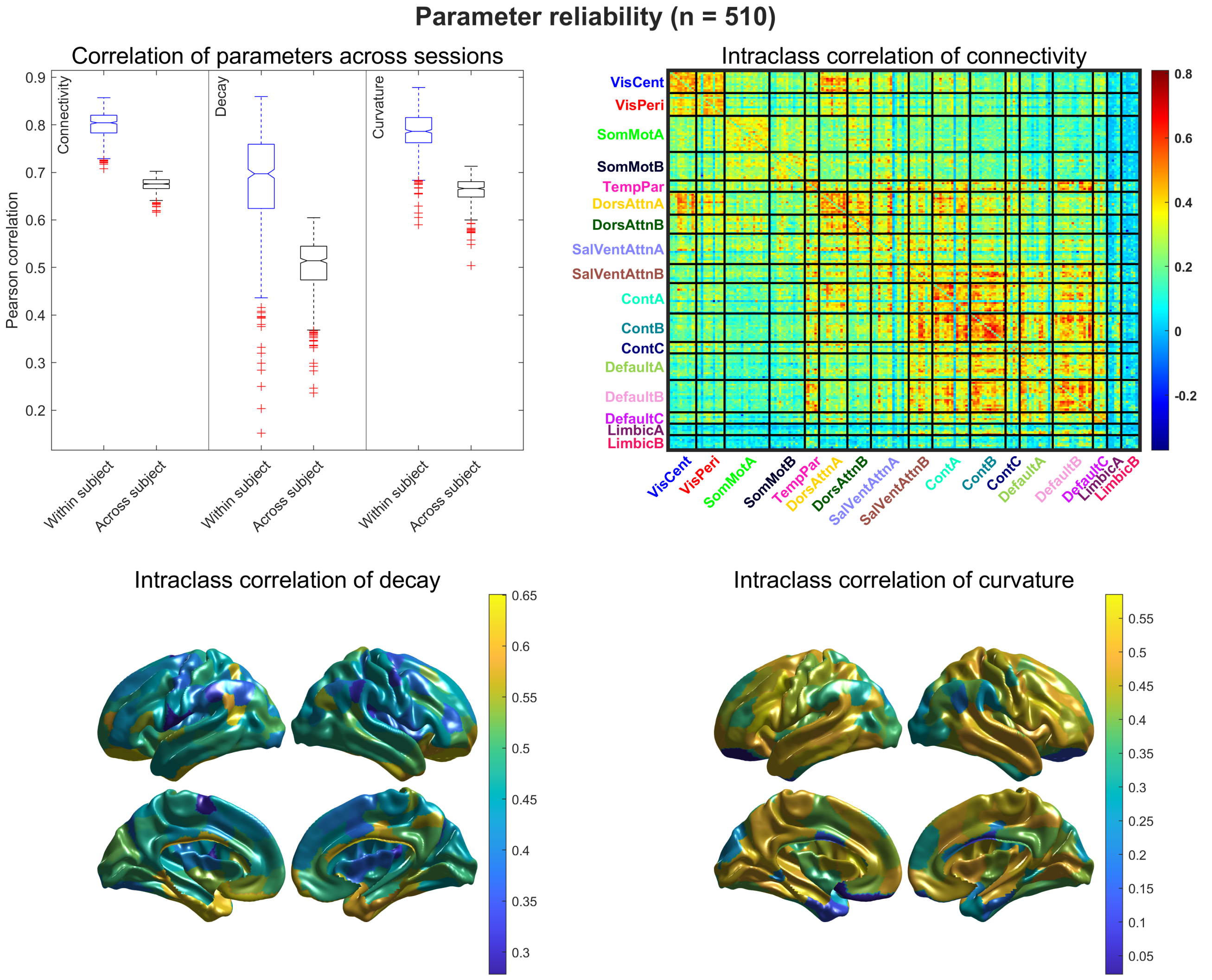


Fig. S3. Obtained parameters were individualized and consistent across the population. Top-left: Correlation between sets of parameters (connectivity, curvature or decay) from models from different sessions. ‘Within subject’ is the correlation between the two models from a same participant, while ‘across subject’ is the mean similarity with all other models. Top-right: intraclass correlation coefficient (ICC) of each connectivity parameter. Text labels indicate the functional networks, separated by the black thick lines. Bottom: ICC for decay and curvature parameters, respectively.

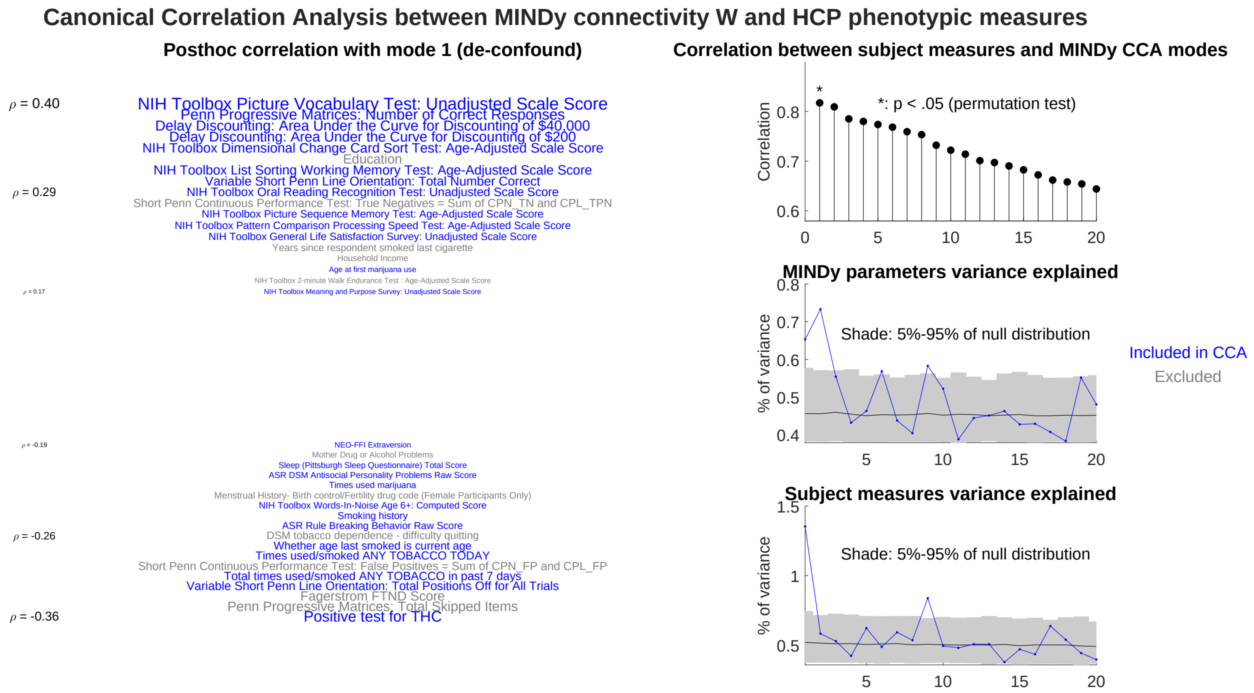


Fig. S4. MINDy parameters encoded reliable individual differences. Please compare with Figure 1 in (3). Left: Post-hoc correlation between the behavioral mode identified by CCA and the phenotypic measures. We listed the most correlated measures with font size scaled by the correlation. Note that the Y axis is ordinal but not scalar. Top-right: correlation between CCA-identified connectivity and behavioral modes. Statistical significance is determined by permutation test with 1000 permutations (same for other panels). Mid-right: Variance of the connectivity explained by the CCA modes. Shaded region indicates the null distribution with 1000 permutations. Lower-right: variance of the behavioral measures explained by the CCA modes.

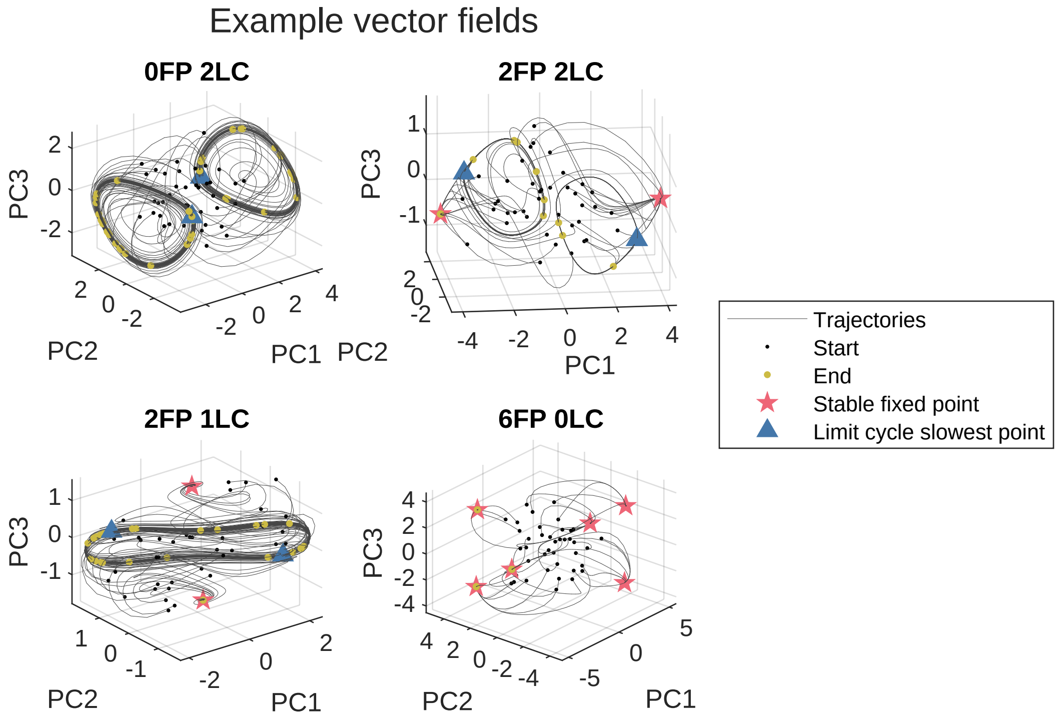


Fig. S5. More examples of dynamical landscapes. See Figure 1 in main text.

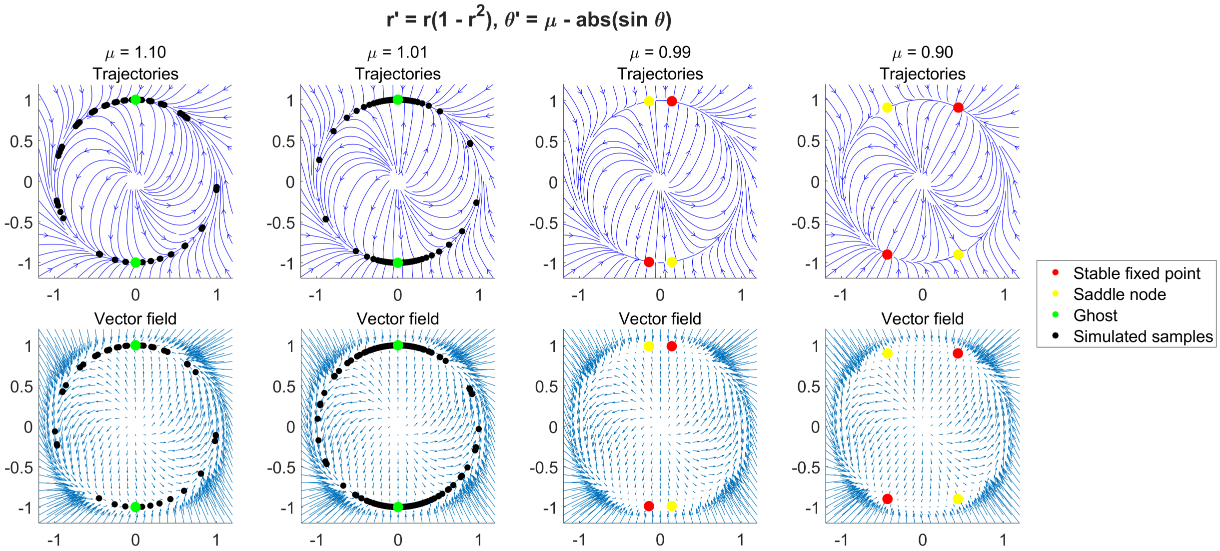


Fig. S6. Toy model for infinite period bifurcation. Top row: trajectories of the models with different value for the bifurcation parameter μ. Red, green and gray dots indicate stable equilibria, saddles and ghost attractors respectively. Black dots are the simulated samples on the limit cycle. Bottom row: vector fields of the models.

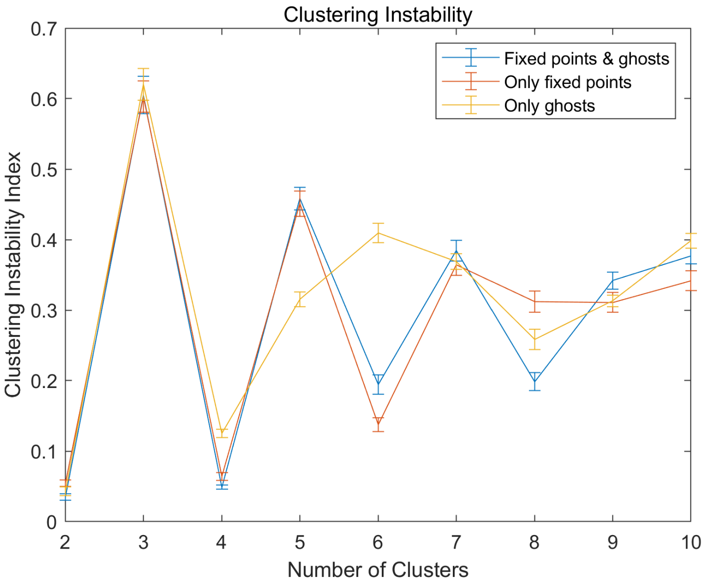


Fig. S7. Stable clustering solution for two and four clusters.

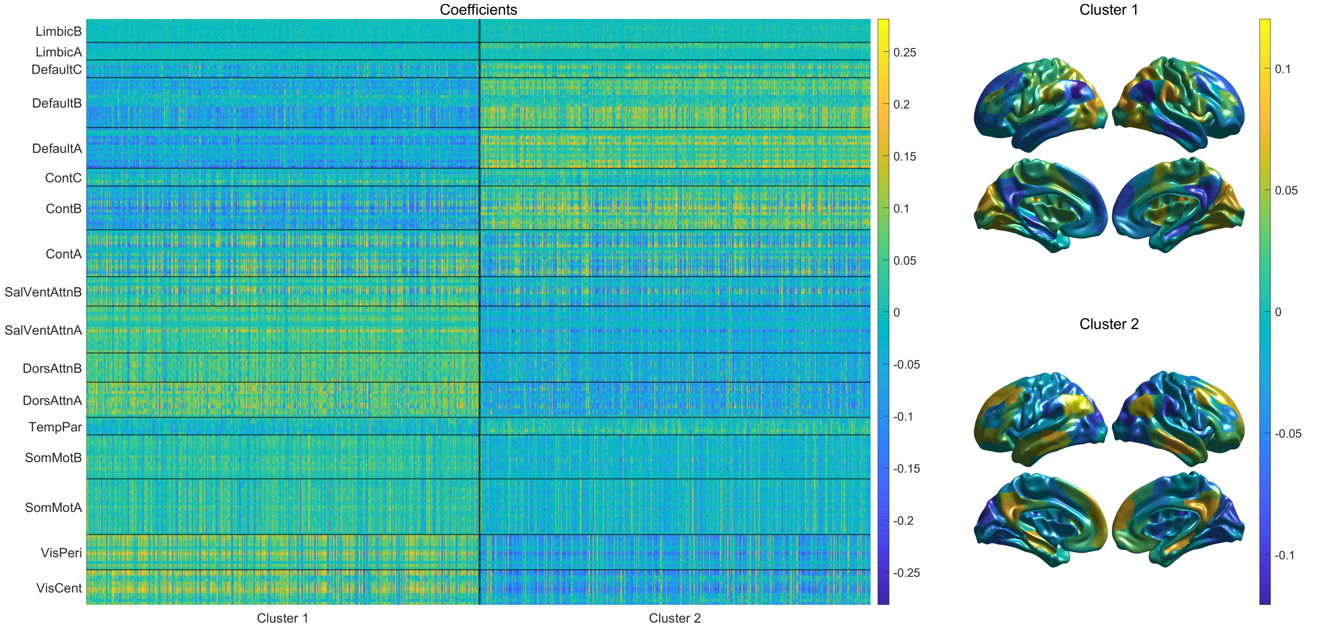


Fig. S8. Two-cluster solution of the attractor clustering.

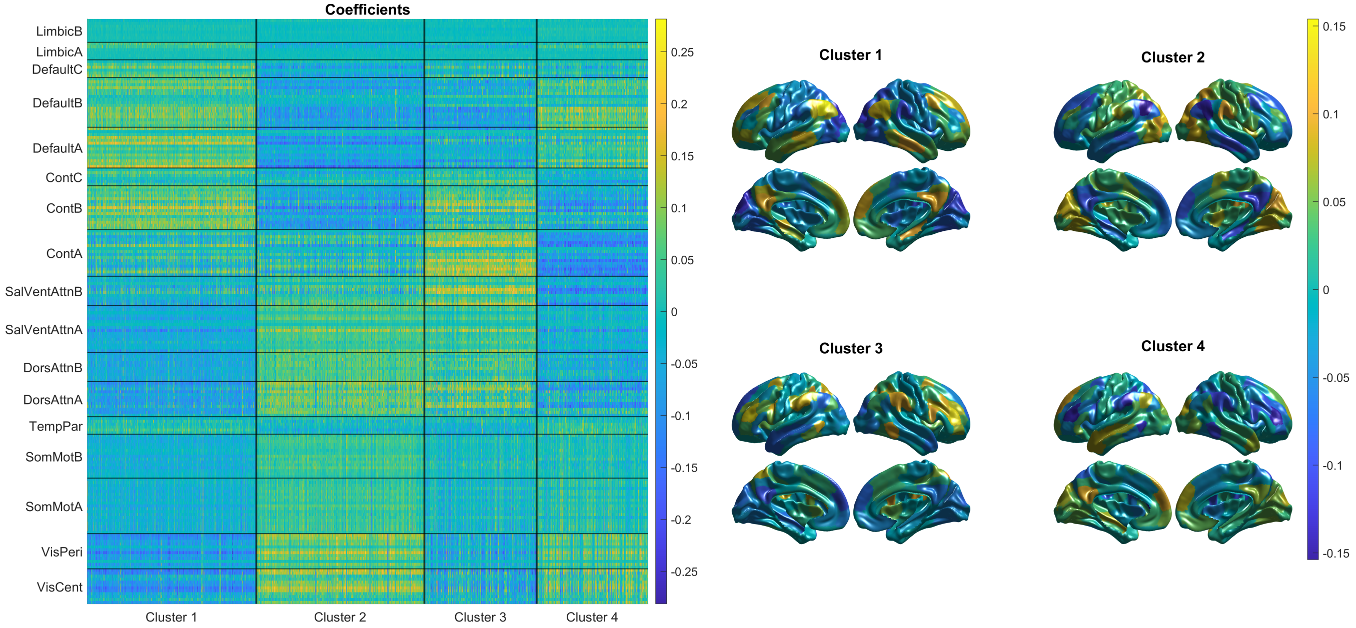


Fig. S9. Attractor clustering with only the stable equilibria.

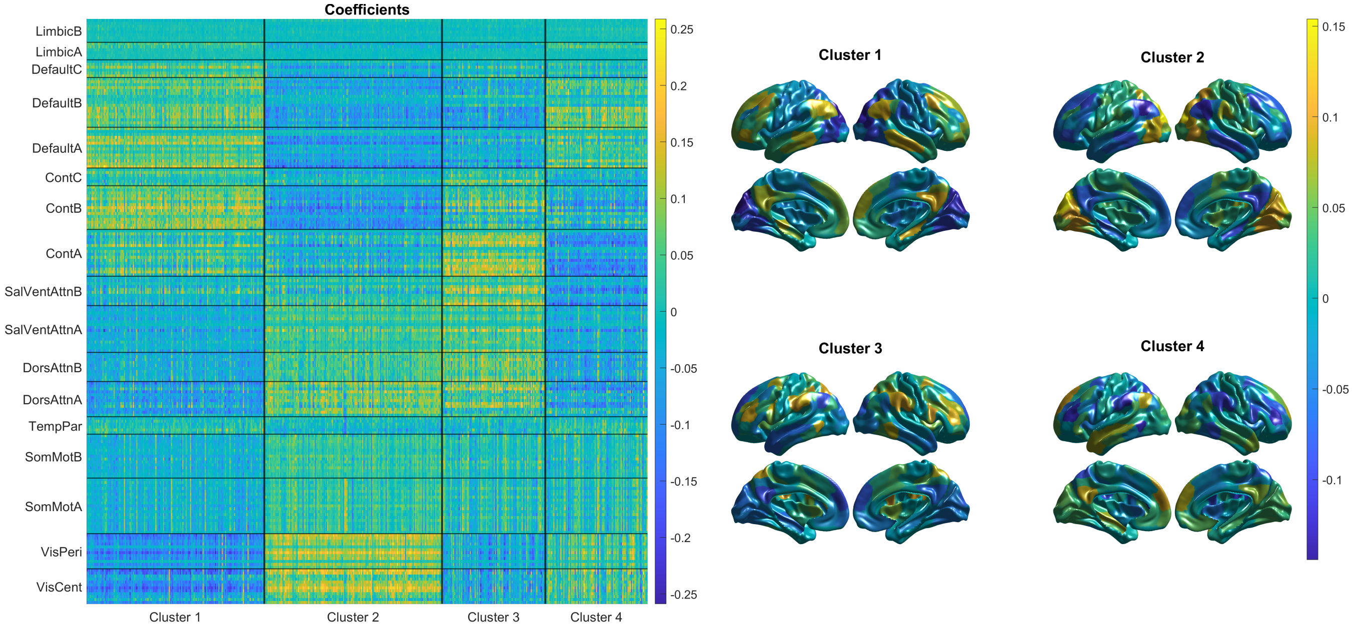


Fig. S10. Attractor clustering with only the limit cycle slowest points.

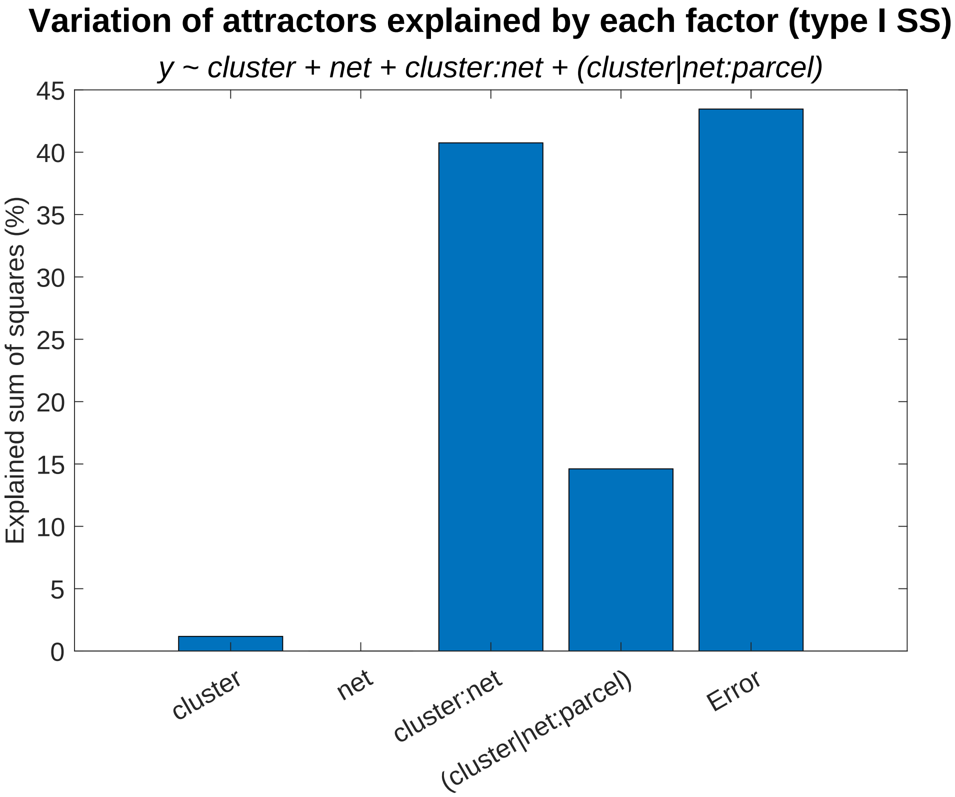


Fig. S11. Variance of parcel activation explained by attractor clusters and functional networks.

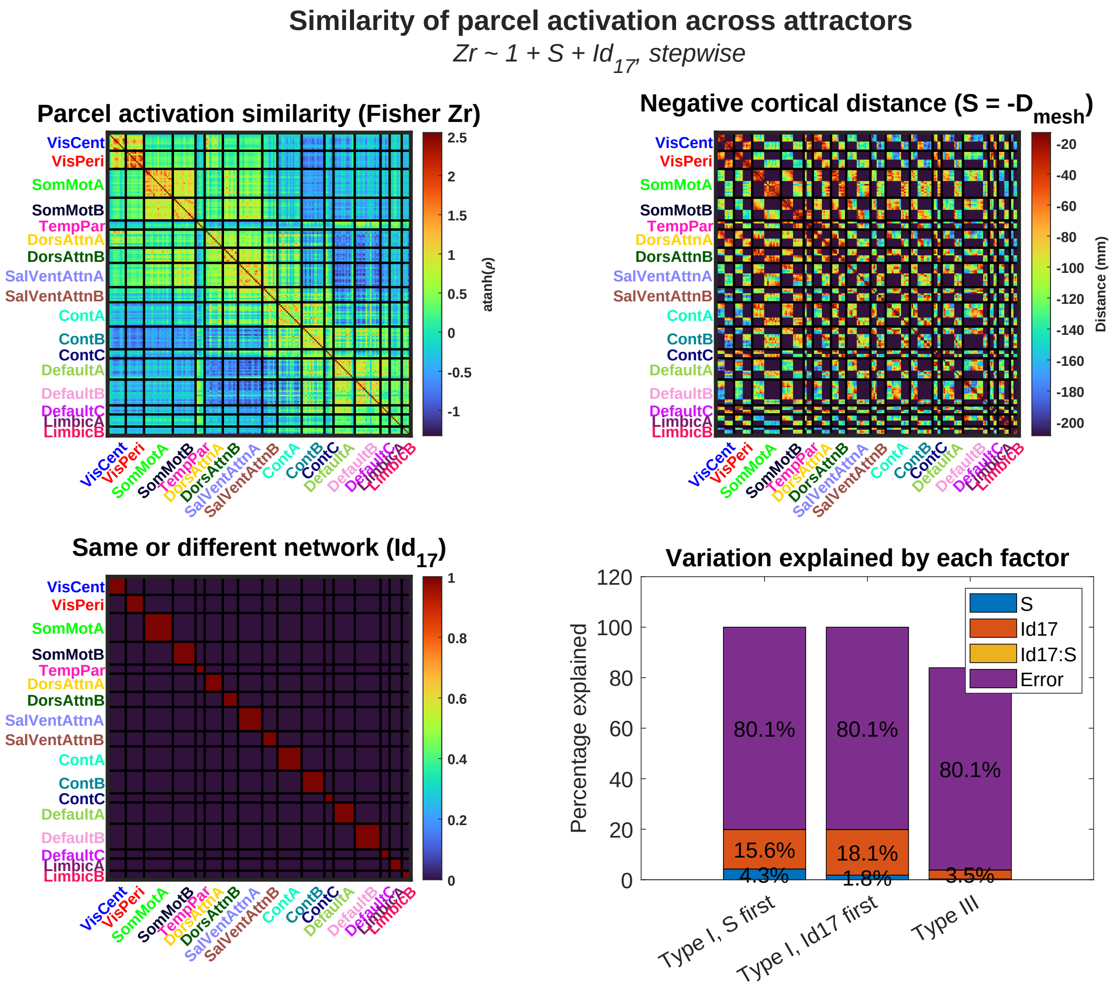


Fig. S12. Parcel activation similarity across all attractors. Top-left: similarity between all parcels' activation across all attractors. Similarity is quantified by Pearson correlation followed by Fisher's transformation. Top-right: the negative of the distances between parcel centroids along the cortical surface. Inter-hemisphere entries were omitted. Bottom-left: similarity between parcels based on (4) functional network segmentation. Similarity is one if the two parcels belong to the same network and zero otherwise. Bottom-right: variance of parcel activation similarity explained by distance or network structure. Column one: type I (hierarchical) sum of squares (SS) where cortical distance precedes network structure (see the text). Column two: type I SS where network structure precedes cortical distance. Column three: type III SS. The SS for the interaction between distance and network structure is less than 1%.

References

1. TO Laumann, et al., On the Stability of BOLD fMRI Correlations. *Cereb. Cortex* **27**, 4719–4732 (2017).
2. N Goyal, D Moraczewski, P Bandettini, ES Finn, A Thomas, Computationally replicating the Smith et al. (2015) positive-negative mode linking functional connectivity and subject measures (2020).
3. SM Smith, et al., A positive-negative mode of population covariation links brain connectivity, demographics and behavior. *Nat. Neurosci.* **18**, 1565–1567 (2015).
4. A Schaefer, et al., Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cereb. Cortex* **28**, 3095–3114 (2018).