Thalamic Influence Over Adaptive Cortical Dynamics Across Conscious States

Authors

Eli J. Müller[1,2*], Brandon R. Munn[1,2], Giulia Baracchini[5], Ben D. Fulcher[3], Vicente Medel[4], Michelle J. Redinbaugh[6,7,8], Yuri B. Saalmann[7,8], Bingni W. Brunton[9], Steven L. Brunton[10] and James M. Shine[1,2]

Affiliations

1 Brain and Mind Centre, The University of Sydney, Sydney, NSW, Australia

2 Centre for Complex Systems, The University of Sydney, Sydney, NSW, Australia

3 School of Physics, The University of Sydney, Sydney, NSW, Australia

4 Latin American Brain Health Institute, Universidad Adolfo Ibañez, Santiago, Chile

5 Montreal Neurological Institute, McGill University, Montreal, QC, Canada

6 Department of Biology, Stanford University, Stanford, USA

7 Department of Psychology, University of Wisconsin – Madison, Madison, WI, USA

8 Wisconsin National Primate Research Center, Madison, WI, USA

9 Department of Biology, University of Washington, Seattle, WA, USA

10 Department of Mechanical Engineering, University of Washington, Seattle, Washington, USA

Corresponding author:

* Eli J. Müller – eli.muller@sydney.edu.au

Summary

The human brain must support both stable and flexible neural dynamics in order to adapt to changing contexts that are inherently non-linear. The thalamus has been linked to the coordination of these opposing dynamical regimens in the cerebral cortex, however existing methodological approaches have not integrated sufficient neurobiological details with a sensitive measure of neural dynamics that permits sensitivity to time-series non-linearities. Inspired by the field of fluid dynamics, we use a novel approach to show that spontaneous fMRI data exhibits non-trivial fluctuations in predictability over time, akin to a river that has sections of smooth and predictable (laminar) versus rough and unpredictable (non-laminar) fluid flow. We use a combination of pharmacological fMRI, macaque electrophysiology and a large-scale biophysical model of the thalamocortical system to provide robust evidence that the thalamus provides versatile control over globally linear dynamics in the cerebral cortex that characterize conscious states.

Introduction

The human brain supports a wide-ranging behavioural repertoire, with dynamic, context-dependant recruitment of specialized neuronal assemblies emerging from coordinated interactions amongst tens of billions of cells. The computational capacities that support these behaviours are in large part related to this cellular complexity, however, this high-dimensional neural activity is of minimal adaptive benefit unless the neural activity that unfolds over time is both stable and yet flexible. In other words, neural dynamics must be reliable for repeated precise execution, but also rapidly adaptable to changing contingencies. How the intricate organization of the brain supports these opposing capacities remains to be understood.

The thalamus is a highly-conserved subcortical structure that has been argued to play a central role in shaping and constraining the flexible neural dynamics that characterise the conscious, awake brain[1–3]. At the microscale, there is evidence that specific cell types in the thalamus act to relay information from sensory systems[4,5], coordinate motor actions[6], and also to integrate and distribute information from subcortical structures[7–9] to decentralized corticothalamic networks. At the macroscale, recent work has shown that diffusely-projecting subtypes of thalamocortical nuclei promote variability of dynamics across the cerebral cortex[7,8,10] through the instantiation of quasi-critical brain state regimes[11] which support shifts in cognition[12] and consciousness[13–16]. These emerging perspectives place the thalamus in a unique topological position within the brain, wherein it is able to both drive the cerebral cortex with targeted inputs, but also to modulate on-going cortical dynamics through diffuse projections[7]. Despite this privileged position, to date there are few direct empirical links between the thalamocortical system and precise dynamic trade-offs characteristic of adaptive systems[1].

A primary reason for this lack of direct evidence is that few existing neuroimaging analysis techniques are capable of providing a time-resolved estimate of the stability (or flexibility) of brain state dynamics that is sensitive to the inherently non-linear waxing and waning nature of our waking lives[17]. Most of the current statistical approaches have either been designed to capture distributional moments of neural time-series or have heavily relied on assumptions of stationarity[18,19–25]. While some approaches have softened assumptions of stationarity – e.g., by carving data into shorter windows that are swept over time[26,27] – these techniques are typically model-based, which makes the detection of non-linear shifts in brain states inherently difficult. Other approaches have used inventive means for tracking functional neuroimaging data at a fine temporal resolution[28–30], however the direct interpretation of these measures with respect to stability and flexibility (which naturally require longer epochs of data to quantify) is non-trivial.

Fortunately, there are existing methods from the field of fluid dynamics that are well-suited to this class of problems, however they require a subtle change in perspective. By way of analogy, if we consider brain dynamics as a flowing river, we need to move from tracking the topography of the river – which is admittedly still highly informative – to characterising the flow of water at different sections of the river: one section might be calm and largely laminar, whereas another might be full of turbulent swirling eddies and so non-laminar (Fig. 1a). Importantly, these local dynamics provide fundamentally different constraints over the local environment, permitting smooth (stable) or disturbed (flexible) flow of local objects, respectively. This dynamics-focussed perspective has led to key insights in fields such as fluid mechanics, mechanical engineering, economics, and sociology[31–33], and here we advocate for its application to problems in neuroscience.

There are several significant historical parallels between the fields of neuroscience and fluid dynamics in the development of measurement techniques that further

motivate advances in data-driven modelling approaches. Both disciplines study complex systems that are high-dimensional, non-linear, and inherently multi-scale, both in space and time, yet both also exhibit dominant spatiotemporal patterns of activity that are meaningful for scientific understanding and engineering[31,34,35]. Importantly, experimental measurements in each field have followed a similar technological trajectory: recordings were first single point (hot-wire anemometry and sharp electrodes), followed by simple arrays (many hot-wires and electrode arrays), and more recently spatial and temporal fields resolved by imaging (laser-based particle image velocimetry PIV[36–38] and magnetic-resonance-based functional brain imaging[39]). Accordingly, mathematical modelling techniques in fluids have evolved from quantifying statistics to extracting spatiotemporal dynamics[40,41]. We note that—with a delay of several decades—a similar progression is now occurring in neuroscience[42]. For this reason, there is a tremendous opportunity to leverage modern dynamical systems modelling techniques to gain insights into neural computation.

In this paper, we repurpose an approach from fluid dynamics for application to functional neuroimaging data. Specifically, we develop a measure of timeseries stability – akin to fluid laminarity – by tracking the capacity of a simple linear dynamic model to accurately predict successive time-points. We show that neural recordings transition through periods of increased linear predictability that irregularly fluctuate through periods of decreased linearity over time (Fig. 1b). The spatial and temporal structure of these linear dynamics reveals coherence maps and intrinsic timescales that are not captured by null models that either scramble temporal order or recreate timeseries based on the covariance and Fourier power spectrum of the original data. We then reveal how pharmacological manipulation of arousal, via propofol-induced anesthesia, results in heightened linearity in human neuroimaging data (Fig. 1c). Next, we demonstrate how targeted and diffuse thalamocortical subtypes can both promote or diminish linearity, respectively. We

underscore the importance of thalamocortical interactions in a detailed biophysical model which further predicts matrix thalamic stimulation re-instantiates deviations from linear dynamics concomitant with conscious arousal, which we then confirm by applying our approach to multielectrode cortical recordings from an anesthetized macaque that was awoken by electrical stimulation of the diffuse thalamic projections within the central lateral thalamus[14,15]. In this way, we reveal the crucial role of the thalamus in shaping and constraining global linear activity patterns distributed across the cerebral cortex, supporting an essential feature of stable, and yet adaptive and flexible cognition.

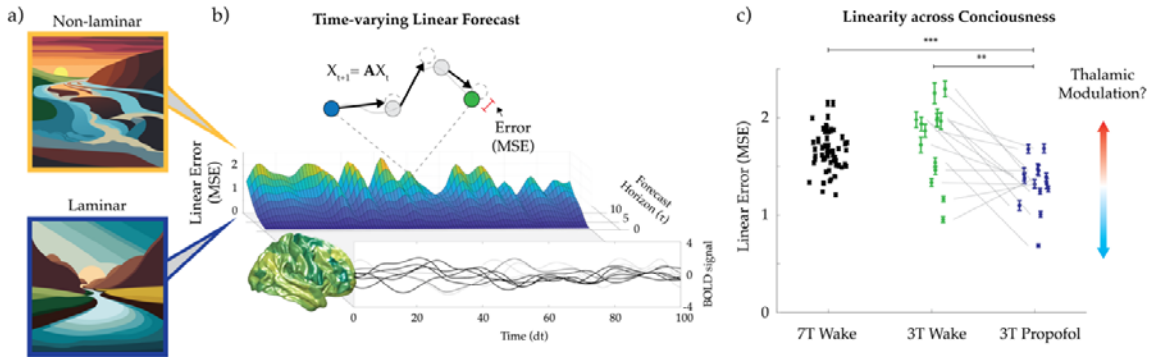


Figure 1 – Linear dynamics in large-scale brain recordings. a) Depiction of laminar and non-laminar river flows forming an analogous axis for large-scale brain dynamics – this analogy provides a useful means for appreciating the approach used in our experiment. b) Top: Schematic showing how a global linear model can be compared to ground-truth timeseries as a time-varying measure of linearity – here, a low-dimensional representation of changes in the BOLD signal across the cerebral cortex is used to predict changes to the state of the brain in the next time point (black lines to dotted circles); the actual brain state (closed circles) is recorded, and the distance between these two states in state space is quantified as the "Error" of the model – linear dynamics are associated with consistently low Error values, whereas non-linear dynamics have variable Error values. Bottom: A Horizon plot showing the time-varying linear forecast performance in 7T resting-state human BOLD[43] from an example subject – note the large fluctuations in Linear Model Error (MSE), particularly at longer delays (Forecast Horizon,τ). c) 7T human resting-state linear prediction performance (N=59; black), 3T human resting-state linear prediction performance (N=14; green), 3T human resting-state under propofol anesthesia linear prediction performance (N=14; dark blue).

Results

Linear dynamics in large-scale cortical recordings

To quantify the time-varying stability of brain state dynamics, we applied a simple linear model estimation technique (Fig. 1b; Supp. Fig. 1) to a high spatiotemporal resolution human 7T fMRI dataset captured during the resting state. First, a global linear propagator tracking moment-to-moment shifts in a recording is generated for each subject's timeseries by utilizing the singular-value-decomposition (See Supp. Fig. 1a). Then, for every timepoint in the recording, the linear propagator is used to forecast state dynamics at a set of future timepoints (τ; 1-10 TR) – note that we explicitly chose to fit a global propagator (i.e., we did not redefine the propagator in a sliding-window fashion). The linear dynamics predicted by the global propagator are then compared to the ground-truth dynamics via the mean-squared-error, providing a time-resolved read-out of how effective (or not) the linear model was at predicting the upcoming brain state. We denote this term as the Forecast Horizon of the linear model prediction (Fig. 1b) and the resultant graph as a Horizon plot.

We reasoned that, if the resting brain were stable (i.e., linear) throughout the recording, then we should observe relatively strong, consistent predictability throughout the entire recording session. In stark contrast, the Horizon plot in Fig. 1b shows that neural dynamics cycle through periods of increased and decreased linearity over the course of a recording session. This observation is at direct odds with the typical assumption of stationarity that is applied in standard neuroimaging analyses[21]. The results are observed for every individual in the dataset (N = 59; Supp. Fig. 2), robust across validation datasets (which had different temporal resolutions; see Supp. Fig 2) and do not correlate with frame-wise displacement (r = 0.01 +- 0.03 across 10τ). Crucially, these results extend previous findings of strong linear

dynamics in human neuroimaging[44–46] and large-scale electrophysiological recordings[46,47], by showing substantial yet irregular deviations from linear predictability in conscious states that is pervasive across subjects.

Adaptive timescales and dynamic brain modes

Intrinsic timescales of the brain have been of interest since the discovery of oscillatory dynamics in the first human EEG recordings[48]. Human neuroimaging studies have revealed a hierarchy of temporal and spatial autocorrelation scales across the cortex[49] and that these capture a large number of existing topological metrics[22]. Linear flow analysis enriches this approach by quantifying intrinsic modal timescales that drive linear dynamics in neural activity – and which may be recruited in an adaptive and time-varying manner reflecting changing cognitive demands. To do this, we leveraged another method commonly used in the field of fluid dynamics – *Dynamic Mode Decomposition* (DMD) – that is designed to efficiently characterize the dynamics of time-varying processes through eigendecomposition of the linear propagator matrix[31,34,35,50], reformulating the linear dynamics as a function of spatiotemporal modes [32,35,50]. Each mode contains four key characteristic properties: the spatial eigenvector defines a spatial pattern of coherent activity, and a corresponding spatial pattern of delays relative to each of these coherence patterns – called a dephasing map. In addition, each mode has a corresponding eigenvalue defining its oscillatory frequency, and an exponential gain parameter – which determines whether a mode will grow or decay in time (i.e., captures its temporal stability) – representing two key dimensions of the system's timescales.

By applying DMD to each subject's fMRI recording, we formulate a subject-specific set of spatiotemporal modes capturing on-going linear dynamics. In order to gain insight into how these dynamic modes relate across all subjects, we leverage a two-pronged clustering approach to aggregate modes across timescales, and then spatial coherence, respectively. First, we consider timescales by utilizing *k*-means clustering

of the eigenspectrum collated across all subject recordings. A peak in adjusted mutual information between clustering repetitions was then used to select k = 5 clusters (1000 repetitions; see Supp. Fig. 4), resulting in 5 distinct timescale groupings – the corresponding average of the coherence maps within these clusters (Supp. Fig. 4). Timescale cluster 1 shows somatomotor and visual activation antiphase with frontal, parietal and temporal cortices across a broad range of frequencies – spanning much of the time resolution of BOLD recordings. These modes are also weakly damped – i.e., have a strong gain relative to the other temporal clusters. Timescale cluster 2 shows visual, premotor, and dorsal lateral prefrontal antiphase with other cortices at slower frequencies (< 0.03 Hz; Fig. 4 g-h middle) and with a stronger damping rate, i.e., weaker gain, than cluster 1. And timescale cluster 3 shows strongly damped temporal cortex activity at frequencies < 0.02 Hz.

Next, we considered the spatial patterns of coherence, defined by each mode, aggregated across subjects and recording sessions to define consistent modal groupings[50]. Again, we leveraged k-means clustering (adjusted mutual information peaked between clustering repetitions for k = 5; See Supp. Fig. 4): Figure 2m shows 3 cluster centroids and their defining coherence patterns – i.e., the spatial patterns of the linear modes across all subjects fall within distinct classes. These classes show a significant overlap with resting-state networks used broadly in the literature[51] – including the default mode, and somatomotor and visual networks. We further find that the damping rates and frequencies of the modes within each class have comparable probability distributions – demonstrating these spatial classes did not demonstrate unique linear timescales at this granularity in resting-state.

Empirical fluctuations are distinct from surrogate models

To further investigate the fluctuations in linear dynamics observed in resting state fMRI data, we contrast these against two surrogate timeseries that conserve distinct

features of the empirical fMRI data. The first, referred to as the *spectral surrogate*, consists of white-noise timeseries imbued with a covariance and Fourier frequency spectrum matching those of the empirical data (i.e., a stationary surrogate[21]; see Supp. Fig. 1d). Crucially, this surrogate shares the empirical correlative structures between brain regions which forms the basis of many approaches in the neuroimaging literature[18]. In addition, this surrogate includes timescale information captured by the Fourier spectrum, which is predictive of future dynamics, since the sinusoidal modes comprising this time averaged measure oscillate across time. The second surrogate, referred to as the *shuffled surrogate*, is simply the initial empirical fMRI data uniformly scrambled in time – i.e., the regional activations relative to other regions at a given moment are conserved, matching empirical regional correlations, but *when* they occur with respect to any other activations in time is scrambled. Since the empirical data and both surrogates have the same correlation structure, they also share the same principal components (i.e., covariance defining regional activation patterns). Despite these similarities, Fig. 2b & 2c show that the temporal expression of these patterns no longer matches that seen in the empirical data (Fig. 2a).

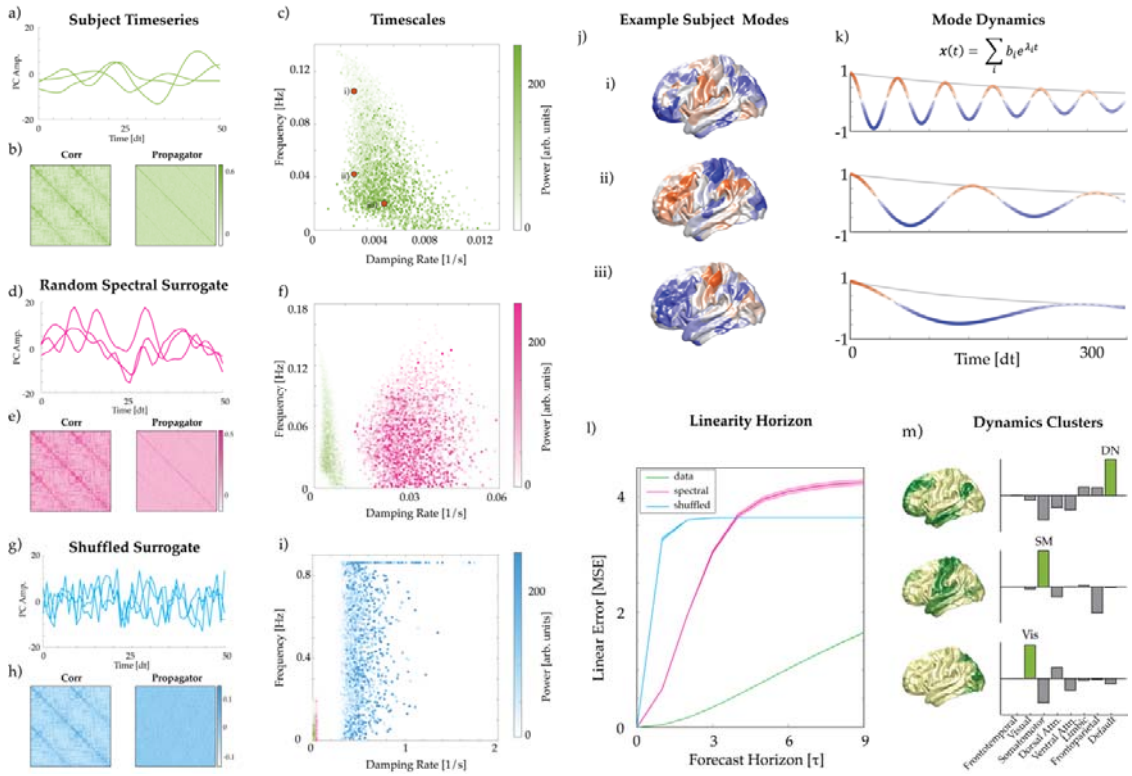


Figure 2 – Spatiotemporal modes of global linearity. a) Example subject time-series of first 3 principal components. b) Example subject correlation matrix (left) and linear propagator matrix (right) capturing linear dynamics. c) Timescale properties of dynamic modes defined by the linear propagator of 7T resting-state data (N=59). Colour bar shows the total power within each mode across the entirety of the recording. Orange dots correspond to the example subject's dynamic modes in j). Random Spectral Surrogate is generated from a white-noise timeseries with a covariance and Fourier power spectrum matching the empirical data in a-b). d) Principal component time-series for the random spectral surrogate. e-left) Random spectral surrogate correlation matrix, (e-right) linear propagator matrix, and (f) dynamic mode timescales. Shuffled Surrogate is generated by uniformly randomizing each timepoint of the regional timeseries – i.e., only randomizing time and not space. g) Principal component time-series for the shuffled surrogate. h) Left: shuffled surrogate correlation matrix; Right: linear propagator matrix, and (i) dynamics mode timescales. j) Example subject dynamic modes capturing the most (j-i), median (j-ii), and least (j-iii) amount of globally-linear dynamics and (k) these modes' corresponding temporal dynamics. l) Average linear errors as a function of future time points across all (N=59) subjects (green), and all corresponding Random Spectral (pink), and Shuffled (blue) surrogates ($p < 0.001$; line = mean across subjects; shaded error bar = variance across subjects). m) k-means spatial coherence clusters of the dynamic modes across all subjects (N=59) - showing 3 clusters (m-left) of k=5 – see Supp 3 for all clusters. (m-right) Cluster power within each of the 7 resting-state networks[51].

Again, leveraging *Dynamic Mode Decomposition*, we find the empirical fMRI data are clearly distinct from the surrogates in two key ways: firstly, the fMRI data consist of slowly fluctuating, sustained, spatiotemporal dynamic modes (Fig. 2c), as compared to the surrogates. Due to its matched oscillatory spectrum, the spectral surrogate contains a similar distribution of modal frequencies (0-0.14Hz) to the empirical data (Fig. 2f). However, these oscillations are strongly damped, as the spectral surrogate mostly excludes moment-to-moment information predictive of future states with only the time-averaged Fourier modes contributing to continuity of its temporal trajectories. In contrast, the shuffled surrogate displays a much wider range of high frequency (0-0.8Hz) strongly damped modes than both the empirical data and the spectral surrogate (Fig. 2i). This is due to the shuffled surrogate's time-point scrambling favouring fast and short-lived shifts in moment-to-moment activity.

Secondly, the empirical fMRI data contains significantly more linear dynamics than either surrogate (Fig. 2l). That is to say, the empirical fMRI data contain information in their moment-to-moment dynamics that is predictive of future globally-linear trajectories. Since both surrogates share the same covariance as the empirical data, and the spectral surrogate additionally includes Fourier spectral structure, we find these time-averaged features are not sufficient to capture the observed linear dynamics seen in the neural recordings. This finding was consistent at the subject level for all subjects (N=59; Supp. Fig. 2) and cross-validated in a 3T resting-state imaging dataset (N=14; Supp. Fig. 2[52]). To ensure that this effect was not caused by fixed patterns enduring within the timeseries, we calculated the time-averaged mean-squared-displacement (MSD) for successive lags – comparable to temporal autocorrelation of the entire brain state – and found the linear errors are significantly smaller than those expected from general variability within the signal (Supp. Fig 3). In short, we have presented a simple technique for clearly demonstrating that, despite substantial fluctuations between linear and non-linear flow, resting state

fMRI data remains distinctly *more* globally-linear than either the spectral or shuffled surrogate.

How does this perspective of linear dynamics extend traditional investigations into the resting brain? Further insight can be gathered by exploiting the dynamic modes corresponding to the propagator used to predict the linear (or non-linear) temporal epochs (Fig. 2l). Figure 2j shows an example of three dynamic modes from a single subject's recording: the dynamic mode with the largest average amplitude across the time series (Fig.2j-i), median (Fig.2j-ii), and the smallest (Fig. 2j-iii). The spatial coherence maps for these modes are shown in Fig. 2j and have corresponding temporal dynamics defined by a complex exponential (Fig. 2k). These modes can be collated across all subjects via clustering approaches (See Methods) and show spatial coherence groupings that recapitulate common resting-state networks[51] (Fig. 2m; Supp. Fig. 4), consistent with previous findings[53], and temporal groupings defining cross-regional dynamics at distinct frequencies and temporal stabilities – i.e., damping rates (Supp. Fig. 4).

Together these results clearly demonstrate that fMRI data contains predictable linear dynamics that are not observed in stationary surrogates – that is, the dynamics of human fMRI contains information in its moment-to-moment fluctuations that is predictive of its future dynamics. Despite this predictability, resting-state neuroimaging data also irregularly cycles through periods of predictable and unpredictable flow (Fig. 1b).

Thalamic axis of cortical linearity

Armed with this simple yet effective approach, we were now able to test our original hypothesis: namely, that axes of anatomical variation in the thalamus facilitate shifts between stable and flexible dynamical modes in the brain. The thalamus itself is comprised of a rich diversity of cell types which have unique projection profiles to

both the cerebral cortex and other subcortical structures[54–56]. Although there is no single schema that effectively captures the diversity of thalamocortical interactions[9,57–60], recent empirical studies in rodents have demonstrated the utility of considering a non-binary continuous spectrum of thalamic cells that differ in terms of the topography and targets of their axonal arbours[61]. At one end of this continuum sit the calbindin-rich matrix nuclei, which have thalamocortical projections spanning multiple cortical areas, typically targeting superficial layers of agranular association cortices (Fig. 3a). This organization contrasts with the parvalbumin-rich core nuclei, which typically target granular layers of the cerebral cortex, providing feed-forward input to cortical pyramidal neurons[8,9] (Fig. 3a-b).

Although there is evidence of individual thalamic cells that contain axonal features of both core and matrix types (even in the same axon[54]), there are good reasons to consider a spectrum that extends between the two extreme phenotypes[3,8]. For one, simplifying assumptions at the microscale are required in order to conduct tractable work at the macroscale of neuroimaging[1,8,11,13]. Secondly, there is now robust evidence that core and matrix cell types also differ in their functionality[3]: recent empirical and theoretical work has shown that electrical stimulation of the central lateral thalamus – which itself is abundant with matrix nuclei – but not the ventral lateral thalamus (which is a predominantly core-rich nucleus) can drive recovery of consciousness in propofol-anesthetized macaque monkeys by leveraging the non-specific diffuse projection profiles of these matrix cells[11,13–16]. From these empirical vantage points, we thus ask: what type of cortical dynamics are facilitated by the thalamus, and how do these dynamics relate to the expression of core and matrix thalamic cell subtypes (Fig. 3)?

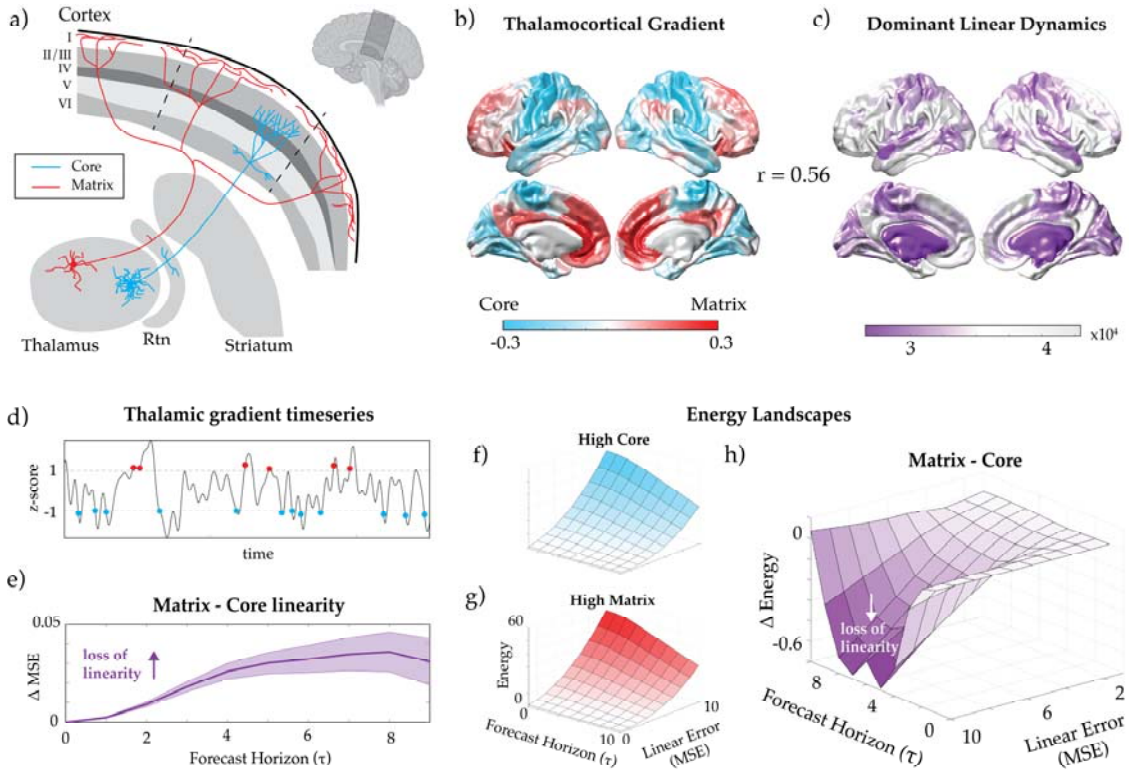


Figure 3 – Thalamic axis of cortical linearity. a) Schematic of characteristic projection profiles of core and matrix thalamocortical projection profiles (adapted from [54]). b) Core-Matrix thalamocortical expression gradient (adapted from [8]). c) Population average cortical map showing regions of dominant linear dynamics 10 time-points into the future. d) Thalamic timeseries for each subject are projected along the core-matrix gradient and epochs showing one standard deviation above the mean (high matrix) and below the mean (high core) are collated. e) The difference of the mean-squared-error of the linear model following high matrix compared to high core thalamic activity. These mean-squared errors can be reformulated into corresponding energies[62] (logarithm of inverse state probability) following high core (f) and high matrix (g) thalamic activity. h) The subtraction of the core energy landscape from the matrix energy landscape. Negative values indicate strong linearity following core activity as compared to matrix activity.

To answer this question, we applied a multitiered approach. First, we utilized the difference between the ground-truth data and the predicted linear dynamics – calculated for each region, each timepoint and each subject. We then sum these errors across regions to create a forecast horizon $\tau = 10$ (ensuring sufficient time for autocorrelation to decay) across all time points and all subjects. This defines a

cortical map revealing a spatial pattern of dominant globally-linear dynamics (Fig. 3c). Figure 3c shows that predictable linear flow was predominant in sensory cortices, and that these dynamics were strongly negatively correlated with the core-matrix expression gradient[8] ($r = -0.56$; $p < 10^{-10}$; Supp. Fig. 6a). That is, cortical brain regions with strong core thalamocortical projections also showed strong linear dynamics. However, both of these spatial maps also covary with the first principal component of the imaging data, which itself captures a significant proportion of the empirical timeseries' overall covariance, thus making strong inferences about the thalamic involvement difficult.

To test our hypothesis more precisely, we next asked whether the thalamus was temporally related to fluctuations in the predictability of cortical dynamics. Specifically, a time-resolved measure of matrix thalamus (relative to core thalamus) was generated by projecting each subject's thalamic BOLD timeseries onto a core-matrix expression gradient approximated from Allen Human Brain genetic expression data that corresponds to both extreme thalamic cell phenotypes (see methods; Fig. 3d)[8]. In this data, a positive weighting corresponds to high matrix activity relative to core regions of the thalamus (and *vice versa*). Next, for each subject, we identified peaks in this matrix-core timeseries that were one standard deviation above or below the mean (for high matrix and high core activity, respectively). Temporally averaging the prediction errors across all subjects and subtracting high core epochs from high matrix epochs clearly shows that increased linearity of cortical dynamics following peaks in core thalamic activity (Fig. 2e) – i.e., core thalamic activity precedes stable linear flow in the cortex.

To investigate these temporal fluctuations further, we leveraged a recent technique motivated by approaches in statistical thermodynamics that quantifies the likelihood of moment-to-moment state changes in a given timeseries by representing fluctuations as thermodynamic energy[62]. The likelihood for changes in linear flow

(MSE) in the data is generated at each proceeding timepoint (forecast horizon, τ) and produces a corresponding energy landscape for the two thalamic conditions (Fig. 2f-g; see Methods for details). This approach strengthens our previous findings, in that strong core thalamic activity (relative to matrix thalamic activity) is followed by increased predictable dynamics (negative drop seen in Fig 2h) in the cerebral cortex (with a maximum ~3 seconds post-peak).

These results are in keeping with the targeted feed-forward projections from the core thalamus acting as a feed-forward propagator cortical activity, and the more diffusely-projecting axons from matrix thalamic nuclei acting more akin to a modulator of ongoing cortical activity[14–16,63]. Furthermore, previous work[8] has shown that this anatomical axis also covaries with temporal variability of corticocortical connectivity, as captured by dynamic correlations in fMRI[8]. That is, cortical regions receiving more projections from the matrix thalamus showed a greater flexibility in their coordination with other cortical regions. Together, these findings demonstrate that the thalamus can provide both the stabilization of on-going cortical dynamics necessary for robust behaviours, but crucially, also mediate the flexibility of these cortical dynamics, which is essential for adaptive cognition. However, to strengthen the evidence for the thalamic control over cortical linearity, we require approaches that causally influence brain states.

Anesthesia promotes linear flow in the cerebral cortex

We first sought to determine whether alteration of conscious brain states impacted the linear predictability of cortical dynamics. Given the highly susceptible, flexible and context-dependent nature of consciousness[64–66], the critical importance of the thalamus in controlling dynamic brain states[1,14,67] (Fig. 3), and armed with a technique for quantifying this impact (Figs. 1 & 2), we reasoned that awake state should be associated with irregular deviations from linear flow, which would then be diminished during anesthesia. To test this hypothesis, we extended our approach

to an existing human neuroimaging resting-state fMRI dataset of 14 healthy volunteers (4 women; mean age 24 years, SD = 5[52]) who underwent propofol-induced anesthesia. As predicted, we found that anesthesia strikingly disrupted deviations from linear predictability compared to the awake state, with the linear forecasts performing significantly better across all subjects following propofol administration (blue line in Fig. 4d). This population-level observation can be better understood when viewing the Horizon plot of a specific subject recording, where the wake dynamics show periods of increased linear flow (Fig. 4b), which are equivalent to those observed for the same subject under propofol anesthesia (Fig. 4c). In contrast, the waking data is punctuated by periods of sharp deviation from linear flow that are not observed during propofol anesthesia (Fig. 4b). This observation thus provides confirmatory evidence for our hypothesis linking deviations from linear predictability to the conscious, waking state.

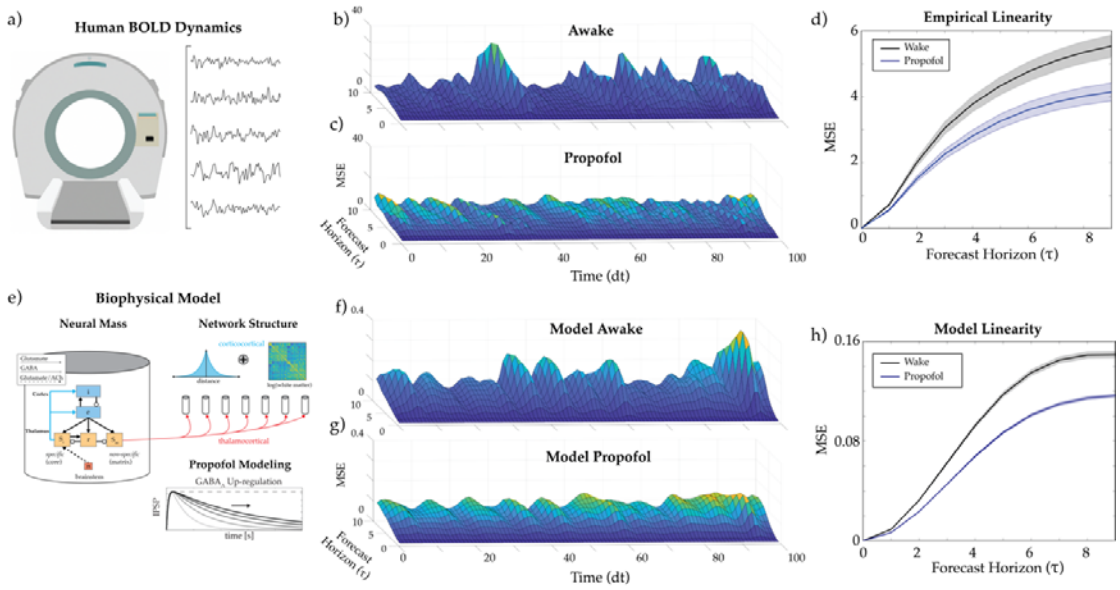


Figure 4 – Linear flow across arousal. a). Schematic of fMRI timeseries. b-c) a Horizon plot of an example single subject resting-state human BOLD showing time-varying linear error during wake (b) and during propofol anesthesia (c). d) Population average linear error of BOLD resting-state data during wake (black) and propofol (dark blue) conditions averaged over each recording – shaded regions show standard-error-of-mean (note: this is a group-level summary of the data in Figure 1c). e) Schematic of large-scale biophysical neural mass model fit to human resting-state BOLD data[13]. As

previous[13], propofol is modelled as a prolongation of inhibitory-postsynaptic-potentials. f) Horizon plot showing time-varying linear error of the corticothalamic model timeseries - transformed using a haemodynamic response function to provide representative BOLD timeseries - fit to wake condition. g) Time-varying linear error of the corticothalamic model timeseries under the propofol condition. h) Average linear error of corticothalamic model during wake (black) and propofol (dark blue) conditions.

To further interrogate this empirical observation, we leveraged a state-of-the-art, large-scale biophysical neural mass model capable of describing consciousness modulation under propofol anesthesia, and its recovery following stimulation of the matrix thalamus[13]. This corticothalamic model includes realistic population-level neuroanatomy, non-linear neural responses, interpopulation connections, long-range white-matter connectivity, and dendritic, synaptic, cell-body, and axonal dynamics[11,13,68–73]. The biophysical constraints incorporated in this model allow putative physiological mechanisms to be explored which provide insights beyond those obtained via purely data-driven approach.

Applying our same approach to the wake and propofol conditions of this model's timeseries – transformed using a haemodynamic response function to provide representative BOLD timeseries – we replicated our empirical observations: namely, relative to the propofol condition, the wake condition is comprised of epochs of linearly-predictable dynamics that were punctuated by sharp unpredictable deviations (Fig. 4f-g). These empirical and theoretical modelling results demonstrate that consciousness is concomitant with fluctuations in linear flow across the cerebral cortex, which are diminished by propofol anesthesia, however the question remains as to whether the thalamus itself plays a causal role in controlling these dynamics.

Causal thalamic drive of cortical non-linear flow across conscious states

A decisive feature of biophysical models is that they can be used to create predictions that can be directly tested in empirical datasets across a range of imaging

modalities and across species. We leveraged this feature to causally test the role of the thalamus in shaping the linearity of cortical dynamics. Specifically, we used an existing dataset comprising multi-electrode recordings from the frontal eye-fields and lateral interparietal cortical areas in a macaque monkey (Fig. 5b)[13,14]. Following propofol-mediated anesthesia, targeted 50Hz electrical stimulation of the matrix-rich central lateral thalamus drove the recovery of conscious arousal[13,14]. Despite the fact that the data was collected in a different species and through a different measurement modality that has a much faster temporal resolution than fMRI, we were able to use the generative nature of the biophysical model to create a testable prediction that mirrored the fMRI findings; Namely, that matrix stimulation should drive a recovery of the linearly unpredictable flow lost under propofol anesthesia (Fig. 5a) by flattening local attractors[11]. Importantly, this mechanism can be exploited while maintaining overall stability and facilitates long range interactions between specific brain areas[11,13].

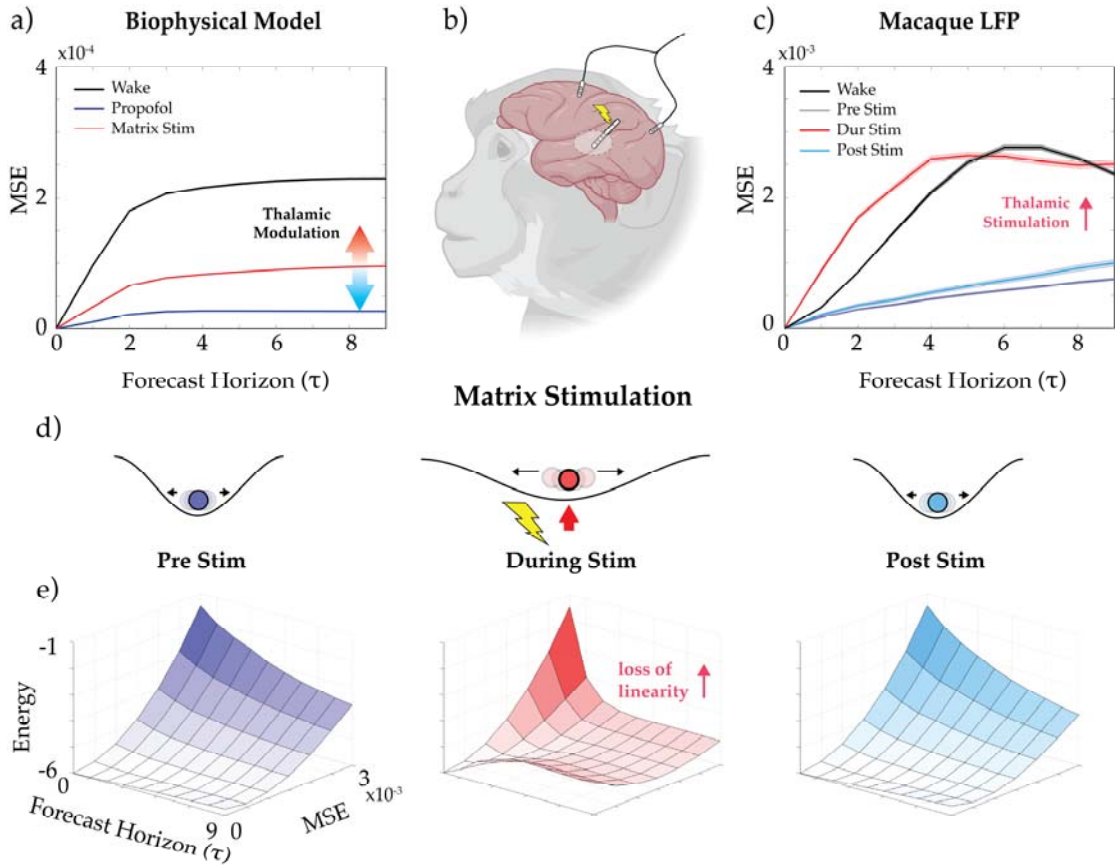


Figure 5 – Causal thalamic influence on cortical linear dynamics. a) Using the corticothalamic biophysical model, we simulated matrix thalamic stimulation by applying a square-wave input to matrix-rich thalamic regions and then applied our linear flow analysis to the cortical electrophysiological time series during wake (black), following simulated propofol administration (blue) and following matrix stimulation (red) – matrix stimulation recovered non-linear cortical dynamics. b) experimental setup, in which a macaque monkey had 50Hz electrical stimulation applied to the matrix-rich central lateral thalamus while local field potentials were recorded from the frontal eye fields and lateral interparietal cortex. c) examination of the empirical recordings from macaque cortex recovered the predicted relationships: namely, central lateral thalamic stimulation recovered cortical non-linearity that matched the awake brain but was lost during propofol anesthesia. d) these effects are analogous to the global cortical state being enslaved to a deep attractor during propofol anesthesia (purple and blue) whose depth was diminished by matrix thalamic stimulation (red), allowing less linear cortical state dynamics. e) energy landscapes[62] (i.e., the logarithm of inverse state probability) during pre-stim (purple), stimulation (red) and post-stimulation (blue) of the central lateral thalamus.

Using multi-electrode recordings from the macaque cortex, we confirmed that propofol had the predicted impact on cortical electrophysiology: namely, the induction of propofol increased linear predictability in the anesthetized monkeys, relative to wake (Fig. 5c), albeit on a shorter time-scale than the one observed in the fMRI data (though see Fig. S7). Next, we calculated this same linear predictability before (pre), during, and after (post) electrical stimulation of the central thalamus (which was effective in recovering the monkey's conscious state; Fig. 5d). As predicted by our biophysical model, we observed substantial propofol-mediated linearity in the pre-stimulation window, a significant decrease in linear dynamics during stimulation (comparable to that seen in the wake condition), and a subsequent return to relatively predictable, linear dynamics following stimulus cessation (Fig. 5d). In addition, while the empirical stimulation recovered proportionally larger amounts of linear errors than the biophysical model (contrast the amplitude of the red lines in Fig. 5a and 5c), we anticipate that this difference is related to the simplified nature of the model, relative the nuanced neurobiological complexity of the brain. Irrespective of these details, our crucial prediction was confirmed: namely, that matrix thalamic stimulation reinvigorated deviations from linear cortical dynamics lost during propofol anesthesia.

As a final analysis, we again utilized the likelihood of changes in linear dynamics across each stimulation condition to formulate a statistical energy landscape (akin to the results observed in Fig. 3f-h). Analogous to an attractor landscape, this approach highlights the stability of on-going predictable cortical linear flow in the pre- and post-stimulation conditions, which is then destabilized by matrix thalamic stimulation (Fig. 5d). Our findings demonstrate that the matrix thalamus drives linearly unpredictable cortical flow by destabilizing local attractors across different states of arousal and provides a key feature of flexible dynamics indicative of versatile and adaptive cognition.

Discussion

Here we showed that the thalamus controls the manner in which cortical dynamics fluctuate between laminar and non-laminar flow across states of consciousness. Using multimodal empirical neural recordings, a sensitive method from fluid dynamics and a biophysical model of the thalamocortical system, we demonstrated how a key principle of cellular organization within the thalamus controls these on-going linear dynamics, and specifically how the diffusely-projecting matrix thalamus destabilizes linear cortical flow. This approach reveals a candidate neuroanatomical mechanism for facilitating the adaptive cognitive benefits of the waking brain: namely, by facilitating brain dynamics that can support the formation of both stable and unstable brain states, which in turn can support reconfiguration as contingencies change across different behavioural contexts.

Previous work has shown how electrical stimulation of the thalamus can modulate conscious arousal[14,15] and implicated the matrix thalamus as driving quasi-critical brain state formation[11,13], thereby facilitating the recovery of conscious awareness from anesthesia. Our present findings enrich these insights and demonstrate how the matrix (diffuse)-core (targeted) cortical projections play a key role in shifting the cortex between laminar and non-laminar dynamics. Our results also represent a key capacity inherent to biophysical models, which is the ability to create mechanistic simulations that allow translations between different imaging modalities and recordings from different species. We anticipate that the addition of further anatomical constraints will only enrich the capacity for these models to provide additional testable predictions, though care must be taken to match the emergent dynamics of these models with the idiosyncratic spatiotemporal constraints imposed by standard cognitive contexts in which higher-order brain functions are interrogated.

Human neuroimaging studies have revealed a hierarchy of temporal and spatial autocorrelation scales across the cortex[49] and that these two features capture a large number of existing topological metrics[22]. Our present findings expand these insights by quantifying intrinsic dynamic modal timescales that drive linear flow in neural activity across imaging modalities. The resulting spatiotemporal dynamic modes are defined by a complex (i.e., imaginary) eigenvector and corresponding eigenvalue[32,35,50]. Each mode contains four key characteristic measures: the spatial eigenvector defines a spatial pattern of coherent activity, and a corresponding spatial pattern of delays relative to each of these coherence patterns – called a dephasing map. In addition, each mode has a corresponding eigenvalue defining its oscillatory frequency, and an exponential gain parameter – which determines whether a mode will grow or decay in time (i.e., captures its temporal stability). The investigation of these features provides rich information regarding the types of dynamic modes that characterize the system. Our results also confirm the long-held notion that BOLD dynamics are predominantly constituted by slow and sustained patterns of activity (hence the popularity of the standard resting state network analyses[74,75]), but extend our ability to characterize and explore the spatial and temporal properties of the modes shaping these dynamics. We also demonstrate that the same anatomical mechanism underpins similar observations in both BOLD and electrophysiological data, which thus paves the way forward for future mechanistic work. Furthermore, our comparison of empirical imaging data to surrogates conserving covariance and Fourier spectra shows these features are not sufficient to explain the observed linear dynamics. An important open question that we were unable to interrogate in these data is whether non-neural sources of noise, such as physiological activity or task contexts, alters the observed relationship between thalamic activity and cortical dynamics, however this line of questioning is left for future, targeted experimental approaches.

It is worth noting that in this implementation, a linear model is generated for each subject over the complete length of that recording. As with calculation of covariance and Fourier power spectrum, this may also be performed at a finer temporal granularity via windowing approaches – this will favour faster timescale effects over slower ones, and thus should depend on the investigators question of interest. Here, the imaging data has been collected during the resting-state, but recordings with specific epochs of interest, such as during a task paradigm, would be served better by a piece-wise approach defining the linear dynamics instantiated by the brain within each epoch, particularly given the time-scale constraints imposed by the vast majority of task-based neuroimaging analysis protocols. Furthermore, our approach allows for the quantification of linear dynamics within a given timeseries, however, it does not make explicit claims about the nature of the residual dynamics. This is because these residuals may be multifaceted. Indeed, it may be that deviations from linearity are driven by non-linearities (in the mathematical sense) within the dynamical system becoming significant, due to the presence of extrinsic or intrinsic drive (i.e., visual/auditory/olfactory stimuli in the brain), or through the instantiation of a different transient linear system on a shorter timescale relative to the epoch of time initially considered. These are both exciting opportunities to be explored in future work.

In conclusion, we have demonstrated fluctuations in globally linear cortical dynamics in multimodal neuroimaging and electrophysiological data across human and non-human primates differentiates conscious arousal states and are controlled by the organization of the thalamus. In this way, we argue that this key feature of the brain's neurobiological architecture helps to shape the robust, yet flexible, adaptive neural dynamics required for effective cognitive function that define our waking lives.

Acknowledgements

E.J.M. was supported by an ARC DECRA Fellowship.
J.M.S. was supported by an NHMRC Investigator Fellowship.

Declaration of Interests

The authors declare no competing interests.

References

1. Shine, J.M., Lewis, L.D., Garrett, D.D., and Hwang, K. (2023). The impact of the human thalamus on brain-wide information processing. Nat Rev Neurosci *24*, 416–430. https://doi.org/10.1038/s41583-023-00701-0.
2. Mofakham, S., Robertson, J., Lubin, N., Cleri, N.A., and Mikell, C.B. (2024). An Unpredictable Brain Is a Conscious, Responsive Brain. J Cogn Neurosci, 1–10. https://doi.org/10.1162/jocn_a_02154.
3. Whyte, C.J., Redinbaugh, M.J., Shine, J.M., and Saalmann, Y.B. (2024). Thalamic contributions to the state and contents of consciousness. Preprint at Cell Press, https://doi.org/10.1016/j.neuron.2024.04.019 https://doi.org/10.1016/j.neuron.2024.04.019.
4. Guillery, R.W., and Sherman, S.M. (2002). Review Thalamic Relay Functions and Their Role in Corticocortical Communication: Generalizations from the Visual System.
5. MacLean, J.N., Watson, B.O., Aaron, G.B., and Yuste, R. (2005). Internal dynamics determine the cortical response to thalamic stimulation. Neuron *48*, 811–823. https://doi.org/10.1016/j.neuron.2005.09.035.
6. Gao, Z., Davis, C., Thomas, A.M., Economo, M.N., Abrego, A.M., Svoboda, K., De Zeeuw, C.I., and Li, N. (2018). A cortico-cerebellar loop for motor planning. Nature *563*, 113–116. https://doi.org/10.1038/s41586-018-0633-x.
7. Shine, J.M. (2021). The thalamus integrates the macrosystems of the brain to facilitate complex, adaptive brain network dynamics. Preprint at Elsevier Ltd, https://doi.org/10.1016/j.pneurobio.2020.101951 https://doi.org/10.1016/j.pneurobio.2020.101951.
8. Müller, E.J., Munn, B., Hearne, L.J., Smith, J.B., Fulcher, B., Arnatkevičiūtė, A., Lurie, D.J., Cocchi, L., and Shine, J.M. (2020). Core and matrix thalamic sub-populations relate to spatio-temporal cortical connectivity gradients. Neuroimage *222*. https://doi.org/10.1016/j.neuroimage.2020.117224.
9. Jones, E.G. (2001). The thalamic matrix and thalamocortical synchrony. Trends Neurosci *24*, 595–601. https://doi.org/10.1016/S0166-2236(00)01922-6.
10. Garrett, D.D., Epp, S.M., Perry, A., and Lindenberger, U. (2018). Local temporal variability reflects functional integration in the human brain. Neuroimage *183*, 776–787. https://doi.org/10.1016/j.neuroimage.2018.08.019.
11. Müller, E., Munn, B., and Shine, J. (2020). Diffuse neural coupling mediates complex network dynamics through the formation of quasi-critical brain states. Nat Commun. https://doi.org/10.1101/2020.06.09.141416.
12. Kosciessa, J.Q., Lindenberger, U., and Garrett, D.D. (2021). Thalamocortical excitability modulation guides human perception under uncertainty. Nat Commun *12*. https://doi.org/10.1038/s41467-021-22511-7.
13. Müller, E.J., Munn, B.R., Redinbaugh, M.J., Lizier, J., Breakspear, M., Saalmann, Y.B., and Shine, J.M. (2023). The non-specific matrix thalamus facilitates the cortical information processing modes relevant for conscious awareness. Cell Rep *42*, 112844. https://doi.org/10.1016/j.celrep.2023.112844.
14. Redinbaugh, M.J., Phillips, J.M., Kambi, N.A., Mohanta, S., Andryk, S., Dooley, G.L., Afrasiabi, M., Raz, A., and Saalmann, Y.B. (2020). Thalamus Modulates Consciousness via Layer-Specific Control of Cortex. Neuron *106*, 66-75.e12. https://doi.org/10.1016/j.neuron.2020.01.005.
15. Tasserie, J., Uhrig, L., Sitt, J.D., Manasova, D., Dupont, M., Dehaene, S., and Jarraya, B. (2022). Deep brain stimulation of the thalamus restores signatures of consciousness in a nonhuman primate model. Sci Adv *8*. https://doi.org/10.1126/sciadv.abl5547.

16. Bastos, A.M., Donoghue, J.A., Brincat, S.L., Mahnke, M., Yanar, J., Correa, J., Waite, A.S., Lundqvist, M., Roy, J., Brown, E.N., et al. (2021). Neural effects of propofol-induced unconsciousness and its reversal using thalamic stimulation. Elife *10*. https://doi.org/10.7554/ELIFE.60824.
17. Deco, G., Kringelbach, M.L., Jirsa, V.K., and Ritter, P. (2017). The dynamics of resting fluctuations in the brain: Metastability and its dynamical cortical core. Sci Rep *7*. https://doi.org/10.1038/s41598-017-03073-5.
18. Bullmore, E., and Sporns, O. (2012). The economy of brain network organization. Preprint, https://doi.org/10.1038/nrn3214 https://doi.org/10.1038/nrn3214.
19. Purdon, P.L., Pierce, E.T., Mukamel, E.A., Prerau, M.J., Walsh, J.L., Wong, K.F.K., Salazar-Gomez, A.F., Harrell, P.G., Sampson, A.L., Cimenser, A., et al. (2013). Electroencephalogram signatures of loss and recovery of consciousness from propofol. Proc Natl Acad Sci U S A *110*. https://doi.org/10.1073/pnas.1221180110.
20. Liégeois, R., Laumann, T.O., Snyder, A.Z., Zhou, J., and Yeo, B.T.T. (2017). Interpreting temporal fluctuations in resting-state functional connectivity MRI. Preprint at Academic Press Inc., https://doi.org/10.1016/j.neuroimage.2017.09.012 https://doi.org/10.1016/j.neuroimage.2017.09.012.
21. Laumann, T.O., Snyder, A.Z., Mitra, A., Gordon, E.M., Gratton, C., Adeyemo, B., Gilmore, A.W., Nelson, S.M., Berg, J.J., Greene, D.J., et al. (2017). On the Stability of BOLD fMRI Correlations. Cerebral Cortex *27*, 4719–4732. https://doi.org/10.1093/cercor/bhw265.
22. Shinn, M., Hu, A., Turner, L., Noble, S., Preller, K.H., Ji, J.L., Moujaes, F., Achard, S., Scheinost, D., Constable, R.T., et al. (2023). Functional brain networks reflect spatial and temporal autocorrelation. Nat Neurosci *26*, 867–878. https://doi.org/10.1038/s41593-023-01299-3.
23. Jones, D.T., Vemuri, P., Murphy, M.C., Gunter, J.L., Senjem, M.L., Machulda, M.M., Przybelski, S.A., Gregg, B.E., Kantarci, K., Knopman, D.S., et al. (2012). Non-stationarity in the "resting brain's" modular architecture. PLoS One *7*. https://doi.org/10.1371/journal.pone.0039731.
24. Abbas, A., Belloy, M., Kashyap, A., Billings, J., Nezafati, M., Schumacher, E., and Keilholz, S. (2019). Quasi-periodic patterns contribute to functional connectivity in the brain Corresponding Author.
25. Vidaurre, D., Smith, S.M., and Woolrich, M.W. (2017). Brain network dynamics are hierarchically organized in time. Proc Natl Acad Sci U S A *114*, 12827–12832. https://doi.org/10.1073/pnas.1705120114.
26. Lurie, D.J., Kessler, D., Bassett, D.S., Betzel, R.F., Breakspear, M., Keilholz, S., Kucyi, A., Liégeois, R., Lindquist, M.A., McIntosh, A.R., et al. (2020). Questions and controversies in the study of time-varying functional connectivity in resting fMRI. Preprint at MIT Press Journals, https://doi.org/10.1162/netn_a_00116 https://doi.org/10.1162/netn_a_00116.
27. Designed Research; D, S.T.G.S.B., Performed Research; D, N.F.W.S.B., and Pnas, J. (2011). Dynamic reconfiguration of human brain networks during learning. *108*. https://doi.org/10.1073/pnas.1018985108/-/DCSupplemental.
28. Faskowitz, J., Esfahlani, F.Z., Jo, Y., Sporns, O., and Betzel, R.F. (2020). Edge-centric functional network representations of human cerebral cortex reveal overlapping system-level architecture. Nat Neurosci *23*, 1644–1654. https://doi.org/10.1038/s41593-020-00719-y.
29. Karahano□lu, F.I., and Van De Ville, D. (2015). Transient brain activity disentangles fMRI resting-state dynamics in terms of spatially and temporally overlapping networks. Nat Commun *6*. https://doi.org/10.1038/ncomms8751.

30. Shine, J.M., Koyejo, O., Bell, P.T., Gorgolewski, K.J., Gilat, M., and Poldrack, R.A. (2015). Estimation of dynamic functional connectivity using Multiplication of Temporal Derivatives. Neuroimage *122*, 399–407. https://doi.org/10.1016/j.neuroimage.2015.07.064.
31. Kutz, J.N., Brunton, S.L., Brunton, B.W., and Proctor, J.L. (2016). Dynamic Mode Decomposition (Society for Industrial and Applied Mathematics) https://doi.org/10.1137/1.9781611974508.
32. Proctor, J.L., Brunton, S.L., and Kutz, J.N. (2016). Dynamic mode decomposition with control. SIAM J Appl Dyn Syst *15*, 142–161. https://doi.org/10.1137/15M1013857.
33. Scherl, I., Strom, B., Shang, J.K., Williams, O., Polagye, B.L., and Brunton, S.L. (2020). Robust principal component analysis for modal decomposition of corrupt fluid flows. Phys Rev Fluids *5*, 1–22. https://doi.org/10.1103/PhysRevFluids.5.054401.
34. Schmid, P.J. (2010). Dynamic mode decomposition of numerical and experimental data. J Fluid Mech *656*, 5–28. https://doi.org/10.1017/S0022112010001217.
35. Brunton, B.W., Johnson, L.A., Ojemann, J.G., and Kutz, J.N. (2016). Extracting spatial-temporal coherent patterns in large-scale neural recordings using dynamic mode decomposition. J Neurosci Methods *258*, 1–15. https://doi.org/10.1016/j.jneumeth.2015.10.010.
36. Willert, C.E., and Gharib, M. (1991). Experiments in Fluids Digital particle image velocimetry.
37. Westerweel, J., Elsinga, G.E., and Adrian, R.J. (2013). Particle image velocimetry for complex and turbulent flows. Preprint, https://doi.org/10.1146/annurev-fluid-120710-101204 https://doi.org/10.1146/annurev-fluid-120710-101204.
38. Adrian, R.J. (1991). PARTICLE-IMAGING TECHNIQUES FOR EXPERIMENTAL FLUID MECHANICS.
39. Shine, J.M., Breakspear, M., Bell, P.T., Ehgoetz Martens, K., Shine, R., Koyejo, O., Sporns, O., and Poldrack, R.A. (2019). Human cognition involves the dynamic integration of neural activity and neuromodulatory systems. Nat Neurosci *22*, 289–296. https://doi.org/10.1038/s41593-018-0312-0.
40. Sirovich, L. (1987). TURBULENCE AND THE DYNAMICS OF COHERENT STRUCTURES PART I: COHERENT STRUCTURES*.
41. Berkooz, G., Holmes, P., and Lumley, J.L. (1993). The proper orthogonal decomposition in the analysis of turbulent flows. Annu Rev Fluid Mech *25*, 539–575. https://doi.org/10.1146/annurev.fl.25.010193.002543.
42. Deco, G., and Kringelbach, M.L. (2020). Turbulent-like Dynamics in the Human Brain. Cell Rep *33*. https://doi.org/10.1016/j.celrep.2020.108471.
43. Hearne, L.J., Cocchi, L., Zalesky, A., and Mattingley, J.B. (2017). Reconfiguration of brain network architectures between resting-state and complexity-dependent cognitive reasoning. Journal of Neuroscience *37*, 8399–8411. https://doi.org/10.1523/JNEUROSCI.0485-17.2017.
44. Gautama, T., Mandic, D.P., and Van Hulle, M.M. (2003). Signal nonlinearity in fMRI: A comparison between BOLD and MION. IEEE Trans Med Imaging *22*, 636–644. https://doi.org/10.1109/TMI.2003.812248.
45. Gultepe, E., and He, B. (2013). A linear/nonlinear characterization of resting state brain networks in fMRI time series. Brain Topogr *26*, 39–49. https://doi.org/10.1007/s10548-012-0249-7.
46. Nozari, E., Bertolero, M.A., Stiso, J., Caciagli, L., Cornblath, E.J., He, X., Mahadevan, A.S., Pappas, G.J., and Bassett, D.S. (2023). Macroscopic resting-state brain dynamics are best described by linear models. Nat Biomed Eng. https://doi.org/10.1038/s41551-023-01117-y.

47. Blinowska, K.J., and Malinowski, M. (1991). Non-linear and linear forecasting of the EEG time series.
48. Berger, H. (1934). Über das elektrenkephalogramm des menschen. DMW-Deutsche Medizinische Wochenschrift *60*, 1947–1949.
49. Fallon, J., Ward, P.G.D., Parkes, L., Oldham, S., Arnatkevičiūtė, A., Fornito, A., and Fulcher, B.D. (2020). Timescales of spontaneous fMRI fluctuations relate to structural connectivity in the brain. Network Neuroscience *4*, 788–806. https://doi.org/10.1162/netn_a_00151.
50. Kunert-Graf, J.M., Eschenburg, K.M., Galas, D.J., Kutz, J.N., Rane, S.D., and Brunton, B.W. (2019). Extracting Reproducible Time-Resolved Resting State Networks Using Dynamic Mode Decomposition. Front Comput Neurosci *13*, 1–17. https://doi.org/10.3389/fncom.2019.00075.
51. Thomas Yeo, B.T., Krienen, F.M., Sepulcre, J., Sabuncu, M.R., Lashkari, D., Hollinshead, M., Roffman, J.L., Smoller, J.W., Zöllei, L., Polimeni, J.R., et al. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. J Neurophysiol *106*, 1125–1165. https://doi.org/10.1152/jn.00338.2011.
52. Kandeepan, S., Rudas, J., Gomez, F., Stojanoski, B., Valluri, S., Owen, A.M., Naci, L., Nichols, E.S., and Soddu, A. (2020). Modeling an auditory stimulated brain under altered states of consciousness using the generalized ising model. Neuroimage *223*. https://doi.org/10.1016/j.neuroimage.2020.117367.
53. Ashourvan, A., Pequito, S., Bertolero, M., Kim, J.Z., Bassett, D.S., and Litt, B. (2022). External drivers of BOLD signal's non-stationarity. PLoS One *17*. https://doi.org/10.1371/journal.pone.0257580.
54. Clascá, F., Rubio-Garrido, P., and Jabaudon, D. (2012). Unveiling the diversity of thalamocortical neuron subtypes. European Journal of Neuroscience *35*, 1524–1532. https://doi.org/10.1111/j.1460-9568.2012.08033.x.
55. Acsády, L. (2017). The thalamic paradox. Preprint at Nature Publishing Group, https://doi.org/10.1038/nn.4583 https://doi.org/10.1038/nn.4583.
56. Halassa, M.M., and Sherman, S.M. (2019). Thalamocortical Circuit Motifs: A General Framework. Preprint at Cell Press, https://doi.org/10.1016/j.neuron.2019.06.005 https://doi.org/10.1016/j.neuron.2019.06.005.
57. Sherman, S.M., and Usrey, W.M. (2024). A Reconsideration of the Core and Matrix Classification of Thalamocortical Projections. Preprint at Society for Neuroscience, https://doi.org/10.1523/JNEUROSCI.0163-24.2024 https://doi.org/10.1523/JNEUROSCI.0163-24.2024.
58. Halassa, M.M., and Sherman, S.M. (2019). Thalamocortical Circuit Motifs: A General Framework. Preprint at Cell Press, https://doi.org/10.1016/j.neuron.2019.06.005 https://doi.org/10.1016/j.neuron.2019.06.005.
59. Jones, E.G. (2002). Thalamic circuitry and thalamocortical synchrony. Philosophical Transactions of the Royal Society B: Biological Sciences *357*, 1659–1673. https://doi.org/10.1098/rstb.2002.1168.
60. Jones, E.G. (1998). Viewpoint: the core and matrix of thalamic organization. Neuroscience *85*, 331–345. https://doi.org/10.1016/S0306-4522(97)00581-2.
61. Phillips, J.W., Schulmann, A., Hara, E., Winnubst, J., Liu, C., Valakh, V., Wang, L., Shields, B.C., Korff, W., Chandrashekar, J., et al. (2019). A repeated molecular architecture across thalamic pathways. Nat Neurosci *22*, 1925–1935. https://doi.org/10.1038/s41593-019-0483-3.
62. Munn, B.R., Müller, E.J., Wainstein, G., and Shine, J.M. (2021). The ascending arousal system shapes neural dynamics to mediate awareness of cognitive states. Nat Commun *12*, 1–9. https://doi.org/10.1038/s41467-021-26268-x.

63. Suzuki, M., and Larkum, M.E. (2020). General Anesthesia Decouples Cortical Pyramidal Neurons. Cell *180*, 666-676.e13. https://doi.org/10.1016/j.cell.2020.01.024.
64. Daniel Clement Dennett (1993). Consciousness Explained (Penguin UK).
65. Tononi, G., Boly, M., Massimini, M., and Koch, C. (2016). Integrated information theory: From consciousness to its physical substrate. Preprint at Nature Publishing Group, https://doi.org/10.1038/nrn.2016.44 https://doi.org/10.1038/nrn.2016.44.
66. Mashour, G.A., Roelfsema, P., Changeux, J.P., and Dehaene, S. (2020). Conscious Processing and the Global Neuronal Workspace Hypothesis. Preprint at Cell Press, https://doi.org/10.1016/j.neuron.2020.01.026 https://doi.org/10.1016/j.neuron.2020.01.026.
67. Setzer, B., Fultz, N.E., Gomez, D.E.P., Williams, S.D., Bonmassar, G., Polimeni, J.R., and Lewis, L.D. (2022). A temporal sequence of thalamic activity unfolds at transitions in behavioral arousal state. Nat Commun *13*. https://doi.org/10.1038/s41467-022-33010-8.
68. Robinson, P.A., Rennie, C.J., and Rowe, D.L. (2002). Dynamics of large-scale brain activity in normal arousal states and epileptic seizures. Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics *65*, 9. https://doi.org/10.1103/PhysRevE.65.041924.
69. Abeysuriya, R.G., Rennie, C.J., and Robinson, P.A. (2015). Physiologically based arousal state estimation and dynamics. J Neurosci Methods *253*, 55–69. https://doi.org/10.1016/j.jneumeth.2015.06.002.
70. Müller, E.J., van Albada, S.J., Kim, J.W., and Robinson, P.A. (2017). Unified neural field theory of brain dynamics underlying oscillations in Parkinson's disease and generalized epilepsies. J Theor Biol *428*. https://doi.org/10.1016/j.jtbi.2017.06.016.
71. Robinson, P.A., Rennie, C.J., Rowe, D.L., and O'Connor, C. (2004). Estimation of multiscale neurophysiologic parameters by electroencephalographic means. Hum Brain Mapp *23*, 53–72. https://doi.org/10.1002/hbm.20032.
72. Robinson, P.A., Rennie, C.J., Wright, J.J., and Bourke, P.D. (1998). Steady states and global dynamics of electrical activity in the cerebral cortex. Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics *58*, 3557–3571. https://doi.org/10.1103/PhysRevE.58.3557.
73. Deco, G., Jirsa, V.K., Robinson, P.A., Breakspear, M., and Friston, K. (2008). The dynamic brain: From spiking neurons to neural masses and cortical fields. PLoS Comput Biol *4*. https://doi.org/10.1371/journal.pcbi.1000092.
74. Smith, S.M., Vidaurre, D., Beckmann, C.F., Glasser, M.F., Jenkinson, M., Miller, K.L., Nichols, T.E., Robinson, E.C., Salimi-Khorshidi, G., Woolrich, M.W., et al. (2013). Functional connectomics from resting-state fMRI. Preprint, https://doi.org/10.1016/j.tics.2013.09.016 https://doi.org/10.1016/j.tics.2013.09.016.
75. Smith, S.M., Fox, P.T., Miller, K.L., Glahn, D.C., Fox, P.M., Mackay, C.E., Filippini, N., Watkins, K.E., Toro, R., Laird, A.R., et al. Correspondence of the brain's functional architecture during activation and rest.
76. Naci, L., Haugg, A., MacDonald, A., Anello, M., Houldin, E., Naqshbandi, S., Gonzalez-Lara, L.E., Arango, M., Harle, C., Cusack, R., et al. (2018). Functional diversity of brain networks supports consciousness and verbal intelligence. Sci Rep *8*. https://doi.org/10.1038/s41598-018-31525-z.
77. Hawrylycz, M.J., Lein, E.S., Guillozet-Bongaarts, A.L., Shen, E.H., Ng, L., Miller, J.A., Van De Lagemaat, L.N., Smith, K.A., Ebbert, A., Riley, Z.L., et al. (2012). An anatomically comprehensive atlas of the adult human brain transcriptome. Nature *489*, 391–399. https://doi.org/10.1038/nature11405.
78. Gryglewski, G., Seiger, R., James, G.M., Godbersen, G.M., Komorowski, A., Unterholzner, J., Michenthaler, P., Hahn, A., Wadsak, W., Mitterhauser, M., et al.

(2018). Spatial analysis and high resolution mapping of the human whole-brain transcriptome for integrative analysis in neuroimaging. Neuroimage *176*, 259–267. https://doi.org/10.1016/j.neuroimage.2018.04.068.

79. Mü, M.C., Waldvogel, H.J., and Faull, R.L.M. (2000). The distribution of calbindin, calretinin and parvalbumin immunoreactivity in the human thalamus.
80. Redinbaugh, M.J., Afrasiabi, M., Phillips, J.M., Kambi, N.A., Mohanta, S., Raz, A., and Saalmann, Y.B. (2022). Thalamic deep brain stimulation paradigm to reduce consciousness: Cortico-striatal dynamics implicated in mechanisms of consciousness. PLoS Comput Biol *18*. https://doi.org/10.1371/journal.pcbi.1010294.
81. Michelle J. Redinbaugh, Mohsen Afrasiabi, Jessica M. Phillips, Niranjan A. Kambi, Sounak Mohanta, Y.B.S. (2021). Thalamic deep brain stimulation as a paradigm to reduce consciousness: implications for cortico-striatal dynamics, absence epilepsy and consciousness studies. 6.
82. Afrasiabi, M., Redinbaugh, M.J., Phillips, J.M., Kambi, N.A., Mohanta, S., Raz, A., Haun, A.M., and Saalmann, Y.B. (2021). Consciousness depends on integration between parietal cortex, striatum, and thalamus. Cell Syst *12*, 363-373.e11. https://doi.org/10.1016/j.cels.2021.02.003.
83. Sanz-Leon, P., Robinson, P.A., Knock, S.A., Drysdale, P.M., Abeysuriya, R.G., Fung, F.K., Rennie, C.J., and Zhao, X. (2018). NFTsim: Theory and Simulation of Multiscale Neural Field Dynamics https://doi.org/10.1371/journal.pcbi.1006387.
84. Abeysuriya, R.G., and Robinson, P.A. (2016). Real-time automated EEG tracking of brain states using neural field theory. J Neurosci Methods *258*, 28–45. https://doi.org/10.1016/j.jneumeth.2015.09.026.
85. Robinson, P.A., Loxley, P.N., O'Connor, S.C., and Rennie, C.J. (2001). Modal analysis of corticothalamic dynamics, electroencephalographic spectra, and evoked potentials. Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics *63*, 1–13. https://doi.org/10.1103/PhysRevE.63.041909.
86. Robinson, P.A., Rennie, C.J., Wright, J.J., Bahiumuli, H., Gordon, E., and Rowe, D.L. (2001). Prediction of electrocnccphulographic spectra from neurophysiology. Phys Rev E Stat Nonlin Soft Matter Phys *63*, 0219031–02190318. https://doi.org/10.1103/PhysRevE.63.021903.
87. Buzsáki, G., Anastassiou, C.A., and Koch, C. (2012). The origin of extracellular fields and currents-EEG, ECoG, LFP and spikes. Preprint, https://doi.org/10.1038/nrn3241 https://doi.org/10.1038/nrn3241.
88. Royer, J., Rodríguez-Cruces, R., Tavakol, S., Larivière, S., Herholz, P., Li, Q., Vos de Wael, R., Paquola, C., Benkarim, O., Park, B., et al. (2022). An Open MRI Dataset For Multiscale Neuroscience. Sci Data *9*, 569. https://doi.org/10.1038/s41597-022-01682-y.
89. Schaefer, A., Kong, R., Gordon, E.M., Laumann, T.O., Zuo, X.-N., Holmes, A.J., Eickhoff, S.B., and Yeo, B.T.T. (2018). Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. Cerebral Cortex *28*, 3095–3114. https://doi.org/10.1093/cercor/bhx179.
90. Fischl, B. (2012). FreeSurfer. Neuroimage *62*, 774–781. https://doi.org/10.1016/j.neuroimage.2012.01.021.
91. Hearne, L.J., Cocchi, L., Zalesky, A., and Mattingley, J.B. (2017). Reconfiguration of brain network architectures between resting-state and complexity-dependent cognitive reasoning. Journal of Neuroscience *37*, 8399–8411. https://doi.org/10.1523/JNEUROSCI.0485-17.2017.

92. Wright, J.J., and Liley, D.T.J. (1996). Dynamics of the brain at global and microscopic scales: Neural networks and the EEG. Behavioral and Brain Sciences *19*, 285–295. https://doi.org/10.1017/s0140525x00042679.
93. Braitenberg, V., and Schuz, A. (2013). Cortex: Statistics and Geometry of Neuronal Connectivity.
94. Deco, G., Ponce-Alvarez, A., Hagmann, P., Romani, G.L., Mantini, D., and Corbetta, M. (2014). How local excitation-inhibition ratio impacts the whole brain dynamics. Journal of Neuroscience *34*, 7886–7898. https://doi.org/10.1523/JNEUROSCI.5068-13.2014.
95. Unzai, T., Kuramoto, E., Kaneko, T., and Fujiyama, F. (2017). Quantitative Analyses of the Projection of Individual Neurons from the Midline Thalamic Nuclei to the Striosome and Matrix Compartments of the Rat Striatum. Cerebral Cortex *27*, 1164–1181. https://doi.org/10.1093/cercor/bhv295.
96. Lliná, R.R., Leznik, E., and Urbano, F.J. (2001). Temporal binding via cortical coincidence detection of specific and nonspecific thalamocortical inputs: A voltage-dependent dye-imaging study in mouse brain slices.
97. Crabtree, J.W., and Isaac, J.T.R. (2002). New Intrathalamic Pathways Allowing Modality-Related and Cross-Modality Switching in the Dorsal Thalamus.
98. Van Der Werf, Y.D., Witter, M.P., and Groenewegen, H.J. (2002). T he intralaminar and midline nuclei of the thalamus. Anatomical and functional evidence for participation in processes of arousal and awareness.
99. Noroozbabaee, L., Steyn-Ross, D.A., Steyn-Ross, M.L., and Sleigh, J.W. (2021). Analysis of the Hindriks and van Putten model for propofol anesthesia: Limitations and extensions. Neuroimage *227*. https://doi.org/10.1016/j.neuroimage.2020.117633.
100. 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., et al. (2016). Situating the default-mode network along a principal gradient of macroscale cortical organization. Proc Natl Acad Sci U S A *113*, 12574–12579. https://doi.org/10.1073/pnas.1608282113.

Methods

Human neuroimaging data

Sixty healthy adult participants (28 females; 18–33 years; right-handed) were recruited and the research was approved by The University of Queensland Human Research Ethics Committee. These data were originally described in [43]. 1050 (~10 minutes) whole-brain 7T resting state fMRI echo planar images were acquired using a multiband sequence (acceleration factor = 5; 2 mm^3 voxels; 586 ms TR; 23 ms TE; 40^0 flip angle; 208 mm FOV; 55 slices). Structural images were also collected to assist functional data pre-processing (MP2RAGE sequence – 0.75 mm^3 voxels 4,300 ms TR; 3.44 ms TE; 256 slices).

DICOM images were first converted to NIfTI format and realigned. T1 images were reoriented, skull-stripped (FSL BET), and co-registered to the NIfTI functional images using statistical parametric mapping functions. Segmentation and the DARTEL algorithm were used to improve the estimation of non-neural signal in subject space and the spatial normalization. From each grey-matter voxel, the following signals were regressed: linear trends, signals from the six head-motion parameters (three translation, three rotation) and their temporal derivatives, white matter, and CSF (estimated from single-subject masks of white matter and CSF). The aCompCor method (Behzadi et al., 2007) was used to regress out residual signal unrelated to neural activity (i.e., five principal components derived from noise regions- of-interest in which the time series data were unlikely to be modulated by neural activity). Participants with head displacement > 3 mm in > 5% of volumes in any one scan were excluded (n = 5). A temporal band pass ($0.001 < f < 0.125$ Hz) was applied to the data. Following pre-processing, the mean time series was extracted from 400 pre-defined cortical parcels using the Schaefer atlas (Schaefer et al., 2018).

For the anesthesia BOLD fMRI dataset, we utilized data derived from previously published research works[52,76] that have been openly shared on the OpenNeuro data repository (doi:10.18112/openneuro.ds003171.v2.0.0). The dataset comprises 17 healthy individuals (Age: 24± 5, M/F: 13/4), all of whom were right-handed, native English speakers, and had no recorded history of neurological disorders. The original study obtained ethical approval from both the Health Sciences Research Ethics Board and Psychology Research Ethics Board of Western University (REB #104,755) and adhered to the principles outlined in the revised declaration of Helsinki (2000). We analyzed two conditions, awake (fully alert and communicative) and deep sedation, which was achieved with an initial target effect-site concentration of 0.6 μg/ml and oxygen titrated to maintain SpO2 above 96%, with increments of 0.3 μg/ml with repeated assessments of responsiveness until deep sedation (Ramsey level 5).

Imaging was performed on a 3T Siemens Tim Trio system with a 32-channel head coil. Subjects resting state fMRI scans using BOLD EPI sequence (33 slices, voxel size: $3mm^3$ isotropic, inter-slice gap of 25%, TR = 2000 ms, TE = 30 ms, matrix size = 64×64, FA = 75°). Resting-state scans had 256 volumes. Anatomical scans were also obtained using a T1-weighted 3D Magnetization Prepared - Rapid Gradient Echo (MPRAGE) sequence (voxel size: $1mm^3$ isotropic, TR = 2.3, TE = 4.25 ms, matrix size = 240 × 256 × 192, FA = 9°).

Preprocessing was completed using *fMRIPrep* standard pipeline which involves the basic preprocessing steps (co-registration, normalization, unwarping, noise component extraction, segmentation, skull stripping, etc.). The extracted time series were denoised specifying motion and physiological signals from white matter and CSF, with high-pass and low-pass band filters set at 0.01 and 0.1, respectively. The global signal was removed by standardizing each time point[39].

Thalamic timeseries representing the 'core' and 'matrix' populations were estimated by estimating the dot-product between the voxel-wise time-series of thalamic BOLD signal and a standardized map of the relative mRNA expression levels for PVALB ('core') and CALB1 ('matrix') provided by the Allen Human Brain Atlas[8,9,77]. Variogram modelling was applied prior to our obtaining the gene maps to correct for potential subject-level differences[78]. Note that there are other calcium binding proteins with non-trivial expression in the thalamus[61,79], however these patterns were not considered here, as they have not been directly linked to the same core-matrix gradient in the thalamus[9]. Note that we do not advocate for a strict dichotomy between core and matrix neurons in this work, but rather that these two projection types represent anatomical extremes at either end of an approximately continuous spectrum[54,56]. See Muller et al 2020[8] for more details.

Macaque electrophysiological dataset

We compared modelling data to electrophysiology data from our previous study[14]. This dataset consists of simultaneous local field potential (LFP) recordings from the frontal eye field (FEF), lateral intraparietal area (LIP) and central lateral thalamus (CL) in the right hemisphere of two macaques (*Macaca mulatta*, 4.3-5.5 years old, 7.63-10.30 kg body weight). We lowpass filtered LFPs to 250 Hz, then linearly detrended and extracted artifacts. Bipolar derivations of the LFPs were then calculated to minimize any possible effects of a common reference and volume conduction. Recordings were performed during general anesthesia – either isoflurane (0.8%–1.5% on 1 L/min O2 flow; 9 sessions) or propofol (0.17-0.33 mg/kg/min i.v.; 9 sessions) – or wakefulness, using 16- or 24-contact linear micro-electrode arrays (LMEAs; MicroProbes). The electrode contacts had a diameter of $12.5\mu m$, and $200\mu m$ spacing between contacts.

We performed deep brain stimulation via simultaneous stimulation of 16 contacts of the LMEAs in the central thalamus. We applied 400 ms bi-phasic pulses of 200 mA,

at 50 Hz stimulation frequency, for a total of 60 s stimulation duration for any given stimulation event. To localize LMEAs, we averaged two 3D T1-weighted structural images of the MRI-compatible electrodes *in situ* (inversion-recovery prepared gradient echo sequence with: FOV = $128mm^2$; matrix = 256 x 256; no. of slices = 166; 0.5 mm isotropic; TR = 9.68 ms; TE = 4.192 ms; flip angle = 12°; inversion time (TI) = 450 ms) and cross-validated the electrode location with electrode depth measurements during recordings/stimulation. We used an arousal index (0-10) based on eye openings, body movements and vital signs, as well as EEG and EMG, to measured stimulation-induced changes in the level of consciousness. We defined effective stimulations, increasing the level of consciousness, as having an arousal index of ≥3, which all experimenters could differentiate from ineffective stimulations having an arousal index of 0-2. See [14,80–82] for more details. The University of Wisconsin – Madison Institutional Animal Care and Use Committee approved all procedures, which conformed to the National Institutes of Health Guide for the Care and Use of Laboratory Animals. Due to hardware constraints, the LFP time-series were down-sampled from 1kHz to 100Hz for linear model estimation – although we note that results in Fig. 5c and subsequent conclusions are not timescale sensitive as shown in Supp. Fig. 7.

Linear Forecast Analysis (LFA)

Here we outline LFA as applied to all datasets within this manuscript. First, each timeseries $\mathbf{Z} = [\boldsymbol{z}_1\ \boldsymbol{z}_2\ \boldsymbol{z}_3\ \dots\ \boldsymbol{z}_n]$ with n observations of p variables is partitioned into a pre $\mathbf{X}$ and post $\mathbf{Y}$ timeseries – where:

$$\mathbf{X} = [\boldsymbol{z}_1\ \boldsymbol{z}_2\ \boldsymbol{z}_3 \dots \boldsymbol{z}_{n-1}]$$
$$\mathbf{Y} = [\boldsymbol{z}_2\ \boldsymbol{z}_3\ \boldsymbol{z}_4 \dots \boldsymbol{z}_n]$$

with $\mathbf{X}$ lacking the final time point of $\mathbf{Z}$ and $\mathbf{Y}$ lacking the initial time point. Note we have dropped the function-of-time notation for simplicity. Singular value

decomposition is then applied, $\mathbf{X} = \mathbf{U} \times \mathbf{S} \times \mathbf{V}^{\mathrm{T}}$, and a rank reduction of the data is made that captures 95% of the explained variance. Although not explicitly necessary for timeseries with significantly more timepoints than variables, some of the data used here violate this condition and thus the results are standardized around this threshold – see Supp. Fig. 4 showing truncation effects on eigenspectrum. A linear propagator matrix $\mathbf{A}$ can be estimated from the data where $\mathbf{Y} = \mathbf{AX}$ as follows:

$$\mathbf{A} \approx \widetilde{\mathbf{A}} = \mathbf{YX}^{\dagger} \triangleq \mathbf{YV\Sigma}^{-1}\mathbf{U}^{*}$$

By stepping through each timepoint of $\mathbf{Z}$ projected into the SVD sub-space, the linear propagator can be used to forecast the proceeding time points. Mean-squared-error is used to compare between the predicted timeseries and the ground-truth timeseries:

$$MSE_{t,\tau} = \langle \left| \mathbf{Z}_{t+\tau} - \tilde{\mathbf{Z}}_{t+\tau} \right|^2 \rangle_k$$

where $\mathbf{Z}$ is the ground-truth data and $\tilde{\mathbf{Z}}$ is the forecasted timeseries using the linear model, and k is the number of SVD modes.

Dynamic mode decomposition

The eigen decomposition of the linear propagator $\mathbf{A}$ gives the dynamic modes of the system, $\mathbf{AW} = \boldsymbol{W\Lambda}$ where $\mathbf{W}$ is the matrix of eigenvectors, and $\boldsymbol{\Lambda}$ is the diagonal matrix of eigenvalues λ_i. The DMD approximation of the data can then be written as the simple dynamic model:

$$\hat{\mathbf{Z}}(t) = \boldsymbol{\Phi} \exp(\boldsymbol{\Omega}\mathrm{t})\, \mathbf{z}_i$$

where $\boldsymbol{\Omega} \triangleq \mathbf{YV\Sigma}^{-1}\mathbf{W}$ with $\boldsymbol{\Omega} = \log(\boldsymbol{\Lambda})/\Delta\mathrm{t}$ and t is time[35]. $\boldsymbol{z}_i$ is the initialization state and can be solved for each timepoint using the pseudoinverse.

Surrogate models

The spectral surrogate utilized here is adapted from [21] and generates a white noise timeseries with the same dimensions as the empirical data. The Fourier transform is then applied to both the surrogate and empirical timeseries, and these are multiplied together in this spectral domain. Following the inverse Fourier transform, the resulting timeseries is projected through the eigenvectors of the empirical covariance matrix to give a null model time series with matched covariance and Fourier power spectrum.

The shuffled surrogate is generated by performing a uniform scrambling of the empirical time-series indices. This ensures regional co-activations at each time point are conserved, but their temporal position relative to activations in scramble – i.e., the same covariance matrix but scrambled moment-to-moment changes.

Linear mode clusters

k-means clustering is used to define clusters of linear mode timescales and spatial coherence independently. In both cases, a sweep of cluster sizes k=2-20 with 100 repetitions for each is performed. A peak in adjusted mutual information (calculated using BCT http://www.brain-connectivity-toolbox.net/) between each repetition's clusters for the same cluster size is used to select a cluster size of k=5. Having selected this cluster size a final run of 1000 repetitions is used to find the optimal clustering. Timescale clustering is performed on the eigenvalues of all subjects (N=59) linear modes, i.e., the 2-dimensional complex numbers defining damping rates and oscillation frequencies. Spatial clustering is performed on the coherence maps defines by each linear mode across all subjects (N=59), i.e., the real component of the complex eigenvectors of the linear propagator.

Energy landscapes

Following methodology from previous work[62], we formulate an energy landscape by calculating the probability of observing a given linear model error (MSE) across the entire timeseries using a Gaussian kernel density estimation,

$$P(MSE, t) = \frac{1}{4N}\sum_{i=1}^{n} K\left(\frac{MSE_{t,i}}{4}\right)$$

where $K(u) = \frac{1}{2\sqrt{\pi}} e^{-\frac{1}{2}u^2}$. As is typical in statistical mechanics the energy of a given state, E_σ, and its probability are related by $P(\sigma) = \frac{1}{Z} e^{-\frac{E_\sigma}{T}}$ where Z is the normalization function and T is the scaling factor equivalent to temperature in thermodynamics[62]. In our analysis $\sum_\sigma P_\sigma = 1 \rightarrow Z = 1$ by construction and we can set $T = 1$ for the observed data. Thus, the energy of each MSD at a given time-lag t, E is then equal to the natural logarithm of the inverse probability, $P(MSE, t)$ of its occurrence,

$$E = \ln\left(\frac{1}{P(MSE, t)}\right)$$

Corticothalamic model

The corticothalamic model consists of 400 coupled neural masses. We outline this architecture by first detailing the corticothalamic neural mass as follows. The corticothalamic neural mass model used in this work contains four distinct populations: an excitatory pyramidal cell, e, and an inhibitory interneuron, i, population in the cortex; and two excitatory nuclei, matrix, s_m; core, s_c, and inhibitory thalamic reticular nuclei, r, population in the thalamus (Fig. 1c). The dynamical processes that occur within and between populations in a neural field model are defined as follows:

For each population, the mean soma potential results from incoming postsynaptic potentials (PSPs):

$$V_a(t) = \sum_b V_{ab}(t) \tag{1}$$

where $V_{ab}(t)$ is the result of a postsynaptic potential of type b onto a neuron of type a and $a, b \in \{e, i, r, s\}$. The postsynaptic potential response in the dendrite is given by

$$D_{ab}V_{ab}(t) = \nu_{ab}\phi_{ab}(t - \tau_{ab}) \tag{2}$$

where the influence of incoming spikes to population a from population b is weighted by a connection strength parameter $\nu_{ab} = N_{ab}s_{ab}$, with the mean number of connections between the two populations N_{ab} and s_{ab} is the mean strength of response in neuron a to a single spike from neuron b. τ_{ab} is the average axonal delay for the transmission of signals, and ϕ_{ab} is the mean axonal pulse rate from b to a.

The operator D_{ab} describes the time evolution of V_{ab} in response to synaptic input,

$$D_{ab} = \frac{1}{\alpha\beta}\frac{d^2}{dt^2} + \left(\frac{1}{\alpha} + \frac{1}{\beta}\right)\frac{d}{dt} + 1 \tag{3}$$

where β and α are the overall rise and decay response rates to the synaptodendritic and soma dynamics.

The mean firing rate of a neural population $Q_a(t)$ can be approximately related to its mean membrane potential, $V_a(t)$, by

$$\begin{aligned} Q_a(t) &= S_a[V_a(t)] \\ &= \frac{Q_a^{max}}{1 + \exp[-\{V_a(t) - \theta_a\}/\sigma']} \end{aligned} \tag{4}$$

which define a sigmoidal mapping function S_a with a maximal firing rate Q_a^{max}, a mean firing threshold θ_a, and a standard deviation of this threshold $\sigma'\pi/\sqrt{3}$.

The mean axonal pulse rate is related to the mean firing rate by,

$$D_a(t)\phi_a(t) = Q_a(t) \tag{5}$$

$$D_a(t) = \frac{1}{\gamma_a^2}\frac{\partial^2}{\partial t^2} + \frac{2}{\gamma_a}\frac{\partial}{\partial t} + 1 \tag{6}$$

Here, $\gamma_a = v_a/r_a$ represents the damping rate, where v_a is the propagation velocity in axons, and r_a is the characteristic axonal length for the population.

A network of 400 corticothalamic neural masses were simulated using the neural field simulation software, *NFTsim*[83]. The parameters for each neural mass were identically set to *"eyes-closed"* estimates given in Table 1[68,69,71,83,84], which results in simulated activity with a 1/f spectrum and a peak in the alpha frequency band (8-13 Hz) in the absence of network coupling. These are example parameters representative of the *"eyes-closed"* state following Bayesian model fits to human EEG power spectra[69]. Many preceding studies[68,69,71,72,84–86] have shown the linear transfer function, which drives the linear spectral content of the corticothalamic model, is derived and shown to be low-dimensional, i.e., only a few loop gains in the system are needed to capture the key features of the power spectra. In this particular context, the eyes-closed state of human EEG has a 1/f slope and a spectral peak in the 8-13Hz alpha frequency band, and this is explained by a weakly damped thalamocortical loop gain (ese). For the present study we have selected characteristic parameters for this power spectrum (Table 7 from [69]). Note that this spectrum describes modulations of firing rate around a fixed point, which are static for the network. These spike rate modulations in neurons drive changes in the extracellular electric field which are then measured via EEG and LFP recordings[87]. In this way, we

are able to compare our model outputs to empirical data through the power spectral density function.

Each simulation was run for a total of 64s with 7.5s of initial transients removed using an integration timestep of $\Delta t = 2^{-13}$s. This minimizes contributions from the model's initial state and ensures the integration algorithm has stabilized before we begin analysing simulation outputs. Longer simulations produced qualitatively identical results, as did shorter simulations, however, many of the analysis measures presented in this paper perform better with more data – i.e., correlation and coherence are noisier with less data. Thus, a balance between metric accuracy, resource allotment, and tractability for dataset manipulations was chosen. All outputs were down-sampled to 200Hz for tractability. All remaining data were used for subsequent analysis.

Structural Connectivity

The structural connectivity used to define the model network consists of a combination of distance dependence and long-range connectivity estimated from white-matter fibre densities measurement[88]. The distance dependence was generated via an exponentially decreasing function. First, the geodesic distance between all nodes, which correspond to parcels from[89] with MNI coordinates, are calculated along the *fsaverage* cortical surface mesh[90] using the Fast-Marching algorithm (Gabriel Peyre (2022)ToolboxFastMarching:https://www.mathworks.com/matlabcentral/fileexchange/6110-toolbox-fast-marching). The geodesic distances are then scaled as an exponentially decreasing function of distance,

$$G_{ij} = e^{-\lambda d_{ij}}$$

where d_{ij} is the geodesic distance in MNI space along the surface mesh and λ is the decay rate. A 200x200 connection matrix is defined in this way for each hemisphere. We further assume that interhemispheric connectivity is symmetric and one-to-one, and thus the full 400x400 distance dependence network is composed as

$$M = \begin{bmatrix} G_{ij}^{LH} & G_{ij}^{RH} \\ G_{ij}^{LH} & G_{ij}^{RH} \end{bmatrix}$$

where G_{ij}^{LH} and G_{ij}^{RH} are the left and right interhemispheric connectivities, respectively.

The complete network connectivity is then formulated via the summation of the distance dependence matrix and the empirically estimated white-matter connectivity (both normalized by their respective maximum values). Since the strength of these connections is not known empirically, we follow other approaches and sweep values of a global scaling of this hybrid connection matrix, as well as the proportion of distance dependence-to-white-matter connectivity and distance decay rate parameter. Functional connectivity of these parameter sweeps is then compared to empirical resting-state BOLD data [91] to define optimal values (see Supp. Fig. 2)

Model balancing

In order to maintain stability in the model, excitatory inputs to a given node, coupled via the structural network connections, must be balanced with a corresponding inhibition. To do this, we first compute the total excitatory connection strength to each corticothalamic node.

We then leverage an assumption from previous neural field models, namely that excitatory and inhibitory synapses in the cortex can be assumed proportional to the number of neurons [86,92]. This random connectivity approximation results in $\nu_{ee} =$

ν_{ie}, and $\nu_{ei} = \nu_{ii}$ which implies $V_e = V_i$ and $Q_e = Q_i$. Inhibitory population variables can then be expressed in terms of excitatory quantities. Whilst we do not make this assumption in the present model, we can leverage it to refine an inhibition scaling that balances the excitatory inputs from our specific structural network.

In the reduced corticothalamic neural mass, the fixed-point attractors, or steady states are found by setting all time derivatives in the above equations to zero. The steady-state values $\phi_e^{(0)}$ of ϕ_e are then given by solutions of

$$\begin{aligned} & S^{-1}\left(\phi_e^{(0}\right) - (\nu_{ee} + \nu_{ei})\phi_e^{(0)} \\ & = \nu_{es} S\left\{\nu_{se}\phi_e^{(0)} + \nu_{sr} S\left[\nu_{re}\phi_e^{(0)} + \frac{\nu_{rs}}{\nu_{es}}\left\{S^{-1}\left(\phi_e^{(0}\right) - (\nu_{ee} + \nu_{ei})\phi_e^{(0)}\right\}\right] + \right. \\ & \left. \nu_{sn}\phi_n^{(0)}\right\}, \end{aligned} \tag{7}$$

where $\phi_n^{(0)}$ is the steady state component of the input stimulus [92,93]. Roots of Eq. (7) are found using the *fzero()* function from MATLAB.

Following similar approaches[94], we leverage Eqn. 7 by setting ν_{ee} equal to each corticothalamic neural masses network coupling defined by the structural connectivity. We then set the cortical firing rate to be 3Hz, in-line with empirical observations, and numerically solve for cortical inhibition ν_{ei}. This results in each neural mass having a 3Hz steady-state cortical firing rate across the network, despite having heterogeneous network connectivity.

Note that the diffuse matrix inputs to each corticothalamic neural mass are excluded from this balancing as they are purely excitatory and only target the excitatory cortical population. Thus, the overall effect of matrix inputs is to distort each local attractor, increasing their firing rates when they are coupling to the network[11]. In addition, matrix nuclei are known to project to the reticular thalamic nucleus in

rodents [9,95–97], albeit weakly[98], but have been excluded from the current model for simplicity.

Modelling Propofol

The effect of propofol is modelled as an up-regulation of GABA-a receptors which prolongs inhibitory postsynaptic response potentials. This is implemented as an increase to the synaptodendritic functions (Eqn. 3) decay rate parameter, α, for all inhibitory connections in the corticothalamic neural mass. In addition, consistent with previous approaches[99], we maintain a constant peak amplitude of the IPSP functions following the rescaling. The solution to Eqn. 3 for a delta function input corresponds to,

$$V = \frac{\alpha\beta}{(\beta-\alpha)}\left[e^{-\alpha t} - e^{-\beta t}\right] \tag{8}$$

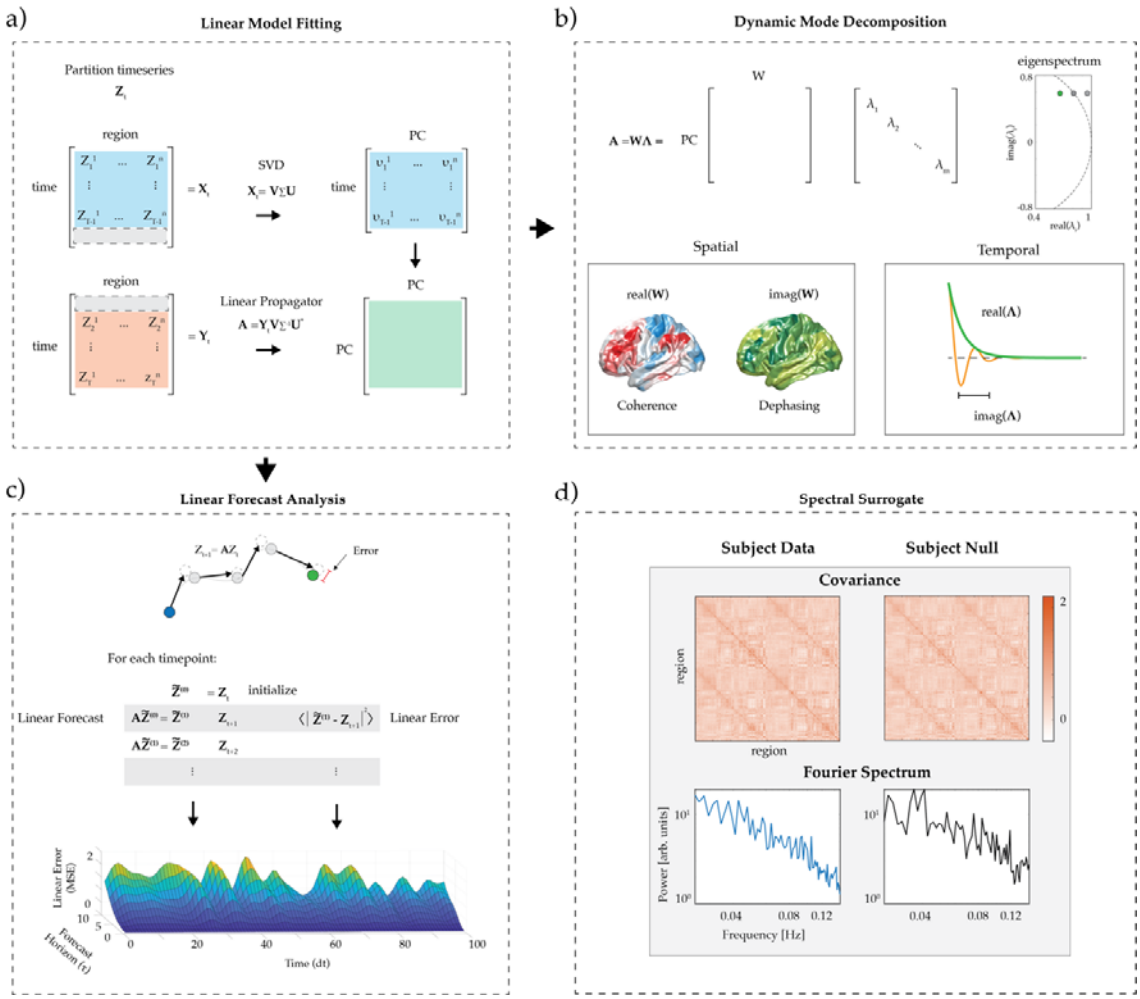
which has a peak amplitude at $t_{\text{peak}} = \frac{\ln^{\beta}/_{\alpha}}{(\beta-\alpha)}$. The rescaling of $\alpha = {}^{\alpha_0}/_{\rho}$ defines a new peak potential at t_{peak} which is renormalized to its pre-propofol value. We leverage the change in coherent alpha-band activity (8-13Hz) observed in [14] between LIP and FEF to optimize propofols effect in the model, which results in $\rho = 1.127$.

Parameter	***Description***	***Value***	***Unit***
γ_e	Cortical damping rate	116	s^{-1}
Q^{max}	Maximum firing rate	340	s^{-1}
θ	Firing threshold	12.9	mV
σ'	Threshold spread	3.8	mV
ϕ_n	Input noise amplitude spectral density	1×10^{-5}	s^{-1}
α	Decay rate of cell-body potential	83	s^{-1}

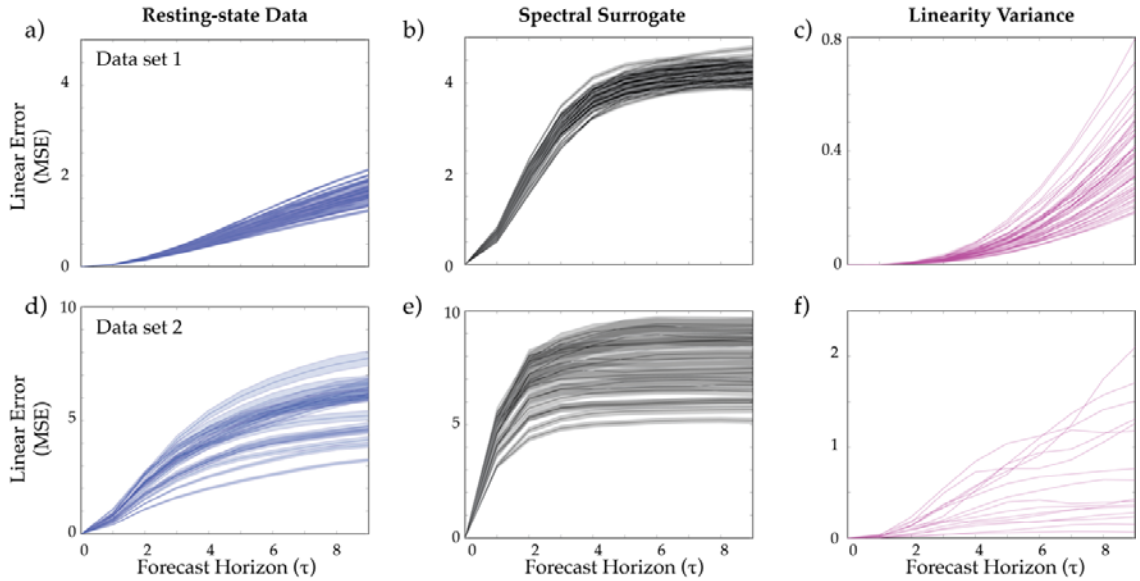
β	Rise rate of cell-body potential	769	s^{-1}
	Intra-node coupling strengths		
ν_{ee}		1.5	mV s
ν_{ei}		-3	mV s
ν_{es_c}		0.57	mV s
ν_{se}		3.4	mV s
$\nu_{s_c r}$		-1.5	mV s
$\nu_{s_c n}$		3.6	mV s
ν_{re}		0.17	mV s
ν_{rs_c}		0.05	mV s
$\tau_{es_c,es_m} + \tau_{s_c e,s_m e}$	Corticothalamic loop delay	85	ms

Table 1 – Corticothalamic neural mass parameters. Adapted from [69].

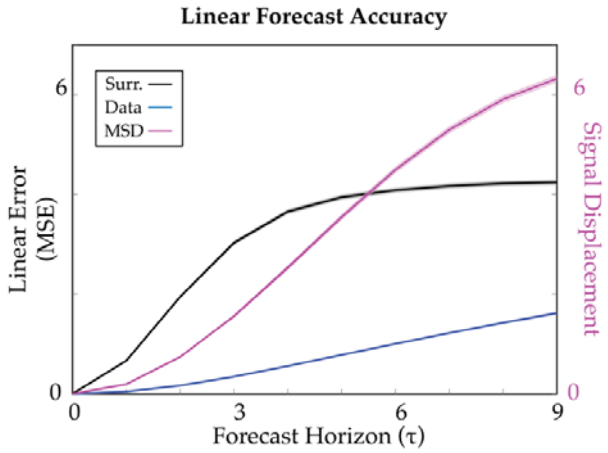
Supplementary Materials



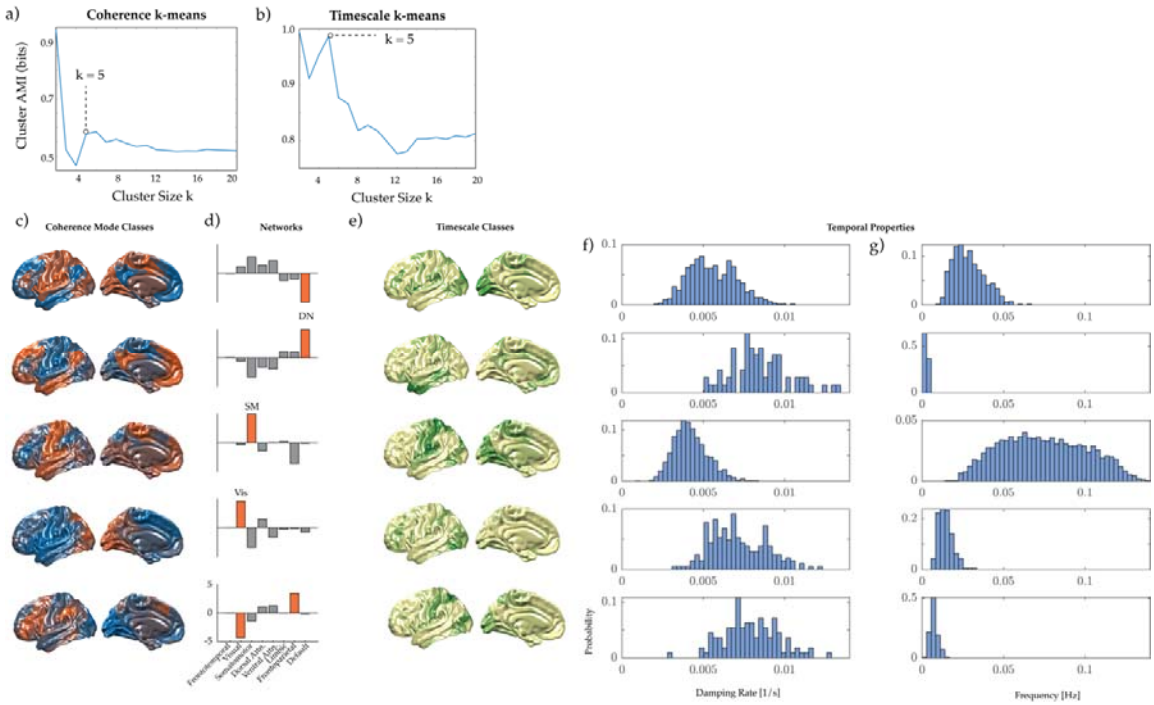
Supp. Figure 1 – Linear Forecasting and Spectral Surrogate. a) Formulation of Linear Propagator via the singular-value decomposition. b) Dynamic mode decomposition – eigenmode decomposition of linear propagator matrix resulting in oscillatory spatiotemporal modes. c) Forecast Horizon Analysis using the current state and the linear propagator matrix to predict future time points. Mean-squared-error is used to compare forecast dynamics to ground-truth empirical timeseries. d) Example spectral surrogate for a single subject showing matched covariance and Fourier power spectrum.



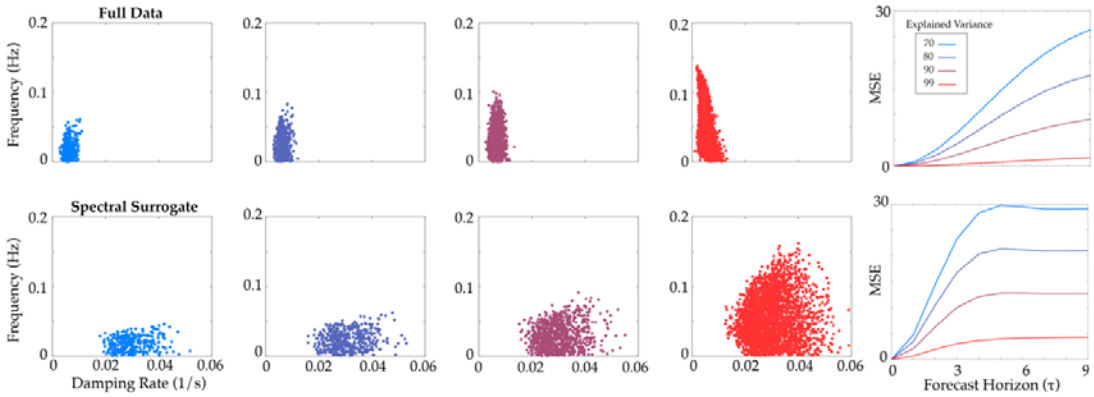
Supp. Figure 2 – LFA cross-validation. First Row: 7T human resting-state BOLD data[43] (N=59). a) Linear Error for each forecast horizon of the empirical data. b) Spectral surrogate for each subject. c) Variance of linear error (MSE) and a function of forecast horizon. Second Row: 3T human resting-state BOLD data[52] (N=14) d) Linear Error for each forecast horizon of the empirical data. e) Spectral surrogate for each subject. f) Variance of linear error (MSE) and a function of forecast horizon.



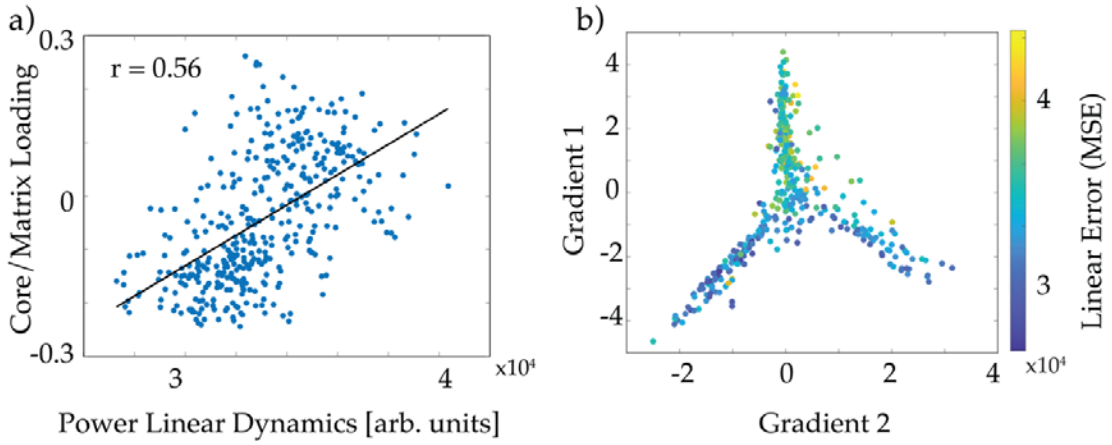
Supp. Figure 3 – Linear Model Performance. Left axis: Linear error as a function of forecast horizon for the subject average 7T resting-state human imaging data(blue), spectral surrogate(black). Right axis: Mean-squared displacement of the subject average 7T resting-state human imaging data showing general variability within the recordings.



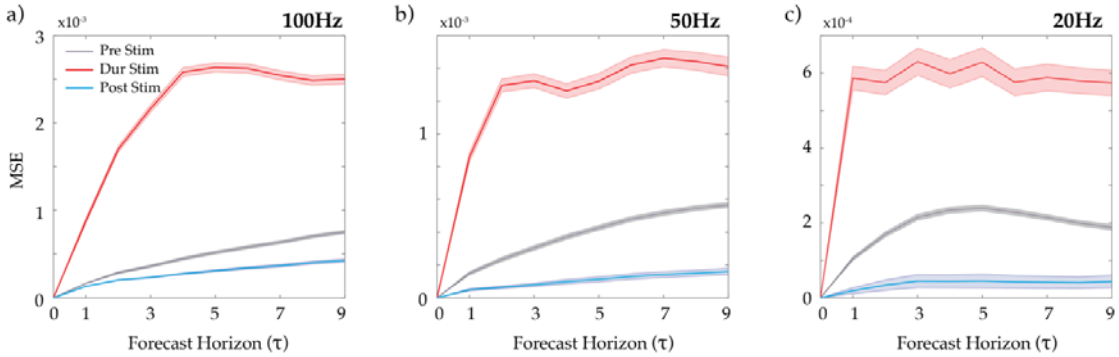
Supp. Figure 4 – K-means optimization and Dynamic mode clusters. a) Coherence map clustering: Adjusted mutual information for 100 k-means repetitions for each cluster size k. b) Eigenmode clustering: Adjusted mutual information for 100 k-means repetitions for each cluster size k. c) Coherence cluster centroids d) Coherence cluster dot-product with resting-state networks e) Average coherence maps of eigenvalue clusters f) Timescale cluster probability distribution for damping rates. g) Timescale cluster probability distribution for oscillatory frequencies.



Supp. Figure 5 – Linearity, dynamics modes and explained variance. Dynamic mode timescales and linearity for the a) full data (7T resting-state N = 59) and b) covariance and Fourier spectrum matched surrogates, as a function of the included variance explained.



Supp. Figure 6 – Linear cortical dynamics. a) Corticothalamic Core-Matrix correlation to linear cortical dynamics. b) Spatial loading of linear cortical dynamics to a low-dimensional gradient from [100].



Supp. Figure 7 – Monkey down-sampling effects. Linearity shifts of macaque multielectrode electrophysiological data pre, during, and post thalamic stimulation for a) 100Hz, b) 50Hz, and c) 20Hz sampling.

Adaptive timescales and dynamic brain modes

Intrinsic timescales of the brain have been of interest since the discovery of oscillatory dynamics in the first human EEG recordings[48]. Human neuroimaging studies have revealed a hierarchy of temporal and spatial autocorrelation scales across the cortex[49] and that these capture a large number of existing topological metrics[22]. Linear flow analysis enriches this approach by quantifying intrinsic modal

timescales that drive linear flow in neural activity. The resulting dynamic modes are defined by a complex (i.e., imaginary) eigenvector and corresponding eigenvalue[32,35,50]. Each mode contains four key characteristic properties: the spatial eigenvector defines a spatial pattern of coherent activity, and a corresponding spatial pattern of delays relative to each of these coherence patterns – called a dephasing map. In addition, each mode has a corresponding eigenvalue defining its oscillatory frequency, and an exponential gain parameter – which determines whether a mode will grow or decay in time (i.e., captures its temporal stability). The investigation of these features provides rich information regarding the types of dynamic modes that characterize the system.

In order to gain insight into how these dynamic modes relate across individual subject recordings, we leverage a two-pronged clustering approach to aggregate modes across timescales, and spatial coherence, respectively. First, we consider timescales by utilizing *k*-means clustering of the eigenspectrum collated across all subject recordings. A peak in adjusted mutual information between clustering repetitions was then used to select $k = 5$ clusters (1000 repetitions; see Supp. Fig. 4), resulting in 5 distinct timescale groupings – the corresponding average of the coherence maps within these clusters (Supp. Fig. 4). Timescale cluster 1 shows somatomotor and visual activation antiphase with frontal, parietal and temporal cortices across a broad range of frequencies – spanning much of the time resolution of BOLD recordings. These modes are also weakly damped – i.e., have a strong gain relative to the other temporal clusters. Timescale cluster 2 shows visual, premotor, and dorsal lateral prefrontal antiphase with other cortices at slower frequencies (< 0.03 Hz; Fig. g-h middle) and with a stronger damping rate, i.e., weaker gain, than cluster 1. And timescale cluster 3 shows strongly damped temporal cortex activity at frequencies < 0.02 Hz.

Next, we considered the spatial patterns of coherence, defined by each mode, aggregated across subjects and recording sessions to define consistent modal groupings[50]. Again, we leveraged *k*-means clustering (adjusted mutual information peaked between clustering repetitions for $k = 5$; See Supp. Fig. 3): Figure 2i shows 3 cluster centroids and their defining coherence patterns – i.e., the spatial patterns of the linear modes across all subjects fall within distinct classes. These classes show a significant overlap with resting-state networks used broadly in the literature[51] – including the default mode, and somatomotor and visual networks. We further find that the damping rates and frequencies of the modes within each class have comparable probability distributions – demonstrating these spatial classes did not demonstrate unique linear timescales at this granularity in resting-state.