Title
Structure, Dynamics, and Information Flow Across Brain States

Author
Toker, Daniel

Publication Date
2019

Peer reviewed|Thesis/dissertation

Structure, Dynamics, and Information Flow Across Brain States

By

Daniel Toker

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Neuroscience

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Mark D'Esposito, Chair
Professor Michael Silver
Professor Robert Knight
Professor Friedrich Sommer

Fall 2019

Abstract

Structure, Dynamics, and Information Flow Across Brain States

by

Daniel Toker

Doctor of Philosophy in Neuroscience

University of California, Berkeley

Professor Mark D'Esposito, Chair

A key challenge in neuroscience is to synthesize our understanding of neural structure, dynamics, and information processing in both health and disease. While substantial progress has been made toward such a synthesis at the microscopic scale of single neurons and their connections, significant work remains to be done at the macroscopic scale of interacting brain regions. While the field has successfully mapped the macroscale connections bridging cortical columns and regions, and has also systematically described basic features of macroscale cortical electrodynamics across perceptual, cognitive, and brain states, neuroscientists still currently lack a mathematically-specific understanding of how these macroscale networks and electrodynamics underpin large-scale neural computation and communication. Toward the end of advancing our mathematical understanding of this relationship between large-scale brain networks, dynamics, and information flow, we here present: a tool for quantifying, in bits, how much information is integrated across large-scale brain networks; a tool for tracking the presence and degree of chaos in neural electrodynamics; and evidence that macroscale cortical circuits optimize their information-carrying capacity during conscious states by operating near edge-of-chaos criticality.

Chapter 1

Introduction

Over the last few decades, neuroscientists have amassed an impressive and growing taxonomy of the macro-scale neural correlates of different cognitive, perceptual, and clinical states. In very broad terms, these correlates either describe the *structural* changes to large-scale brain networks associated with different states or conditions, or the changes to the large-scale *dynamics* associated with different states or conditions (Figure 1.1).

Many recent attempts to understand how certain structural and dynamic changes lead to changes in cognition and perception have invoked the notion of "information flow." It is assumed, for example, that the reason the brain is a modular network with integrative hubs is so that it can support specialized *information processing*, with concurrent communication of *information* between specialized modules [62]. Similarly, it is generally assumed that the cognitive deficits associated with autism spectrum disorder, for example, reflect changes to neural information processing and communication [4], and that anesthesia, deep sleep, and generalized seizures obliterate consciousness precisely because they somehow drastically disrupt neural information flow [5].

While this use of "information" as a concept to explain macro-scale neural function is consistent with the basic (and successful) view of the brain as an information processor [6], the invocation of this concept could be made far more mathematically precise. In particular, there remains significant work to be done to bridge the insights of cognitive and systems neuroscience with the insights of formal information theory.

This thesis consists of three independent projects aimed at building such a bridge between information theory and cognitive and systems neuroscience.

The first part of this thesis (Chapter 2) is aimed at alleviating one of the major roadblocks in the application of information theory to neuroscience, namely the restriction of classical information theory measures to communication between transmitter-receiver *pairs* [7] - a framework obviously unsuited to studying information flow through massive *networks* like the brain. While recent advances in information theory have derived a mathematically rigorous measure of network-wide information integration [46], this measure was previously impossible to calculate in real neural recordings. This is because a network's level of integrated information is defined by how much information it can integrate across its informational "weakest link" (analogous to defining the strength of a chain by the strength of its weakest link) - and the search space for this weakest link explodes super-exponentially with network size. The consequence of this exponential explosion is that measuring levels of information integration in real brain networks would take longer than the lifespan of the universe. To enable the calculation of integrated information in real brains, I drew on insights from the literature on functional brain networks, and showed that techniques from functional neural connectomics could solve this computational problem. This reduced the computation time for integrated information in large brain networks to just minutes. I used this solution to demonstrate that the informational weakest link of the macaque cortex splits posterior sensory areas from anterior association areas, and report evidence in favor of the long-standing but untested hypothesis that globally efficient network structures support information integration while modular network structures support information segregation. This work was published in *PLoS Computational Biology*.

While this work provided a tool for quantifying, in bits, large-scale information flow in the brain, and also elucidated the relationship between network *structure* and large-scale information flow, it said little about the relationship between neural *dynamics* and information flow. To more formally study study the relationship between dynamics and information processing across brain states, I turned to the long-standing but difficult to test proposal that dynamical chaos - i.e., exponential sensitivity to inputs - might be key to understanding large-scale neural communication [140, 145, 146, 282, 362]. While extensive work has been done relating neural chaos and information flow in simulations, very little work has been done to study this relationship in real brains, principally because chaos has historically been nearly impossible to detect from noisy time-series recordings [282]. Toward the end of studying the relationship between chaoticity and information flow in real brains, I developed a noise-robust chaos-detection tool, which significantly outperforms existing chaos-detection methods (Chapter 3). That work was recently published in *Nature Communications Biology*.

Finally, I applied this tool to cortical electrophysiology recordings to test the long-lasting hypothesis that macro-scale cortical networks maximize their information-carrying capacity by

operating near the phase transition separating chaotic from periodic (i.e. non-chaotic) dynamics [271–275, 277]. This phase transition is known as edge-of-chaos criticality. It has been known since the 1980s that diverse systems exhibit their highest information-carrying capacity and most complex information processing at this particular critical point, presumably by combining the dynamical stability of periodic systems with the sensitivity to inputs of chaotic systems. While it has long been conjectured that the healthy, waking brain operates near this critical point in order to optimize its information processing capacity, empirical evidence in favor of this hypothesis has been lacking due to the dearth of noise-robust chaos-detection tools. In Chapter 4, I use the tools described in Chapter 3 to present both simulation-based and empirical evidence that macro-scale cortical networks do in fact maintain a high information-carrying capacity during conscious states by operating near this edge-of-chaos critical point. I further provide evidence that GABAergic anesthesia reduces cortical information-carrying capacity by precipitating an excursion away from this critical point into the strongly chaotic regime, that generalized seizures likewise reduce cortical information-carrying capacity by inducing a phase transition into the periodic regime, and that psychedelics increase cortical information-carrying capacity by tuning macro-scale cortical dynamics even closer to edge-of-chaos criticality.

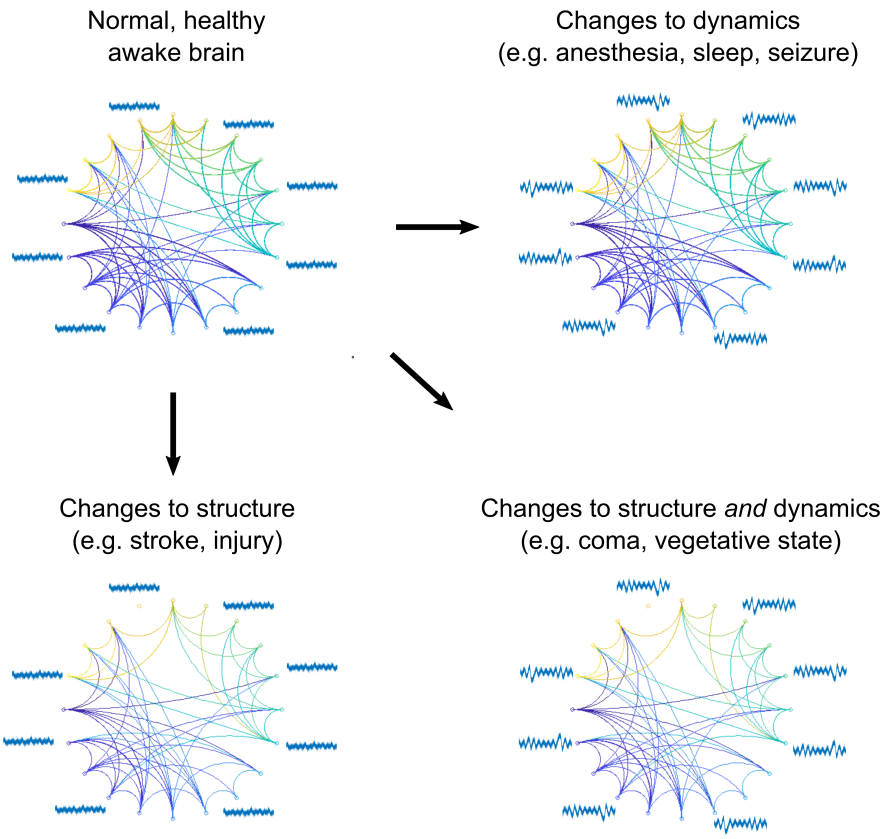


FIGURE 1.1: A key challenge in neuroscience is to understand how changes to macro-scale neural structures and dynamics affect information flow in the brain. For example, it is known that damage to well-connected brain regions following stroke or traumatic brain injury can precipitate pronounced changes to the macro-scale network structure of the brain [8], but it is unclear from an information-theoretic perspective how such structural changes affect neural computation and communication. Similarly, it is presumed that the pronounced changes to macro-scale neural dynamics observed during states like anesthesia [9], sleep [10], and seizures [11] affect neural computation and communication, but the relationship between these dynamical changes and information flow is likewise poorly understood. For many conditions, of course, structural and dynamical changes cannot be treated in isolation: conditions and states such as autism spectrum disorder [12, 13], coma [14, 15], or even healthy aging [16, 17] lead to pronounced changes to both neural structure and dynamics, and again from a mathematical perspective it is unclear how these changes affect neural information processing. The goal of this thesis is to describe computational tools and preliminary empirical results aimed at improving our mathematical understanding of the relationships between neural structure, dynamics, and information flow, so as to improve our understanding of how information processing changes or is disrupted across diverse brain states and conditions.

Chapter 2

Information Integration in Large Brain Networks

Daniel Toker and Friedrich T. Sommer

2.1 Abstract

An outstanding problem in neuroscience is to understand how information is integrated across the many modules of the brain. While classic information-theoretic measures have transformed our understanding of feedforward information processing in the brain's sensory periphery, comparable measures for information flow in the massively recurrent networks of the rest of the brain have been lacking. To address this, recent work in information theory has produced a sound measure of network-wide "integrated information," which can be estimated from time-series data. But, a computational hurdle has stymied attempts to measure large-scale information integration in real brains. Specifically, the measurement of integrated information involves a combinatorial search for the informational "weakest link" of a network, a process whose computation time explodes super-exponentially with network size. Here, we show that spectral clustering, applied on the correlation matrix of time-series data, provides an approximate but robust solution to the search for the the informational weakest link of large networks. This reduces the computation time for integrated information in large systems from longer than the lifespan of the universe to just minutes. We evaluate this solution in brain-like systems of coupled oscillators as well as in high-density electrocortigraphy data from two macaque monkeys, and show that the informational "weakest link" of the monkey cortex splits posterior sensory

areas from anterior association areas. Finally, we use our solution to provide evidence in support of the long-standing hypothesis that information integration is maximized by networks with a high global efficiency, and that modular network structures promote the segregation of information.

2.2 Introduction

Information theory, which largely measures communication between transmitter-receiver *pairs* (for e.g. a telephone sender and receiver) [30], has been key to understanding information transmission in the feedforward paths of the brain's sensory periphery [31–37]. But, traditional information-theoretic measures are of limited utility as soon as signals enter the recurrent networks that form the rest of the brain. That is because these measures are designed to quantify *feedforward* information flow. Until very recently, no theoretically sound measures were available to quantify and analyze information that is integrated by entire recurrent *networks*.

Recent work in information theory has risen to meet the challenge of quantifying the integration of information across the recurrent networks that bridge spatially distributed brain areas. Over the last decade, several measures of network-wide information integration have been proposed [38–45], which all generally define information integration as how much more information flows in a *whole* network than in the sum of its parts. The intuition can be phrased like this: if you cut a network into disconnected parts, forcing those parts to evolve over time independently of one another, how much less information is carried over time in the network? If we can estimate this difference accurately, we'd have a value - in bits - of how much information is integrated in a network.

Most of these measures of information integration have faced serious theoretical issues, such as exceeding the total information in a network, falling below 0 bits, or being impossible to estimate from time-series data [40]. To remedy this problem, mathematicians have recently derived a new, theoretically sound measure of information integration called "geometric integrated information," which is immune to the criticisms leveled against most previous measures [46, 47] (that said, we note that a mathematically similar measure called "stochastic interaction" was derived almost two decades ago [38], and that its time-reverse equivalent was recently lauded as a theoretically sound option for measuring information integration [40], but that this measure has been shown to exceed a system's total mutual information in time [43] - a criticism to which geometric integrated information is immune. We also note that there might be other sensible upper-bounds for a measure of integrated information, such as channel capacity or "effective

information," as in [48].). This means that, in principle, neuroscientists could use geometric integrated information to push past the feedforward circuits of the brain's sensory periphery, and begin to make sense of the information being integrated across the recurrently connected modules of the rest of the brain.

But there's a hitch. Calculating any of the proposed measures of information integration, including geometric integrated information, is computationally intractable for networks with more than about 20 nodes (e.g. 20 neurons or voxels). That is because all such measures of information integration require identifying what is called the "minimum information bipartition" (MIB) of a network, which is the bipartition that splits the network into two maximally independent sub-communities [38–47]. This makes measuring integrated information in large networks impossible, because finding the MIB requires a brute-force search through all possible bipartitions of a network - a combinatorial search whose computation time explodes super-exponentially with network size.

The reason we need to find the MIB is that a network's capacity for information integration is characterized by where information integration is lowest, which is very much like defining the strength of a chain by the strength of its weakest link: if one link is weak, then the whole chain is weak. For example, if a network has unconnected sub-networks, then the integrated information of that network is 0 bits. In general, to accurately determine a network's value of integrated information, one has to find the MIB of that network. Note that, in principle, the partition that yields *the* global minimum of integrated information might split a network into more than two sub-communities. But, because the number of possible n-partitions explodes with the Bell number (e.g. a network of 8 nodes can be partitioned 4,140 ways, a network of 10 nodes can be partitioned 115,975 ways, and a network of 12 nodes can be partitioned 4,213,597 ways), we follow most of the Integrated Information Theory literature [38–47] and restrict partitions to *bi*partitions, which still capture a network's overall capacity for information integration, and are at least computationally tractable for small networks. But, even with the restriction to bipartitions, the application of Integrated Information Theory is computationally challenging. As mentioned above, a brute-force search to find the bipartition that minimizes integrated information becomes computationally intractable quickly (e.g. a 20-node network can be bipartitioned 524,287 ways and a 30-node network can be bipartitioned 536,870,911 ways). Given the computational intractability of finding the MIB of large networks, our question is this: for a given set of time-series data recorded from nodes in a connected network, is there a way to approximate the minimum information bipartition without a brute-force search?

There have been several proposed solutions to this problem. In our own earlier work [49], we proposed using graph clustering to quickly find the MIB - a proposal also voiced by others

[40] - though neither we nor others have yet successfully demonstrated that graph clustering does in fact find good partitions across which to calculate integrated information. Other proposed solutions have used optimization algorithms to find the MIB [50], but these are either prohibitively slow or split brain networks into one-vs-all partitions, which do not reflect how complex biological systems are likely organized [48, 51]. Here, we build upon and empirically validate our earlier proposal that the MIB can be identified through graph clustering.

We show that a network partitioning method called "spectral clustering" [52, 53], when applied to correlation matrices of neural time-series data (Fig. 2.1), reliably identifies or approximates the MIB of even large systems. We demonstrate this in several steps. First, we show that spectral clustering can find the exact MIB in small, brain-like networks (14-16 nodes) of coupled oscillators. Then, we move onto large networks of coupled oscillators (50-300 nodes), where we forced the MIB onto the networks by structurally severing them in half, and show that spectral clustering can find good approximations of the MIB in these large oscillator networks as well. Third, we show that spectral clustering can find the exact MIB in small samples of monkey ECoG data. Fourth, we apply spectral clustering to data from all available recording sites in two monkey brains - which are so large that it would likely take centuries to determine their ground-truth MIB - and show that spectral clustering quickly finds a partition across which integrated information is smaller than or nearly equivalent to the value of integrated information across partitions identified by an optimization-based solution to this search problem (which can take weeks to run).

We note that we also tried using two other community detection algorithms, namely the Weighted Stochastic Block Model algorithm [54] and the Louvain Algorithm for modularity maximization [55], but that our early experimentation with these algorithms did not yield results nearly as strong as did spectral clustering in identifying the MIB. That said, we leave open the possibility that other community detection algorithms might approximate networks' MIBs as well as spectral clustering does.

We use our spectral clustering-based method to report two novel empirical findings: 1) The MIB of ECoG recordings in the macaque cortex splits posterior sensory areas from anterior association areas, and 2) Supporting predictions from neural connectomics research, we show that networks with a high global efficiency (i.e. a short average path length) produce high integrated information and that strongly modular networks produce low integrated information.

Because we believe that this measure will be empirically valuable for understanding how different brain states or task conditions rely on different modes of information integration between

neurons or brain regions, we have made our code publicly available as a toolbox at https://figshare.com/articles/Information_Integration_in_Large_Brain_Networks/7176557.

2.3 Results

2.3.1 Geometric Integrated Information

As mentioned in the Introduction, a number of measures of integrated information based on time-series data have been proposed. Only very recently [46, 47], a measure was derived that is at the same time computable from time-series data and properly bounded between zero bits and the total mutual information in time and space in a system. This measure, called "geometric integrated information," or Φ^{G}, is defined as the minimized Kullback-Leibler divergence between the "full model" p of a system X, which fully characterizes all the spatiotemporal influences within the system, and a "disconnected model" q. In the disconnected model, the network of interest is partitioned into statistically disconnected sub-communities, which evolve over time independently of one another:

$$q(X_t^i|X_{t-\tau}) = q(X_t^i|X_{t-\tau}^i)\ \forall i \tag{2.1}$$

where the index i labels the statistically disconnected sub-communities (so, for a bipartition, i iterates from 1 to 2), and X_t and $X_{t-\tau}$ describe present and past states of the system, respectively (t and $t-\tau$ are discrete time indices). X_t^i and $X_{t-\tau}^i$ refer to non-empty subsets (corresponding to sub-communities) of the variables constituting X_t and $X_{t-\tau}$; X_t and $X_{t-\tau}$ are n-dimensional real-valued random vectors, i.e. $X_t := (X_{t_1}, X_{t_2}, ..., X_{t_n})$, where X_{t_j} for $j = (1, ..., n)$ are real-valued random variables. In other words, for a given multivariate time-series, with n variables (e.g. neurons, electrodes, or voxels) and m time-points, $X_{t-\tau}$ is a matrix of observations of all n variables from time 1 to time $m-\tau$, and X_t is a matrix of observations of al n variables from time τ to time m. Geometric integrated information is then defined as:

$$\Phi^{\mathrm{G}} = \min_q D_{KL}[p(X_t, X_{t-\tau})||q(X_t, X_{t-\tau})] \tag{2.2}$$

where $D_{KL}[p, q]$ stands for the Kullback-Leibler divergence between two distributions p and q.

Geometric integrated information has a simple and quick-to-compute formulation for multivariate Gaussian signals [46], and all data analyzed in this paper are approximately multivariate normal (S2 Fig). (We note that for Gaussian variables no recourse to information geometry is necessary to minimize the KL divergence in Eq. 2, and so arguably there is no direct sense in

which this measure is "geometric" for Gaussian variables. That said, because the framework of information geometry is necessary for calculation of this measure in the non-Gaussian case, we follow [46] and still call this measure "geometric integrated information" in the Gaussian case). Like many information-theoretic measures, geometric integrated information can be computed in Gaussian data using the framework of linear regression. As is commonly done in time-series analysis across a range of fields, we can model the evolution in time of a Gaussian system using a simple linear regression model:

$$X_t = AX_{t-\tau} + E \tag{2.3}$$

where X_t corresponds to the present of the system and $X_{t-\tau}$ corresponds to the past of the system, A corresponds to the regression matrix estimated from the data, and E corresponds to the error or residuals in the linear regression. Both A and E can be computed from the covariance matrices of the data. The regression matrix A is given by the normal equation:

$$A = \Sigma_{X_t X_{t-\tau}} (\Sigma_{X_{t-\tau} X_{t-\tau}})^{-1} \tag{2.4}$$

where $\Sigma_{X_t X_{t-\tau}}$ is the variance between the present and the past of the system X. The covariance of the error matrix E can also be computed from the covariance of the data, and is precisely equivalent to the *conditional variance* of the present, given the past of the system:

$$\Sigma_{EE} = \Sigma_{X_t|X_{t-\tau}} \tag{2.5}$$

where

$$\Sigma_{X_t|X_{t-\tau}} = \Sigma_{X_{t-\tau} X_{t-\tau}} - A(\Sigma_{X_t X_{t-\tau}})^T \tag{2.6}$$

The covariance of E is all we'll need for the *complete* model of the system's evolution. Oizumi et al [46] prove that the disconnected model of the system can also be expressed in terms of linear regression:

$$X_t = A'X_{t-\tau} + E' \tag{2.7}$$

where A' is a regression matrix like A, but all elements describing interactions across the MIB have been set to zero (i.e. A' is a diagonal block matrix). If we have correctly identified the MIB of the network, and therefore set all the right elements of A' to zero, then the covariance of E', which is the only thing we now need to calculate integrated information, is:

$$\Sigma_{E'E'} = \Sigma_{EE} + (A - A')\Sigma_{X_{t-\tau} X_{t-\tau}} (A - A')^T \tag{2.8}$$

There is no (known) closed-form solution for A' and $\Sigma_{E'E'}$, but these matrices can be estimated

using iterative methods. In this paper, we estimate these matrices using the augmented Lagrangian method provided by [56]. Finally, we insert Eqs. 6 and 8 into the standard formula for the Kullback-Leibler divergence between two Gaussians with identical means. After a simple algebraic transformation the estimate of integrated information, in bits, can be written:

$$\Phi^{\mathrm{G}} = \frac{1}{2}\log\frac{|\Sigma_{E'E'}|}{|\Sigma_{EE}|} \tag{2.9}$$

where $|\Sigma_{E'E'}|$ refers to the determinant of the error matrix in our disconnected model, and $|\Sigma_{EE}|$ refers to the determinant of the error matrix in the connected model.

If the sub-communities of a network evolve in time mostly independently of one another, then these determinants will be close and Φ^{G} will be small. If, on the other hand, there are strong inter-dependencies between the sub-communities of a network, then these two determinants will diverge and Φ^{G} will be large.

To find the minimum information bipartition, we need to perform a brute-force search through all possible bipartitions of a network, and find the bipartition that minimizes integrated information. Unfortunately, this will usually lead to strongly asymmetrical partitions, in which one or two nodes are split from the rest of the system - and such partitions are usually of little functional relevance [40, 43–45]. While how to best handle such asymmetric partitions remains an open problem in the Integrated Information Theory literature [40, 43], there have been a number of proposed solutions for finding more balanced and functionally meaningful partitions. Here, we use the solution originally suggested in [48] and also used in [44, 45], which is to find the bipartition that minimizes integrated information, normalized by the factor K:

$$K = \min_{k}\left[H(M^k)\right] \tag{2.10}$$

where $H(M^k)$ refers to the entropy of a sub-community M^k. For a multivariate Gaussian system M, the entropy $H(M) = \frac{1}{2}\ln(|2\pi e\Sigma(M)|)$, where the bars denote the matrix determinant and $\Sigma(M)$ is the covariance matrix of the variable M. Normalized integrated information thus equals $\frac{\Phi^{\mathrm{G}}}{K}$. Minimizing the *normalized* version of integrated information biases the search toward partitions that are more balanced in the number of nodes, and away from partitions in which a single node is isolated from the rest of the network. Thus, strictly speaking, the MIB of a network is the bipartition, out of all possible bipartitions, that minimizes $\frac{\Phi^{\mathrm{G}}}{K}$, and the integrated information of that network is Φ^{G}, *not* normalized by K, across that partition. (Note that normalization was not discussed in the paper in which geometric integrated information was originally derived [46], but that it has already been shown that without normalization,

the bipartition that minimizes geometric integrated information is often the one-vs-all partition [50].)

Recall that earlier, we mentioned that a previously proposed optimization-based solution for quickly finding the MIB often splits networks into one-vs-all partitions, which are difficult to interpret in terms of biological function. This solution, proposed by [50], makes use of the Queyranne algorithm for minimizing sub-modular functions, and was shown to accurately identify bipartitions that minimize *non*-normalized integrated information. Problematically, these bipartitions are often one-vs-all splits - which is precisely what normalization was designed to avoid. Thus, finding the MIB using the Queyranne algorithm can be considered a valid option if a researcher wants to find a partition that minimizes non-normalized integrated information, as opposed to normalized integrated information. Our goal, however, is to find a quick and accurate method for identifying bipartitions that minimize *normalized* integrated informaton, because we share others' conviction [48] that this yields more biologically meaningful results.

Finally, note that Φ^{G} is calculated over a time-lag τ (Eqs. 1-8). If, for example, τ is set to 50 ms, then Φ^{G} will tell you, in bits, how much information is carried over 50 ms using the network connections that cross the MIB of your system. While the choice of a partition across which to calculate integrated information (i.e., the MIB) is well-defined, the choice of a time-lag τ is not. For the purposes of this study, we chose a time-lag that, on average, maximized integrated information for the system at hand (S3 Fig). This choice was based on previous papers [42, 57], which, based on phenomenological arguments, maintain that the time-scale of neural information integration that is most relevant to cognitive and perceptual processes is the scale that maximizes integrated information - a claim about which we are agnostic, but which our method could help elucidate in future research. That said, we note that in general, it is common to estimate time-delayed information measures such as transfer entropy for various time-lags, and then to choose the time-lag that maximizes the information measure of interest. This procedure has been shown to accurately capture the time scales of delayed system interactions [58].

2.3.2 Identifying The MIB With Graph Clustering

As a critical innovation, which enables the estimation of Φ^{G} for large networks, we propose to reduce the search space for the MIB using graph clustering on the correlation matrix of neural time-series data (Fig. 2.1). We searched the literature for a graph clustering algorithm that is biased toward balanced partitions, like the search for the MIB. We therefore chose to use spectral clustering [52] to partition our networks, because it is known to quickly find bipartitions that

approximately but robustly minimize the "normalized cut function" in graph theory, which is the sum of weights that cross a partition normalized by the sum of weights between the entire network and the communities on either side of that partition (see Methods for more details). While the normalized cut function is mathematically distinct from the function being minimized in search for the MIB (i.e., $\frac{\Phi^G}{K}$), in both cases normalization is being used to find roughly equal-sized communities, and so we hypothesized that both should yield similar partitions.

To use a network partitioning algorithm, we need a way to estimate network structure from time-series data. To address this challenge, we drew on insights from neural connectomics research. Network neuroscientists often treat the correlation matrix of neural time-series data as a "functional network" describing neural interactions, and apply graph clustering algorithms like spectral clustering to neural correlation matrices to partition the brain into distinct functional sub-networks [51, 59, 61–64, 338]. Following this insight, our method takes the correlation matrix of time-series data, transforms it using a power adjacency function (following [65]) and thresholds the transformed matrix across a range of cutoffs (following [66–70]), applies spectral clustering at each threshold, calculates Φ^G (normalized) across each resulting candidate network partition, and picks as the estimate of the MIB the partition that yields the lowest value of Φ^G (normalized). See the Methods for more details on how we used spectral clustering to approximate the MIB.

2.3.3 Spectral Clustering Finds the MIB in Small Brain-Like Networks of Coupled Oscillators

As a first step in assessing how well spectral clustering on the correlation matrix of time-series data recorded from a network can find the MIB of that network, we begin with a simulation of coupled oscillators. Among the variety of existing oscillator models, we chose to test our method in brain-like networks of coupled stochastic Rössler oscillators [71] because, when weakly coupled, their activity approximates a multivariate normal distribution [72] (S2A-C Fig, S2F-K Fig), similar to the ECoG data we analyze later in this paper (2SD-E Fig). Besides oscillators' frequency and the amplitude of noise injected into the oscillators, all parameters in the model were taken from previous literature (see Methods).

We simulated 25,000 time-points of oscillatory signals from 50 14-node networks and 50 16-node networks. These networks were generated using a novel algorithm based on Hebbian plasticity, which produces connectivity patterns that recapitulate basic features of brain connectomes, including a modular structure and rich between-module connectivity [51], and a log-normal degree distribution [73] (see Methods).

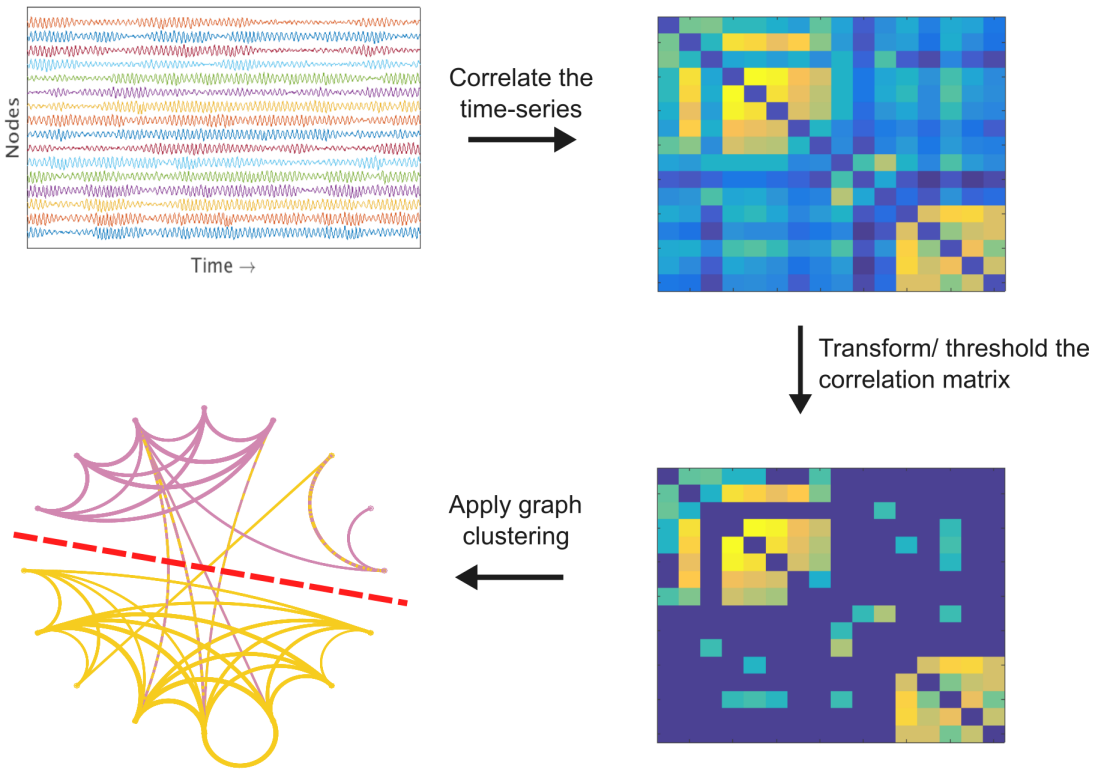


FIGURE 2.1: This is a summary of our method for approximating the minimum information bipartition (MIB) of large systems, which is necessary for calculating integrated information, without a brute-force search. We assume that the MIB of a brain network is not random, but instead is delineated by the network's functional architecture. To identify the functional architecture of brain networks from time-series data, we draw on work from functional brain connectomics, in which "functional brain networks" are often constructed by taking correlation matrices of neural time-series data, thresholding those correlation matrices to produce weighted adjacency matrices, and applying community detection algorithms like spectral clustering to those adjacency matrices. This procedure partitions the brain into functionally distinct sub-networks [64]. Our hypothesis is that the MIB of a brain network should be delineated by the functional boundaries identified through graph clustering. Out of the range of approaches to clustering brain networks, we chose spectral clustering because it is particularly well-suited for *normalized* partitioning problems, in which (just as with the search for the MIB), the goal is to find sub-networks of roughly equal size (i.e., to avoid partitioning a network into one node isolated from the rest of the network). See Methods for details on how spectral clustering was used to approximate the minimum information bipartition of brain networks.

To assess the performance of spectral clustering in identifying the MIB from time-series data, we need a best guess at the "ground truth" MIB of a system. When the underlying transition probabilities of a system are known, the ground-truth MIB can simply be determined by a brute-force search through all possible bipartitions of a system and identifying the bipartition that minimizes normalized integrated information. Identifying the ground-truth from time-series data, however, requires infinite observations. Thus, when we refer to the "ground-truth" MIB throughout this paper, we simply mean the bipartition, identified through a brute-force search through all possible bipartitions, that minimizes an estimate of normalized integrated information from finite observational data.

We found that in 95/100 of our small simulated networks, there was a difference of 0 bits between Φ^{G} (normalized) across the spectral clustering-based bipartition and the lowest value of Φ^{G} (normalized) identified through a brute-force search through all possible bipartitions (Fig. 2.2a). In other words, in almost all networks tested, our spectral clustering-based approach gives the exact same result as does a brute-force search for the MIB. We further found that the Rand Index [74] (a common measure of partition similarity) between the ground-truth MIB and the spectral bipartition was 1 (indicating a perfect match) for those same 95 networks (Fig. 2.2b). Finding partitions that are highly similar to the MIB in these networks is important, since the more dissimilar a partition is from the MIB, the larger Φ^{G} (normalized) will tend to be across that partition; in other words, the further off you are from the MIB, the less accurate your estimate of integrated information will tend to be (S4A-B Fig). To test the statistical stability of these results, we computed running averages of both the Rand Indices and the differences between estimated Φ^{G} values (e.g. the running mean Rand Index of the first two 14-node networks, then the first three 14-node networks, then the first four 14-node networks, etc.). We then took the approximate derivatives of the running averages for both network sizes, and used two-sample t-tests to accept the null hypothesis (α=0.05) that the approximate derivatives were indistinguishable from 0 for both tests, for both network sizes. This means that the results reported in Fig. 2.2 are statistically stable at a sample size of 50 networks (i.e. adding more samples would not likely change the means significantly, as the differences in the running average are already approximately zero at just 50 networks). To further check whether this result generalizes across different network dynamics, we used the same networks to generate multivariate autoregressive simulations and performed the exact same analysis, and found that spectral clustering also accurately identifies the MIB for autoregressive data (S7 Fig). We used the same running average and approximate derivative test to confirm that our results for the autoregressive dynamics in S7 Fig are also statistically stable at a sample size of 50 networks.

Finally, we also compared our approach to another proposed method for quickly identifying the

MIB from time-series data. This method uses the Queyranne algorithm for fast minimization of sub-modular functions [50]. Though in past work the Queyranne algorithm has been successfully used to minimize non-normalized integrated information, we used the Queyranne algorithm to try to find a bipartition that minimizes *normalized* integrated information. The difference between Φ^{G} (normalized) across the Queyranne bipartition and Φ^{G} (normalized) across the MIB was 0 bits (indicating a perfect match) in only 1/50 14-node networks (mean difference = 0.0031 bits) and in 2/50 16-node networks (mean difference = 0.0026 bits). The Rand Index between the Queyranne partition and the MIB was 1 for the same networks for which the difference in Φ^{G} (normalized) was 0; the mean Rand Index was 0.576 across all 14-node networks and 0.582 across all 16-node networks. The Queyranne algorithm also performed poorly in minimizing normalized integrated information in autoregressive simulations generated from these same small brain-like networks (S7 Fig). Thus, spectral clustering does a better job of estimating the MIB in small brain-like networks than does the Queyranne algorithm. Moreover, even when trying to minimize normalized integrated information, which is biased toward balanced partitions, the Queyranne algorithm often found partitions that isolate one node from the rest of the network. This occurred in 23/50 of the 14-node networks (while none of the MIBs identified through a brute-force search yielded one-vs-all partitions) and in 26/50 of the 16-node networks (while only one of the MIBs identified through a brute-force search was a one-vs-all partition). Such partitions are usually of little functional relevance - hence why normalization is introduced in searching for the MIB [48]. Moreover, the partitions found by the Queyranne algorithm were also generally dissimilar from the partitions found by our spectral clustering approach: the mean Rand index between the spectral partitions and the Queyranne algorithm partitions was 0.57 for the 14-node networks and 0.59 for the 16-node networks.

2.3.4 Spectral Clustering Approximates the MIB in Large, *Cut* Brain-Like Networks of Coupled Oscillators

Having passed this basic test in small networks, we next asked whether spectral clustering can accurately identify the MIB in large systems. To test this, we used our algorithm for generating brain-like connectivity (see Methods) to create networks which ranged from 50 to 300 nodes in size. Networks of these sizes cannot be exhaustively searched for their MIB, so we forced the MIB onto these networks by cutting them in half. If a network is cut into two parts, then, with infinite data, the MIB will converge onto where the network has been cut and Φ^{G} across this cut will be 0 bits. For these networks, we generated 100,000 time-points of data using the stochastic Rössler oscillator model, since in larger systems more data are necessary for more accurate estimation of multivariate information measures. We were unable to test the accuracy

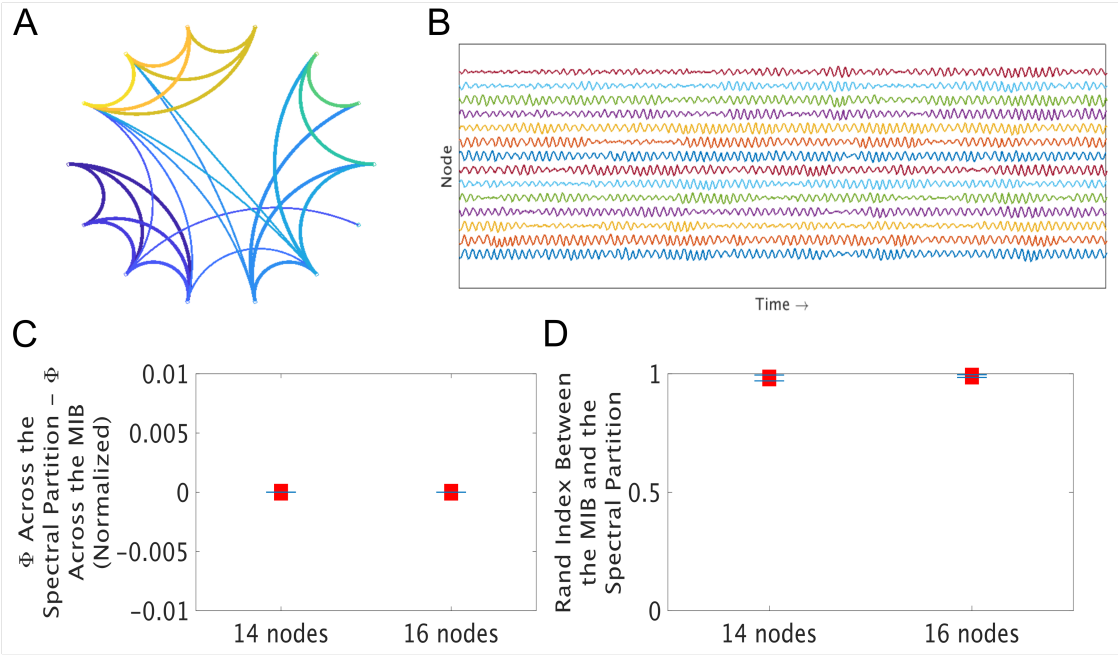


FIGURE 2.2: We first tested our spectral clustering-based approach in small simulations. **A** This is an example of a small brain-like network we generated using a novel algorithm based on Hebbian plasticity. This algorithm produces networks that are loosely brain-like, in that they are modular, show rich cross-module connectivity, and display a log-normal degree distribution with long right tails. We used this algorithm to generate 50 14- and 16-node networks. See Methods for more details on network generation. **B** This is a sample of oscillatory data generated from the network in **A**. We generated these data using a stochastic coupled Rössler oscillator model. In the Rössler oscillator model, each node stochastically oscillates according to its own intrinsic frequency, and dynamically synchronizes with other nodes it is connected to. The resulting data are multivariate normal (S2 Fig), allowing for the fast computation of integrated information. **C** As a first test of our spectral clustering-based approach to identifying the MIB from time-series data, we subtracted Φ^{G} (normalized) across the ground-truth MIB, identified through a brute-force search through all possible bipartitions, from Φ^{G} (normalized) across the partitions identified through spectral clustering. In this test, a perfect match between values would yield a difference of 0 bits. Red squares indicate the mean across 50 networks, and the blue bars indicate standard error of the mean. **D** As a second test of our spectral clustering-based approach, we computed the Rand Index [74], which is a common measure of partition similarity, between the spectral partitions and the ground-truth MIBs of these networks. A Rand Index of 1 indicates a perfect match between partitions, and a Rand Index of 0 indicates maximum dissimilarity between partitions. Red squares indicate the mean across 50 networks, and the blue bars indicate standard error of the mean. These results show that spectral clustering finds the MIB of small networks of coupled oscillators. We found similar results using the same networks but different network dynamics (S7 Fig).

of the Queyranne algorithm for these networks, because the computation time for using the algorithm to minimize normalized integrated information increased exponentially, making its application to networks with more than 50 nodes prohibitively expensive; that said, we note that the algorithm is far faster in minimizing *non*-normalized integrated information, as shown

in [50].

Spectral clustering again performed remarkably well. The mean absolute difference between Φ^{G} across the spectral partition and Φ^{G} across the ground-truth cut was less than 0.001 bits (normalized) for all network sizes (Fig. 2.3A), indicating a close match. Note that, objectively, Φ^{G} should be zero in these cut networks, and we would expect estimates of Φ^{G} to converge to zero bits with infinite data; as a sanity check, we utilized a well-established method for extrapolating estimates of information measures to what they would be if infinite data were available, and found that this brought estimates significantly closer to zero bits for these cut networks, as expected (S1 Fig). The Rand Index between the spectral partition and the ground-truth cut was greater than .8 for 37/40 of the 50-node networks, 39/40 of the 100-node networks, 29/40 of the 150-node networks, 21/40 of the 200-node networks, 18/40 of the 250-node networks, and 10/40 of the 300-node networks. Given that the estimates of Φ^{G} across the spectral partition and the ground-truth cuts were very close even in the 200- to 300-node networks (for which the spectral partitions were similar to the ground-truth cut less often) and also both extrapolated to around the ground-truth of zero bits, these results suggest that there are sometimes multiple minima for normalized integrated information (i.e. in these cut networks, there are sometimes several bipartitions across which there is little to no information integration). To test the statistical stability of these results, we computed running averages of both the Rand Indices and the differences between estimated Φ^{G} values (e.g. the running mean Rand Index of the first two 50-node networks, then the first three 50-node networks, then the first four 50-node networks, etc.). We then took the approximate derivatives of the running averages for each network size, and used two-sample t-tests to confirm that the approximate derivatives were statistically indistinguishable from 0 for both tests, for each network size. This means that the results reported in Fig. 2.3 are statistically stable at a sample size of 40 networks (i.e. adding more samples would not likely change the means significantly). Finally, we again checked whether this result generalizes across different network dynamics, by generating autoregresive simulated data from these large, cut networks. We found that spectral clustering performed even better (nearly perfectly) for the autoregressive simulations (S8 Fig), again supporting the robustness and generalizability of our method. We again used a running mean of the results, together with approximate derivatives, to confirm that the results for the autoregressive data in S8 Fig were also statistically stable at a sample size of 40 networks.

2.3.5 Spectral Clustering Approximates the MIB in the Macaque Cortex

We next applied the same spectral clustering method to one minute of ECoG data from two macaque monkeys, Chibi and George [373]. After pre-processing (see Methods), data for 125 electrodes distributed across the left cortex of each monkey were available. These data were multivariate normal (S2D-E Fig). To enable comparison between graph clustering-based partitions and the ground-truth MIB, we divided these data into overlapping sets of fourteen electrodes each, resulting in 112 sets of electrodes for each monkey. The difference between Φ^{G} across the MIB and Φ^{G} across the partitions identified by spectral clustering was 0 (indicating a perfect match) for 46/112 of the datasets from Chibi's brain (mean difference = 0.0001 bits) and in 67/112 of the datasets from George's brain (mean difference = 0.0002 bits) (Fig. 2.4A). The Rand Index comparing the spectral partition and MIB was 1 for those same datasets (Chibi mean Rand Index = 0.79, George mean Rand Index = 0.87) (Fig. 2.4B). As was the case for our simulated networks, the more dissimilar partitions in the monkeys' brains were from the MIB, the larger Φ^{G} (normalized) tended to be across those partitions (S4 Fig). The Queyranne algorithm again performed worse than spectral clustering, yielding perfect matches to the ground-truth in only 18/112 of the datasets from Chibi's brain (mean Rand Index = 0.6) and 22/112 from George's brain (mean Rand Index = 0.64). Moreover, as was the case for our simulated data, the Queyranne algorithm separated one node from the rest of the system in the majority (145/224) of all ECoG datasets (as opposed to the ground-truth MIBs, which separated one node from the rest of the system in only 39/224 datasets). Finally, the partitions found by the two algorithms were generally dissimilar: the mean Rand index between the spectral partitions and the Queyranne algorithm partitions was 0.65 for the electrode clusters in Chibi's brain and 0.67 for George's brain.

As a test of how well spectral clustering could approximate the MIB for all electrodes, we asked whether it could minimize Φ^{G} (normalized) in the whole cortex of each monkey. We therefore calculated Φ^{G} across the spectral clustering-based bipartition of the entire left cortex for both monkeys. We found that this estimate of the MIB split posterior sensory areas from anterior association areas in both brains (Figs. 2.4C, 2.4E). To test the statistical robustness of this result, we compared both our estimated Φ^{G} (normalized) values and our estimated MIBs for both monkey cortices to results from 100 Amplitude Adjusted Fourier Transform surrogate datasets [165]; we found that our estimated Φ^{G} (normalized) values were significantly higher than the surrogate distributions for both monkeys, and that the similarities between the MIBs estimated for the monkey cortices and the MIBs estimated for the surrogate datasets were at chance levels, suggesting that the results for the full monkey brains are not artifactual (S9 Fig). We then compared Φ^{G} across the spectral clustering-based partitions to Φ^{G} values calculated

across partitions identified by a Replica Exchange Markov Chain Monte Carlo (REMCMC) algorithm. The REMCMC method for estimating the MIB is described in detail in [50]; the algorithm used in this paper is the same as that used in [50], except that it searched for a bipartition that minimized normalized (rather than non-normalized) integrated information. We also terminated the algorithm after 10 days, since it failed to reach convergence for either monkey dataset by that point. Since the algorithm tries to minimize normalized integrated information across six parallel sequences, it produces six guesses for the MIB. We also tried using the Queyranne algorithm for the monkey brains, but the algorithm failed to terminate even after two weeks of running, and so we did not include the Queyranne algorithm in this analysis.

For George's brain, normalized integrated information across the spectral clustering-based partition was lower than it was across all six bipartitions identified by the REMCMC method (Fig. 2.4E). In Chibi's brain, the REMCMC algorithm found two partitions across which normalized integrated information was very slightly lower (0.0002 bits) than it was across the spectral clustering-based partition; interestingly, the two REMCMC partitions (which yielded the same value of normalized integrated information) were not only dissimilar to each other (Rand Index = 0.5), but were also both dissimilar to the spectral clustering-based partition (Rand Indices = 0.5, 0.55), suggesting that there were several local minima of normalized integrated information for Chibi's brain. In all, these results show that our spectral clustering-based method reliably minimizes Φ^{G} (normalized) of the entire macaque cortex, suggesting that it successfully finds or approximates the MIB in large neural data.

2.3.6 Network Structure and Information Integration

The ability to quickly measure information integration in large networks allowed us to assess what network architectures best support information integration, and what that might imply about how brains could be organized to integrate information. We here test for the first time, *in silico*, several graph-theoretic measures that have been hypothesized to track neural information integration. Note that in the neural connectomics literature, these graph-theoretic measures are often applied to either *structural* networks, such as the physical connectivity between brain regions that might be revealed through diffusion tractography, or to *functional* networks, such as correlation matrices calculated from functional magnetic resonance imaging recordings [77]. Because analyses of structural networks are more straightforward than analyses of functional networks (primarily because there is considerable debate surrounding what constitutes a functional network), we here focus on the relationships between structural networks and integrated

information. We hope to more systematically investigate the relationship between integrated information and *functional* networks in future work.

The most commonly invoked graph-theoretic measure of a network's capacity to integrate information is *global efficiency* [66, 77–80]. Global efficiency is related to the inverse of the average shortest path between nodes in a network. Formally, the global efficiency E of a network G is defined as follows:

$$E(G) = \frac{1}{n(n-1)} \sum_{i \neq j \in G} \frac{1}{d(i,j)} \tag{2.11}$$

where n is the number of nodes in the network and $d(i,j)$ is the shortest path between given network nodes i and j. In high efficiency networks, any node can be reached by any other node with only a few steps. For about a decade, network neuroscientists have assumed that the global efficiency of a brain network quantifies its ability to concurrently exchange information between its spatially distributed parts; for this reason, it has been assumed that global efficiency sets an upper limit on neural information integration [66, 77, 79, 80].

Conversely, it has been assumed that the modularity of brain networks (and of complex networks more generally) *limits* the integration of information, primarily by segregating network dynamics [51, 62, 80]. The modularity of a network is defined by Newman's Q:

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i k_j}{2m}\right] \delta(c_i, c_j) \tag{2.12}$$

where A_{ij} is the adjacency between nodes i and j, k_i and k_j are the sums of the adjacencies involving i and j, respectively, c_i and c_j are the modules to which nodes i and j have been assigned, respectively, $m = \frac{1}{2}\sum_{ij} A_{ij}$, and $\delta(c_i, c_j)$ equals 1 if $c_i = c_j$ and 0 otherwise. Networks that can be easily subdivided into distinct sub-communities or modules will have a high Q, whereas networks with little community structure (such as random networks) will have a low Q. We used the Brain Connectivity Toolbox's [80] `modularity_und.m` function, which implements Newman's spectral community detection algorithm [81], to compute network modularity.

To directly study the relationship between network efficiency, modularity, and integrated information, we followed the network generation procedure introduced by Watts and Strogatz in their canonical paper on small-world networks [82]. In their paper, Watts and Strogatz begin with completely regular lattice networks, in which nodes are only connected to their neighbors; they then systematically increase a parameter p, which is the probability that a given node will re-wire a local connection and connect to any random node in the network. A p of 0 yields a

completely regular lattice network, a p of 1 yields a completely random network, and intermediate values of p yield "small-world" networks, which are highly clustered like regular lattice networks but also have short characteristic path lengths like random networks (Fig. 2.5A). The parameter p also systematically controls the global efficiency of the network: higher values of p produce networks with higher global efficiency [78] (Fig. 2.5B). We also show that p systematically decreases network modularity (Fig. 2.5C).

Since up until this point we have only shown that our spectral clustering-based approach can find the MIB of brain-like networks of coupled oscillators, autoregressive signals generated from brain-like networks, and in real brain data, we first checked whether spectral clustering can also find the MIB in small lattice networks, small-world networks, and random networks of coupled oscillators. Consistent with our earlier results, we found that spectral clustering found the exact MIB (determined through a brute-force search) in almost all 14- and 16-node Rössler oscillator networks of these types that we tested (S6 Fig). As such, we felt confident that it would also give us accurate estimates of integrated information in large networks of these types. We therefore iterated through 19 values of p: the first 10 values were logarithmically spaced between 0.001 and 0.1 (following [82]), and the following nine values were linearly spaced between 0.1 and 1. For each value of p, we created 50 100-node networks, which all had the same number of edges and a mean degree of 6, and ran the Rössler oscillator model on those networks to produce 25,000 time-points of oscillatory signals. To ensure that any differences in integrated information in the resulting network dynamics were attributable to network connectivity rather than coupling strength, we set the oscillators' coupling parameter to 0.25 for all networks in this analysis (rather than determine the coupling strength through a master stability function, as we do elsewhere - see Methods).

We found that, as predicted by work in neural connectomics [66, 77, 79, 80], networks' global efficiency was tightly coupled to their capacity for information integration. Increasing the rewiring probability p systematically increased both a network's global efficiency (Fig. 2.5B) and how many bits of information are integrated across that network (Fig. 2.5D), and decreased the networks' structural modularity. Interestingly, both global efficiency and integrated information reach a plateau around $p = 0.4$, though it is unclear from our present results why this is the case. Finally, when looking across all networks, there was a strong and significant correlation (r=0.91, $p< 10^{-324}$) between the networks' global efficiency and how much information they integrate (Fig. 2.5D) and a strong and significant anti-correlation (r=-0.90, $p< 10^{-324}$) between the networks' structural modularity and how much information they integrate (Fig. 2.5E). This supports the widely held hypothesis that global efficiency determines how many bits of information a network can integrate and that modularity limits information integration, at least in

the case of coupled oscillator networks. It would be interesting to see whether this relationship between network efficiency and integrated information extends to systems with non-Gaussian dynamics - a possibility we hope to explore in future work.

2.3.7 Run Time Analysis

The results reported thus far show that our spectral clustering-based approach can accurately approximate the MIB of a system from time-series data. As a final analysis, we show that it is also much faster to run than either a brute-force search or the Queyranne algorithm for large systems, since its run time scales much less steeply (Fig. 2.6). We simulated 25,000 time points of data using our Rössler oscillator model and artificial brain-like networks (see Methods) ranging from 10 to 120 nodes in size. We estimated integrated information using a brute-force search for the MIB in the 10- to 18-node networks, used the Queyranne algorithm for networks of 10- to 50-nodes in size, and used our spectral clustering approach for all network sizes. We empirically measured how long it took to run each of these algorithms on Matlab, using a 64-bit linux CentOS. In Fig. 2.6 we plot the average run time across five samples of each network size. We found that, as expected, the run time for the brute-force search for the MIB scales super-exponentially; we further found that the run time for our approach scales much less steeply than does the run time for the Queyranne algorithm, which means that our method is not only more accurate than the Queyranne algorithm in finding bipartitions that minimize normalized integrated information, but is also much faster for large systems. That said, we again emphasize that the Queyranne algorithm is a valid and fast option for minimizing *non*-normalized integrated information [50].

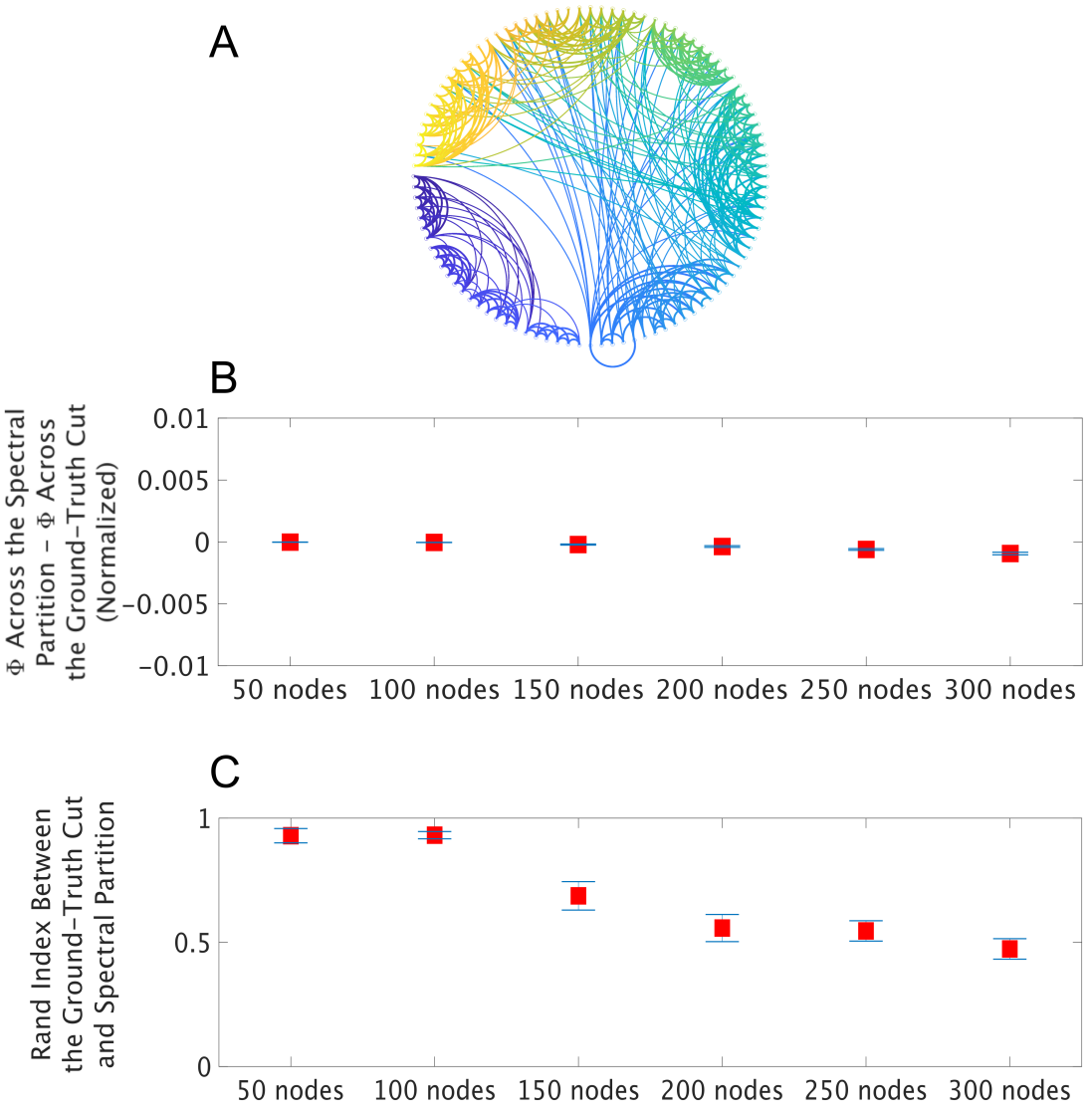


FIGURE 2.3: **A** Having shown that spectral clustering can find the MIB in time-series data from small networks, we next asked whether it could find the MIB of large simulated networks. While large networks cannot be exhaustively searched for their MIB, the MIB can be forced onto them by cutting them in half. We generated 40 such cut networks for each network size. Network sizes ranged from 50 nodes to 300 nodes. **B** Here, we show Φ^{G} (normalized) across the ground-truth cut subtracted from Φ^{G} (normalized) across the partition identified through spectral clustering. Red squares indicate the mean across 40 networks, the absolute value of which never exceeded 0.001 bits (normalized), and the blue bars indicate standard error of the mean. **C** Here, we show the mean and standard error of the Rand Index between the ground-truth cut and the spectral clustering-based partition of the correlation matrix estimated from each network. The Rand Index between the spectral partition and the ground-truth cut was greater than 0.8 (indicating high similarity) for the majority of networks of all network sizes, except for the 200- to 300-node networks. Despite this dip in Rand Index, spectral clustering still found partitions across which Φ^{G} (normalized) was extremely close to Φ^{G} (normalized) across the ground-truth cut in the 300-node networks (**A**), which suggests that in these networks, there was sometimes several possible partitions that minimized normalized integrated information.

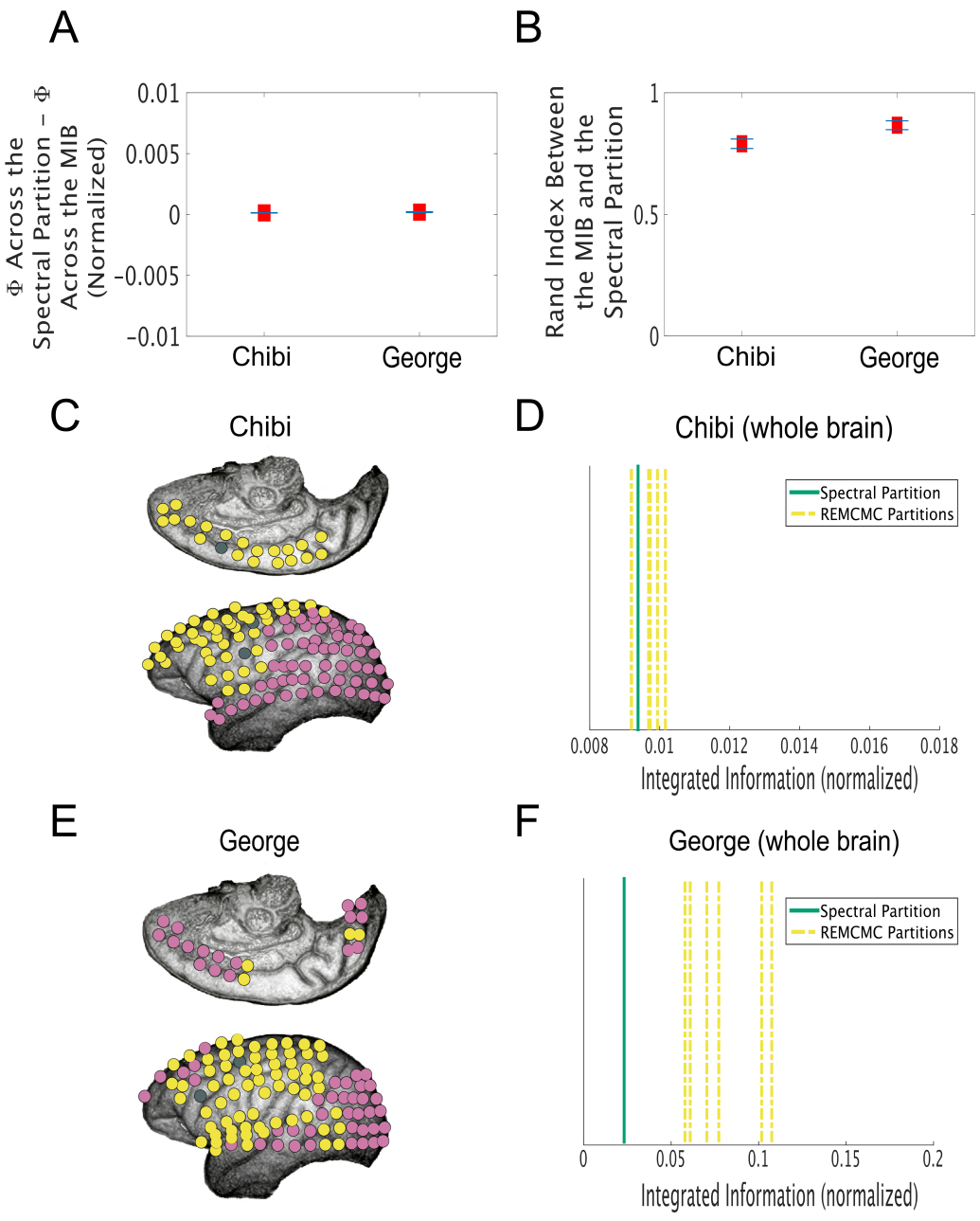


FIGURE 2.4: Caption on the following page.

(continued) A We split the available ECoG electrodes in two macaque monkeys into overlapping sets of 14 electrodes. The ground-truth MIB of 14 electrodes can be identified through a brute-force search, and compared to the spectral partition estimated from the correlation matrix of data from those electrodes. Here, we subtracted Φ^{G} (normalized) across the ground-truth MIB from Φ^{G} (normalized) across the spectral partition. There was a difference of 0 bits for 67/112 (mean difference=0.0002 bits) datasets from George's brain, and a difference of 0 bits in 46/112 (mean difference=0.0001 bits) datasets from Chibi's brain. Red squares indicate the mean difference in Φ^{G} (normalized) across all datasets from one brain, and blue bars indicate standard error of the mean. **B** Spectral clustering found the exact MIB for the same 67/112 datasets in George's brain (mean Rand Index=0.87) and 46/112 datasets in Chibi's brain (mean Rand Index=0.79). **C** We used our spectral clustering approach to estimate the MIB of Chibi's entire left cortex, and found that it split posterior sensory areas from anterior association areas. Electrodes are colored according to the community in which they are clustered; the electrodes that were excluded from the analysis because they displayed consistent artifacts are colored grey. **D** Φ^{G} (normalized) across the spectral partition of Chibi's left cortex (solid green line) was lower than it was across 4/6 partitions identified by the Replica Exchange Markov Chain Monte Carlo (REMCMC) method (yellow dashed lines) [50]. The other 2/6 partitions yielded values of normalized integrated information that were very slightly lower (0.0002 bits) than the value across the spectral clustering-based partition, and were dissimilar both to each other (Rand Index=0.5) and to the spectral partition (Rand Indices=0.5, 0.55), suggesting that there were several local minima of normalized integrated information in Chibi's brain. We ran the algorithm for 10 days. **E** Our estimate of the MIB of George's left cortex using spectral clustering also (largely) split posterior sensory areas from anterior association areas. **F** Φ^{G} (normalized) across the spectral partition of George's left cortex was lower than it was across all bipartitions identified by the REMCMC method. Note the difference in scale on the x-axes of **D** and **F**; it is unclear why this scale should differ between the two brains.

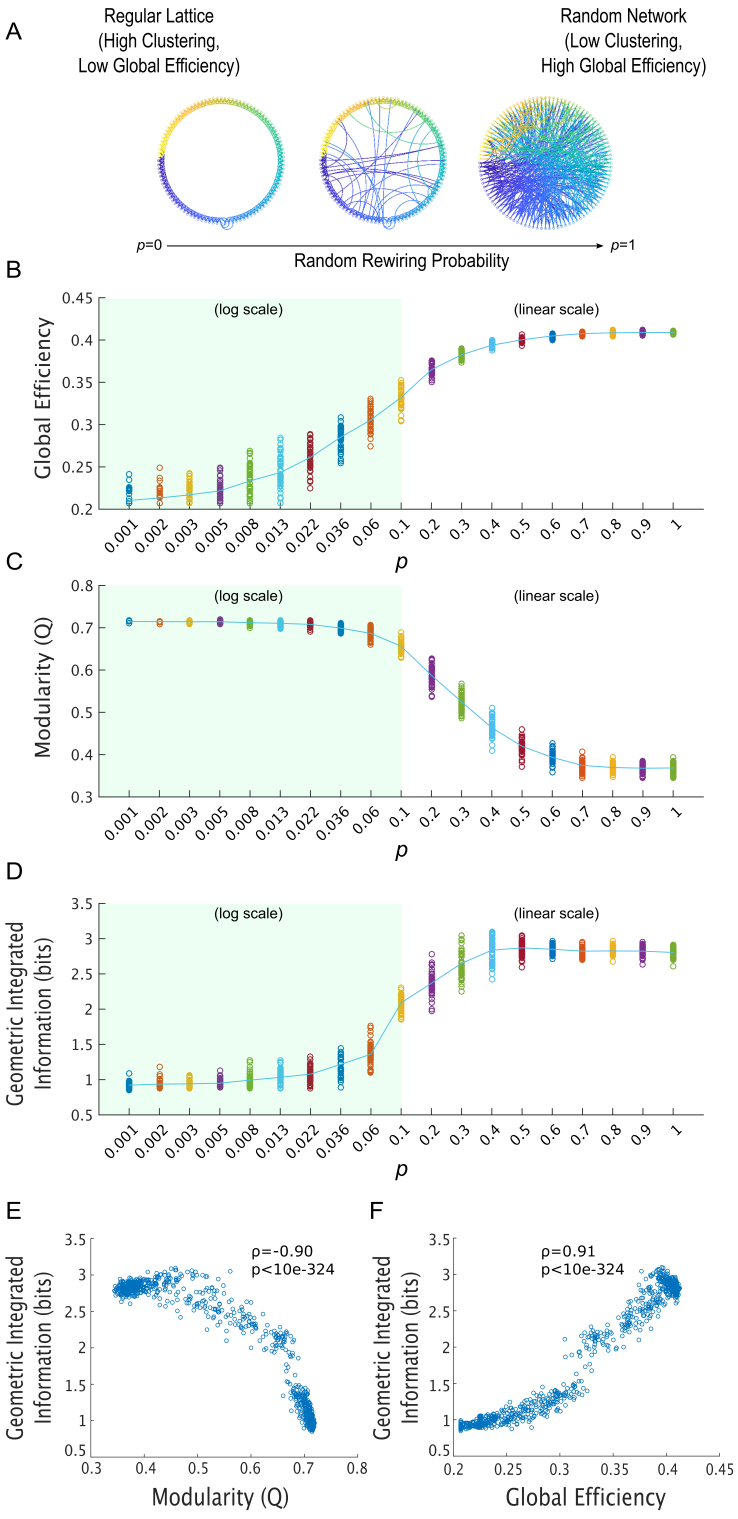


FIGURE 2.5: Caption on the following page.

(continued) The method presented in this paper for quickly identifying a network's MIB using spectral clustering makes it possible to quickly measure integrated information in large brain networks. A straightforward first-pass at an application for our method is to evaluate the long-held and untested assumptions that the "global efficiency" of a network reflects its capacity for information integration and that the modularity of a network underpins the segregation of information. **A** Following the procedure introduced by Watts and Strogatz [82], we systematically increased the global efficiency of our networks by increasing their rewiring probability p. Following Watts and Strogatz [82], we varied p on a log-scale between 0.001 and 0.1; to explore the full parameter space, we also linearly varied p between 0.1 and 1. For each value of p, we generated 50 100-node networks, and generated time-series data for each of those networks using the stochastic Rössler oscillator model. We then used our spectral clustering-based technique to measure geometric integrated information in these networks. **B** As expected [78], increasing p increased the global efficiency of the networks. Here, each dot corresponds to the global efficiency of one network of coupled Rössler oscillators with that particular value of p. The green line passes through the mean across networks. **C** Increasing p also systematically *decreased* the modularity Q of the networks. **D** A higher probability p of forming long-distance network connections, which increases global efficiency, led to higher integrated information (*non*-normalized). **E** There was a strong negative correlation between the networks' structural modularity and how much information they integrate, in bits (Spearman's ρ=-0.90, p$< 10^{-324}$). Note that the gap around $Q = 0.65$ occurs at the transition from the log variance of p to the linear variance of p (**C**). **F** There was a strong positive correlation between the networks' global efficiency and how much information they integrate, in bits (Spearman's ρ=0.91, p$< 10^{-324}$). Note that the gap around $E = 0.32$ occurs at the transition from the log variance of p to the linear variance of p (**B**). These results support the hypothesis that network modularity supports the segregation of information, while global efficiency supports the integration of information.

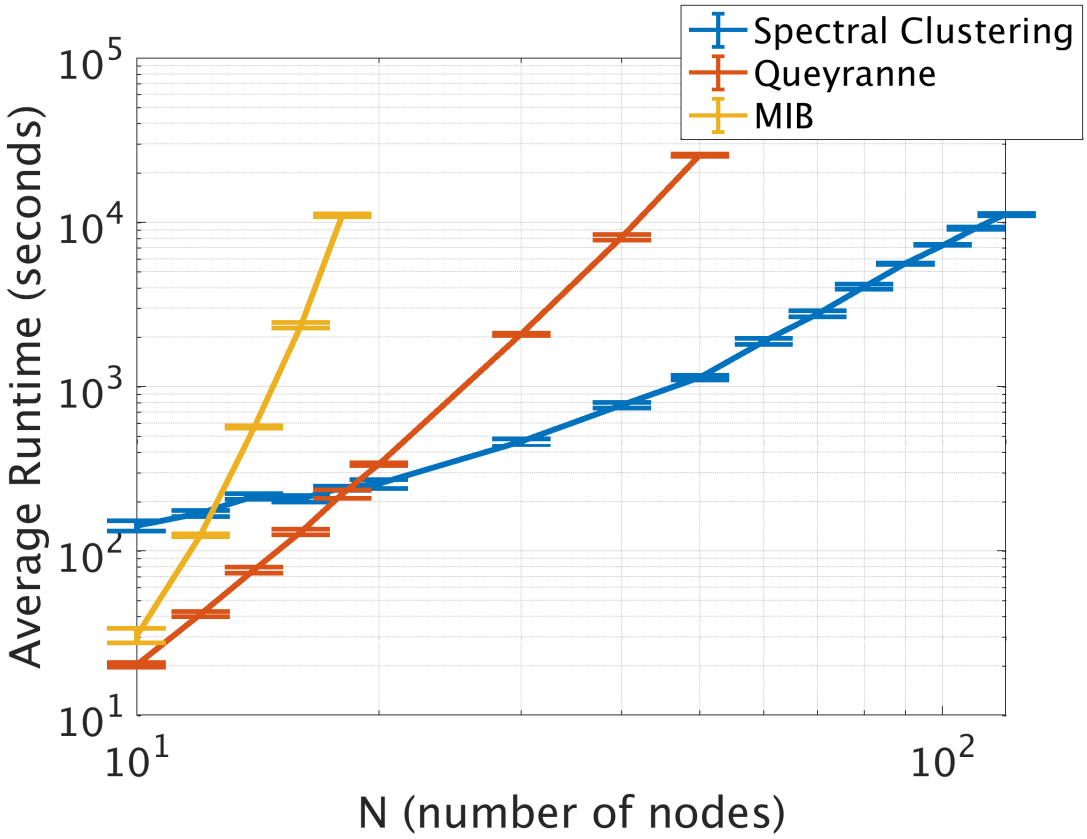


FIGURE 2.6: Average run time across for the three algorithms as a function of network size. Error bars indicate standard error of the mean across five networks of coupled oscillators of a given size. For very small brain-like networks (10-14 nodes), our spectral clustering-based approach is slower than either the Queyranne algorithm or a brute-force search for the MIB. This is because our algorithm searches through a fixed number of candidate graph cuts (see Methods). But, this feature is also the algorithm's strength: because our algorithm searches through the same number of candidate partitions for large systems as it does for small systems, its computation time scales much less steeply than that of the other two algorithms. If our algorithm were to search through more partitions (for e.g. by iterating through more threshold values of the correlation matrices - see Methods), then it would be slower, but its run time would still scale far less steeply than the other two algorithms, because the number of candidate partitions would remain fixed.

2.4 Discussion

We have presented in this paper a method for measuring integrated information in large systems, using time-series observations from those systems. Specifically, we presented a robust approximate solution to the search for the minimum information bipartition of large networks, a problem that has impeded efforts to measure integrated information in large brain networks. Our proposed method for quickly partitioning brain networks to find the MIB is drawn from well-established methods in neuroimaging (for a recent review of the use of graph clustering on neural correlation matrices to identify functional sub-networks of the brain, see [51], and for the specific use of spectral clustering in such analyses, see [64, 67]). Although the Queyranne algorithm has previously been shown to successfully find bipartitions that minimize *non*-normalized integrated information [50], the algorithm usually finds one-vs-all network partitions, even when trying to find a partition that minimizes *normalized* integrated information (as we report here). That said, we agree with Kitazono and colleagues [50] that it would be fruitful to consider methods that combine our spectral clustering-based approach with their Queyranne algorithm-based approach.

It is worth pointing out that although spectral clustering found the MIB or partitions close to the MIB in the majority of both real and simulated signals for which the ground-truth MIB could be computed, it did not always yield perfect results. While it is still unclear what conditions ensure that spectral clustering will find the exact MIB, we note that in the analyses performed here, the performance of spectral clustering was correlated with the strength of interactions between units separated by the spectral partition (S4 Fig).

Importantly, our solution passed a number of basic but challenging tests involving artificial and real brain recordings. As a first application of our result, we investigated the relationship between integrated information and network structure. We found that, consistent with earlier predictions [66, 77, 79, 80, 83], networks with a high global efficiency produce high integrated information and that networks with high structural modularity produce low integrated information (Fig. 2.5). This observation may help in pinpointing brain structures with high levels of information integration. For example, it has been assumed that the cerebellum does not integrate much information because of its highly modular architecture, while the rich, recurrent cross-module connectivity of the thalamocortical system has been assumed to allow for high levels of information integration [84–86]. Our simulation-based results support this hypothesis, though the truth of the matter will clearly need to be determined on the anvil of experiment.

We also found that our method for identifying the MIB of large systems split posterior sensory areas from anterior association areas in both monkey cortices we tested (Fig. 2.4). In strict

mathematical terms, this means that activity in posterior and anterior regions evolved largely independently over time. We note that both monkeys were awake and resting while the data we analyzed were collected; it would be interesting to see whether the demarcation of independent information-processing sub-networks might vary as a function of cognitive task or brain state.

Because our solution to the problem of searching for the MIB in large networks has made it possible to measure integrated information in real brains, we envision the described solution becoming a broadly applicable tool for neuroscience. In particular, our solution can help to elucidate the function of recurrent brain networks, just as the information-theoretic measure of channel capacity revealed coding schemes in feedforward brain circuits [31–37]. Our method can also be used to directly test the Integrated Information Theory of Consciousness [42], for example by measuring changes in information integration during states of unconsciousness, like anesthesia. With respect to the applicability of our method to the Integrated Information Theory of Consciousness, it is worth pointing out one fascinating result here, which was that in the macaque brains, integrated information peaked at a time-lag of around 100 ms (S3 Fig), which roughly corresponds to the observed timescale of conscious human perception [87, 88]. This matching of time scales is one prediction of the Integrated Information Theory of Consciousness [42, 57], though this correspondence should be investigated more systematically in future empirical work.

Given the potential usefulness of measuring integrated information in complex systems more generally, our method may also be of use to researchers in other fields as well. To facilitate such research, we have made our Matlab toolbox publicly available.

2.5 Methods

2.5.1 Simulating Connectomes

We here describe our algorithm for generating artificial brain-like networks or "connectomes." First, following insights from the evolutionary neuroscience literature [89], the number of modules in our networks was equal to the log of their number of nodes, rounded up. The sizes of the modules in these networks were random, though the sizes of the modules did not vary significantly because each node had an equal probability of being assigned to any given module. Undirected edges were cast between nodes according to two different probabilities: for a pair of nodes i and j where $i \neq j$, an edge was cast between j and i according to a probability p_{int} if both nodes were in the same module and with probability p_{ext} if they were in different

modules. For a given network with M modules and for a given module with n nodes, if $n \geq 4$, then $p_{int} = \frac{4.5}{n}$ and $p_{ext} = \frac{3.3}{nM}$; otherwise $p_{int} = \frac{4}{n}$ and $p_{ext} = \frac{3.75}{nM}$.

To mimic a basic Hebbian process, the nodes that made the most connections were then rewarded with even more connections and the nodes that made the fewest connections were punished by having their connections pruned. The process works like so: after edges have been cast according to the two probabilities p_{int} and p_{ext}, find q, such that around 38% of nodes have made fewer than q connections (this parameter of 38% was chosen somewhat arbitrarily, but it reliably led to a log-normal degree distribution as desired). Create a vector x with elements $[q-1, q, q+1, ..., f+5]$, where f is the largest number of connections that any node in the network made in the previous step of casting out connections. Create a second vector y of the same length as vector x. The first $\frac{f}{4}$ elements of y are set to 1, and the last $l - f + 1$ elements of y are set to Z, where $Z = \sqrt{N} + \log\frac{N}{7}$, N is the number of nodes in the network, and l is the length of vectors x and y. The middle w elements of y, where $w = f - \frac{f}{4} + 1$, are replaced with the vector $[1, 1 + \frac{Z}{w}, 1 + 2\frac{Z}{w}, ..., Z]$. A sigmoid function S is fit to x and y. For every node in the network, random connections are pruned or added, such that every node now has $S(c)$ connections, where c is the number of edges the node had before pruning or adding connections. All networks were checked to ensure that in a given network, any node could be reached by any other node. The resulting networks recapitulated basic features of brain networks, including a modular structure with rich cross-module connectivity [51], as well as a log-normal degree distributions with long right tails [73].

2.5.2 Simulating Time-Series Data With Coupled Stochastic Rössler Oscillators

To simulate oscillatory brain signals from our artificial networks, we used a stochastic Rössler oscillator model. We chose to simulate data using Rössler oscillators because, as has been previously shown [72], they follow a multivariate normal distribution when weakly coupled (S2 Fig). The system of Rössler oscillators is modeled by the following differential equations:

$$\dot{x}^i = -wy^i - z^i - \sigma \sum_{h=1}^{N} g_{ih} x^h \tag{2.13}$$

$$\dot{y}^i = wx^i + ay^i + d\eta^i \tag{2.14}$$

$$\dot{z}^i = b + (x^i - c)z^i \tag{2.15}$$

where, following previous literature [72, 90], a=0.2, b=0.2, and c=9. The oscillation frequencies w were normally distributed around a mean of 10 with a standard deviation of .1. d was set to 750, and η^i is Gaussian noise. g_{ih} are the coefficients of the network's Laplacian matrix, and σ is the coupling strength between oscillators. For all simulations other than the ones reported in Fig. 2.5 (where the coupling was 0.25 for all networks), σ was determined using a master stability function. Master stability functions give the lower and upper bounds for the coupling strengths that ensure network synchronizability. For networks of coupled Rössler oscillators, the lower-bound for the coupling strength is 0.186 divided by the second top eigenvalue of the network's Laplacian matrix, and the upper-bound is 4.614 divided by the last eigenvalue of the network's Laplacian matrix [91, 92]. For each network, σ was set to the half-way point between these lower- and upper-bounds. The equations were integrated with a Euler algorithm, with dt=0.001. For our time-series, we took the y component of these equations, which yielded rich synchronization dynamics and followed a multivariate normal distribution (S2 Fig).

2.5.3 Reducing the Search Space for the MIB

As shown in [52], spectral clustering provides an approximate but robust solution to the "normalized cut" or Ncut problem in graph theory. The problem is motivated by a body of work on how to partition a graph $\mathbf{G = (V,E)}$, with $\mathbf{V}$ vertices and $\mathbf{E}$ edges, into disjoint subsets $A, B, A \cup B = V, A \cap B = \emptyset$. The Ncut problem entails finding a network cut which minimizes the following measure:

$$Ncut(A,B) = \frac{cut(A,B)}{assoc(A,V)} + \frac{cut(A,B)}{assoc(B,V)} \tag{2.16}$$

where $cut(A,B)$ is the sum of edges (binary or weighted) crossing a particular cut, $assoc(A,V)$ is the sum of edges between community A and the entire network, and $assoc(B,V)$ is similarly the sum of edges between community B and the entire network. Dividing $cut(A,B)$ by the normalization factors $assoc(A,V)$ and $assoc(B,V)$ helps ensure that the clusters separated by the bipartition are relatively balanced in size, and as such serves the same function as the normalization function K (Eq. 10) in the search for the MIB.

Shi and Malik [52] developed a fast spectral clustering algorithm that can quickly find a partition that (approximately but robustly) minimizes the Ncut function. The algorithm applies k-means clustering to the eigenvectors corresponding to the top k eigenvalues of a network's Laplacian matrix, where k is the number of communities being split (so, for a bipartition, k=2). Though many other clustering methods are available, we chose spectral clustering because it is

particularly well-suited for *normalized* clustering problems, and as such is appropriate for the search for the MIB.

The principle contribution of this paper is the empirical finding that the MIB of a network can be approximated by applying spectral clustering to correlation matrices of time-series data. To get a range of candidate partitions from a single correlation matrix, we first applied a power adjacency function [65] to the correlation matrix C, such that every correlation value r_{ij} in C is mapped onto a continuous edge weight w_{ij}:

$$w_{ij} = (\frac{r_{ij} + 1}{2})^{\beta} \tag{2.17}$$

The value chosen for β determines the shape of the power adjacency function. We iterated through 10 values of β, logarithmically spaced between 1 and 10. For every resulting power adjacency transformation of C, we then iterated through a range of cutoff values (from the 0th to the .99th percentile of weights in steps of 0.005), and for every iteration, all edge weights less than that cutoff value were set to 0 (following [66–70]). Spectral clustering was then applied to the Laplacian matrix computed from each adjacency matrix, as well as to the Laplacian matrix computed from the un-thresholded correlation matrix. In total, this resulted in 2189 candidate partitions for each dataset. Φ^{G} (normalized) was calculated for each of these candidate partitions, and we chose among these the partition that minimized Φ^{G} (normalized) as our spectral clustering-based alternative to the MIB (identified through a brute-force search). Note that, to our knowledge, there is no analytic guarantee that the MIB will be among these 2189 candidate partitions, and so the work presented here can be seen as a numerical experiment strongly motivating the proposal that there is a relationship between the MIB and the spectral partition of the correlation matrix of time-series data. In future work, we hope to analytically study this relationship in greater depth.

2.5.4 ECoG Preprocessing

ECoG data from the left cortex of two monkeys, Chibi and George, is publicly available on neurotycho.org [373]. Data from 128 electrodes were available for over an hour of recording from both monkeys. We selected the first 50,000 ms of data from both monkeys. The data were then down-sampled to 500 Hz, demeaned, de-trended, and band-stop filtered for 50 Hz and harmonics, which is the line noise in Japan (where the data were collected). Data were then re-referenced to the common average across electrodes. We then visually inspected the data for artifacts. Segments of data with artifacts that spread across more than one electrode

were removed from all electrodes, and individual electrodes with consistent artifacts that did not spread to their neighbors were removed entirely (electrodes 14, 28, and 80 were removed for George, and electrodes 17, 53, and 107 were removed for Chibi). The pre-processed ECoG data were approximately multivariate normal (S2 Fig), allowing for the fast measurement of integrated information.

2.6 Supporting Information

2.6.1 Extrapolating Integrated Information to Infinite Observations

As we mentioned in the main body of the paper, estimates of information measures from finite time-series data are typically over-estimated. To get around this over-estimation problem, past work in information theory [1, 35, 36], inspired by earlier work in statistical mechanics [4], has produced a simple method for extrapolating information estimates to what they would be if infinite data were available. Though we did not utilize this extrapolation method in the main body of the paper, we present it here for two reasons: 1) we include the method in our Matlab toolbox, and 2) to demonstrate that it successfully mitigates the over-estimated results reported in Fig. 2.3 (i.e. the results from the structurally severed networks, for which the ground-truth value of integrated information should trivially be 0 bits).

Consider a jointly Gaussian variable, described by a linear regression model:

$$X_t = AX_{t-\tau} + E \tag{2.18}$$

where X_t is the present of the system, $X_{t-\tau}$ is the past of the system, A is the regression matrix, and E is the error or residuals in the linear regression. The geometric integrated information of a this Gaussian process can be calculated analytically without time-series data, by constructing a random regression matrix A. To do so, we simply create a random dense positive definite matrix. This can be done in Matlab using `sprandsym.m`. To ensure that the graph is fully connected, we set the density of the graph to 1 (otherwise the MIB will trivially separate disconnected nodes). Since the calculation of geometric integrated information assumes stationarity, we decay the coefficients of the random dense positive definite matrix so that its spectral radius is less than 1 (here, we pick a spectral radius of 0.8). This can be done with the `var_specrad.m` function of the MVGC Multivariate Granger Causality Matlab Toolbox [220].

For simplicity, we assume that the dynamics of the system are fully described by interactions between the variables and that "noise" in the system is uncorrelated. Thus, we set the elements

along the diagonal of the covariance of the error matrix to 1, and set all off-diagonal elements to 0.

In this system, the covariance of the present state of the system can be computed by solving the discrete-time Lyapunov equation. In Matlab, this can be done with `dlyap.m` command. This gives us $\Sigma(X)$, and so we have everything we need to analytically calculate integrated information in this system. To do so, we iterate through all possible bipartitions of the regression matrix A, and for each candidate partition we set all cross-partition connections in the corresponding disconnected regression matrix A' to 0. We then use an augmented Lagrangian method to determine the non-zero values of A', which then lets us analytically calculate the covariance matrix $\Sigma(E')$ of the residuals of the disconnected model. We further use $\Sigma(X)$ to analytically compute the differential entropy of each divided sub-system: if K is the submatrix of $\Sigma(X)$ that corresponds to one sub-community, then the entropy of that sub-community is $\frac{1}{2}\log[(2\pi\mathrm{e})^{\mathrm{n}}|\det(\mathrm{K})]]$, where n is the number of variables/nodes in the sub-community. We then use these entropy values to calculate normalized integrated information. We select the partition that minimizes normalized integrated information as the MIB. The non-normalized integrated information across that partition is the analytic ground-truth for this system.

We can now demonstrate the accuracy of our extrapolation method by generating time-series data for these systems. We do this by simply starting with random initial conditions, and, at each time step, we multiply the preceding values by the regression matrix and add Gaussian noise (which, by constructing $\Sigma(E)$ such that off-diagonal elements are 0, we ensure is uncorrelated).

With this multivariate time-series, we can compute integrated information following the same steps as those described in the main body of the paper. If the widely-used method for extrapolating entropy and mutual information to infinite observations (as described in the main body of the paper) works well for integrated information, then we should expect our extrapolated estimate of Φ^{G} to be close to the ground-truth in these systems.

We followed the above steps to generate analytic ground-truth values, time-series data, and extrapolated estimates of Φ^{G} in 14- and 16-node networks, for which the ground-truth MIB can be established through a brute-force search through all possible bipartitions of the randomly constructed regression matrices. We found that in these networks, the extrapolation method is highly successful, and provides values very close to the ground-truth (Figure S1A-C). Moreover, the majority of MIBs identified by extrapolating on the time-series data were identical to their corresponding ground-truth MIBs (determined by searching through all possible bipartitions of the ground-truth regression matrix) (S1D Fig).

Finally, we applied this extrapolation method to the simulated data generated from the structurally severed networks reported in Fig. 2.3. In principle, the most rigorous method of extrapolation (which was used in S1A-D Fig) would be to extrapolate to infinite observations for *every candidate bipartition*, and to pick as the MIB the bipartition that minimizes *extrapolated* normalized integrated information. This becomes very comutationally expensive for large networks, however, so for the following analyses we only extrapolated to infinite observations using the bipartition estimated from the finite data. (Note that our Matlab toolbox can handle either picking one partition and extrapolating on that partition, or extrapolating on all candidate partitions). We found that although estimates of Φ^{G} in the cut networks diverge from the expected ground-truth of 0 bits (S1E-F Fig), as is expected given the notorious over-estimation bias of information measures, extrapolation reduces this over-estimation bias and brings estimates closer to the ground-truth of 0 bits (S1G-H Fig).

Bibliography

[1] Strong SP, Koberle R, van Steveninck RRdR, Bialek W. Entropy and information in neural spike trains. Physical review letters. 1998;80(1):197.

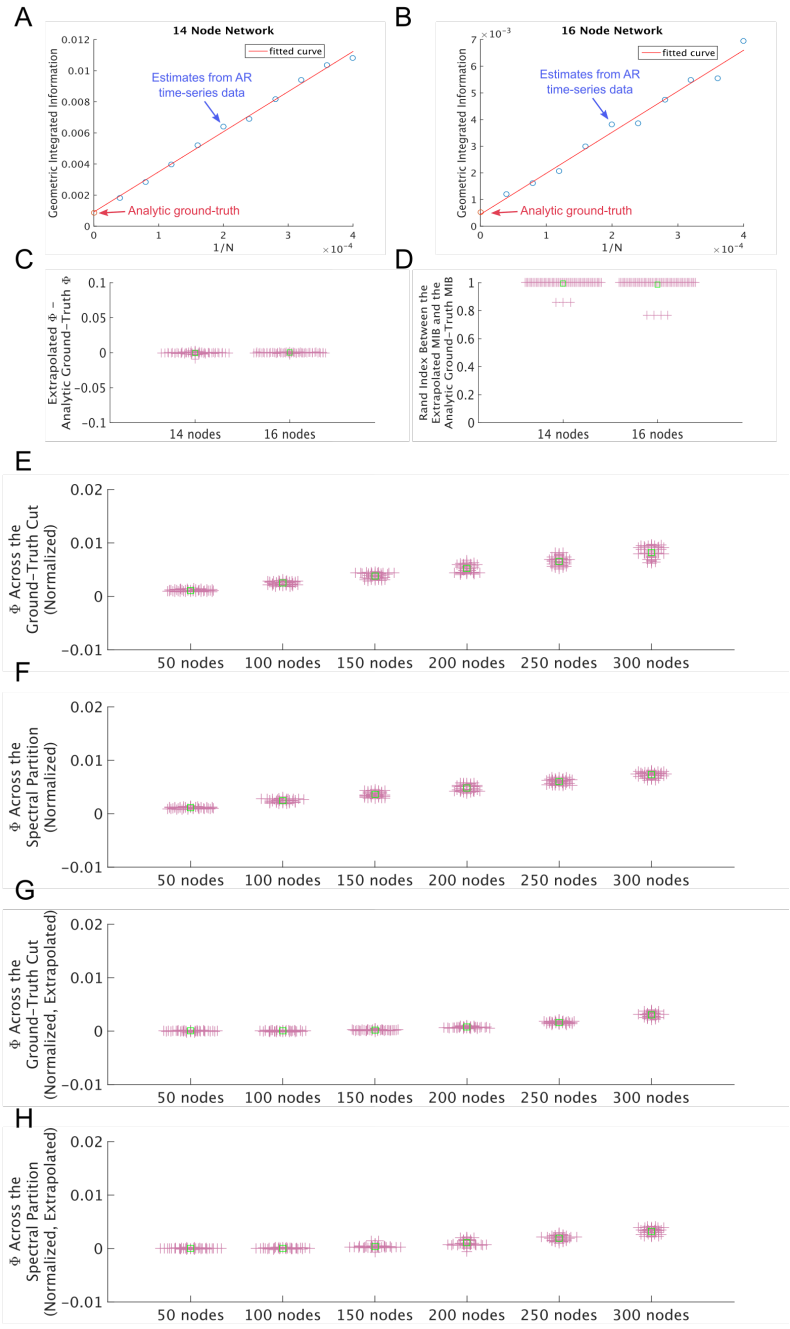
[2] Brenner N, Strong SP, Koberle R, Bialek W, Van Steveninck RRDR. Synergy in a neural code. Neural computation. 2000;12(7):1531–1552.

[3] Dayan P, Abbott LF. Theoretical neuroscience. vol. 806. Cambridge, MA: MIT Press; 2001.

[4] Ma Sk. Calculation of entropy from data of motion. Journal of Statistical Physics. 1981;26(2):221–240.

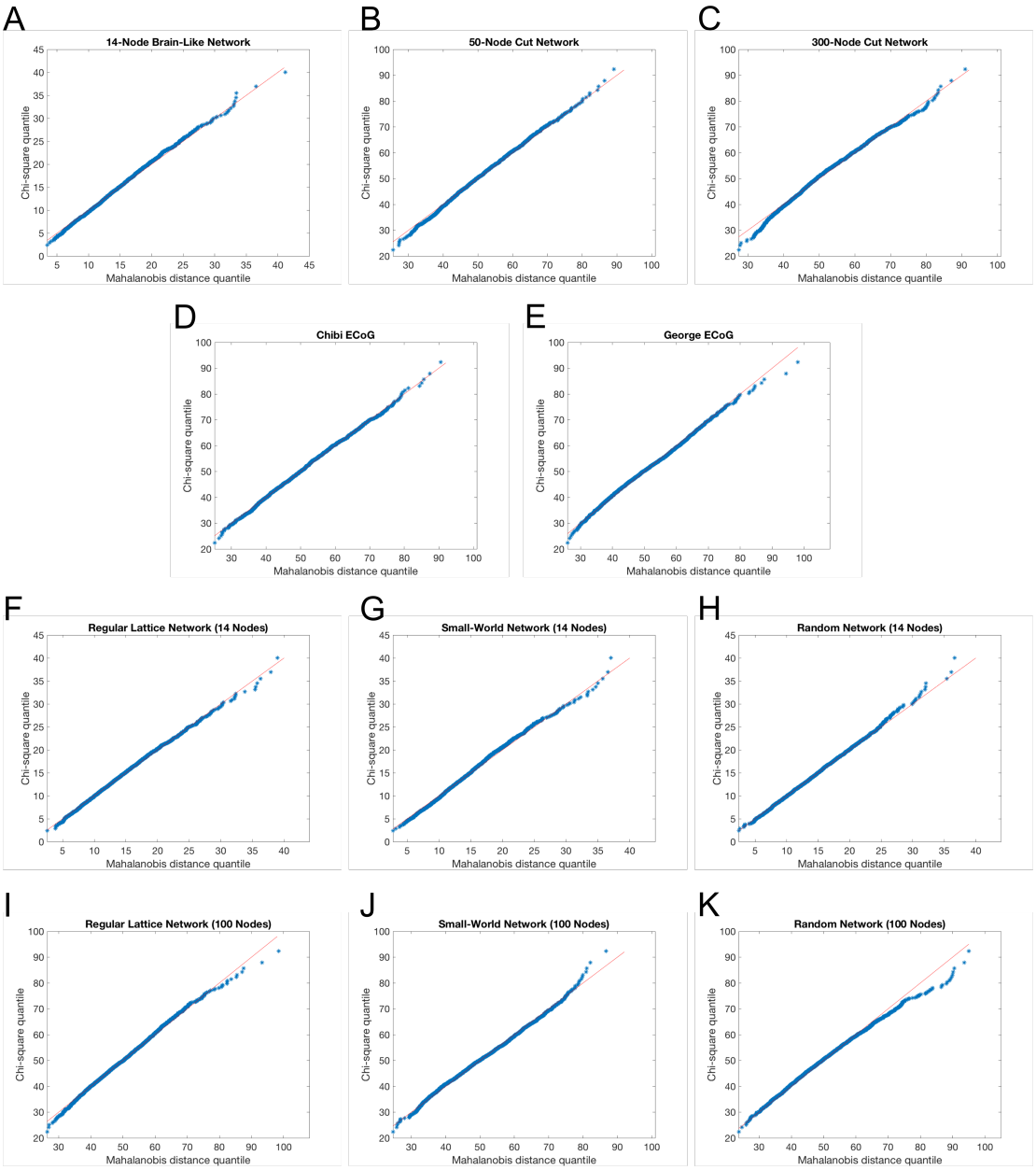
[5] Barnett L, Seth AK. The MVGC multivariate Granger causality toolbox: a new approach to Granger-causal inference. Journal of neuroscience methods. 2014;223:50–68.

2.6.2 Supplementary Figures

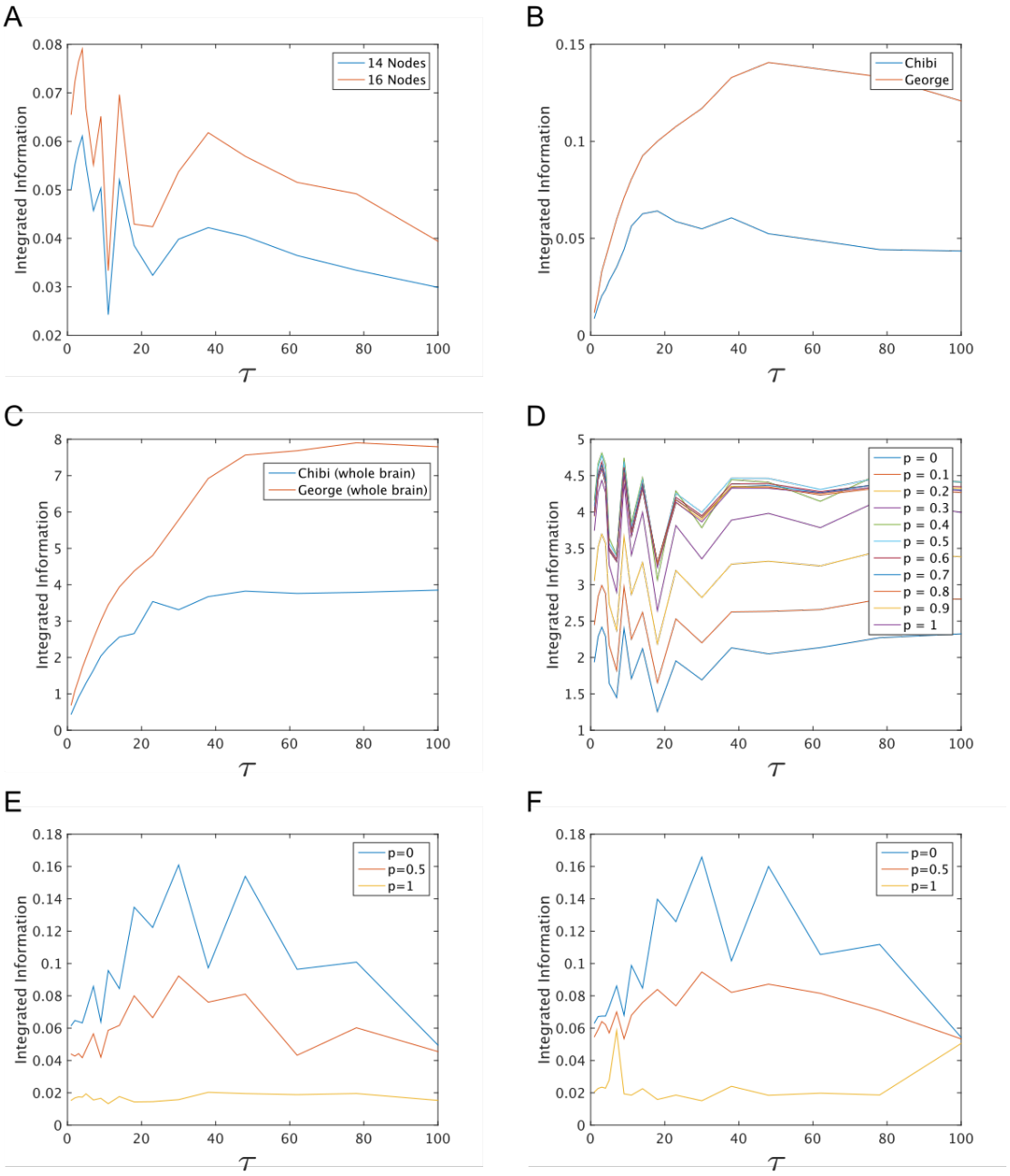


(Caption on next page.)

S1 Figure. Results of extrapolation analysis. In the absence of infinite data, integrated information can be accurately estimated from finite time-series data. Following well-established for estimating entropy and mutual information from finite time-series data, we calculate integrated information from time-series data as a function of $1/N$, where N is the number of observations. We then fit a line to this distribution of estimated values, and estimate integrated information as the value of the line at $1/N = 0$. **A** A sample of measurements of Φ^{G} as a function of 1/N in a 14-node autoregressive system. The orange circle on the y-axis is the analytic ground-truth value of Φ^{G} in this system, and the y-intercept of the line fitted to estimates of Φ^{G} from sub-sampled time-series data is the estimate of Φ^{G} extrapolated to infinite observations (i.e. where 1/N=0). Note that the value of the fitted line at $1/N = 0$ is remarkably close to the analytic ground-truth value. **B** A sample estimate of integrated information from a 16-node autoregressive system. **C** We generated autoregressive time-series data from 60 14-node networks and 60 16-node networks, and then compared our extrapolated estimates of integrated information to the analytic ground-truth for those systems. Our extrapolated estimates were highly accurate: the mean absolute error for the 14-node networks was 0.001 bits, and the mean absolute error for the 16-node networks was less than 0.0005 bits. Red squares are the mean across tested networks, and blue bars indicate standard error of the mean. **D** For our autoregressive time-series, we extrapolated integrated information to infinite observations across all possible bipartitions, and selected the bipartition that minimized *normalized* integrated information extrapolated to infinite observations as the MIB. We then compared these bipartitions to the analytic ground-truth MIBs. Our estimates based on time-series data found the exact (analytic) MIB in 57/60 of the 14-node systems (mean Rand Index = 0.99) and in 56/60 of the 16-node systems (mean Rand Index = 0.98). Thus, we can confidently assume that well-established extrapolation methods work for integrated information. **E** In the structurally severed networks in Fig. 2.3 of the main paper, there should trivially be 0 bits of integrated information. Yet, as we show here, estimates of Φ^{G} in these networks given finite time-series data often yield results greater than zero. In particular, larger networks seem to result in larger over-estimation of Φ^{G}. This same finding holds for non-extrapolated values across the bipartitions identified by our spectral clustering approach (**F**), which is not surprising, because our approach usually found partitions similar to the ground-truth cuts in these networks (Fig. 2.3). Using the extrapolation procedure described here, however, mitigates this over-estimation bias and brings estimates of integrated information close to the ground-truth of 0 bits (**G,H**).

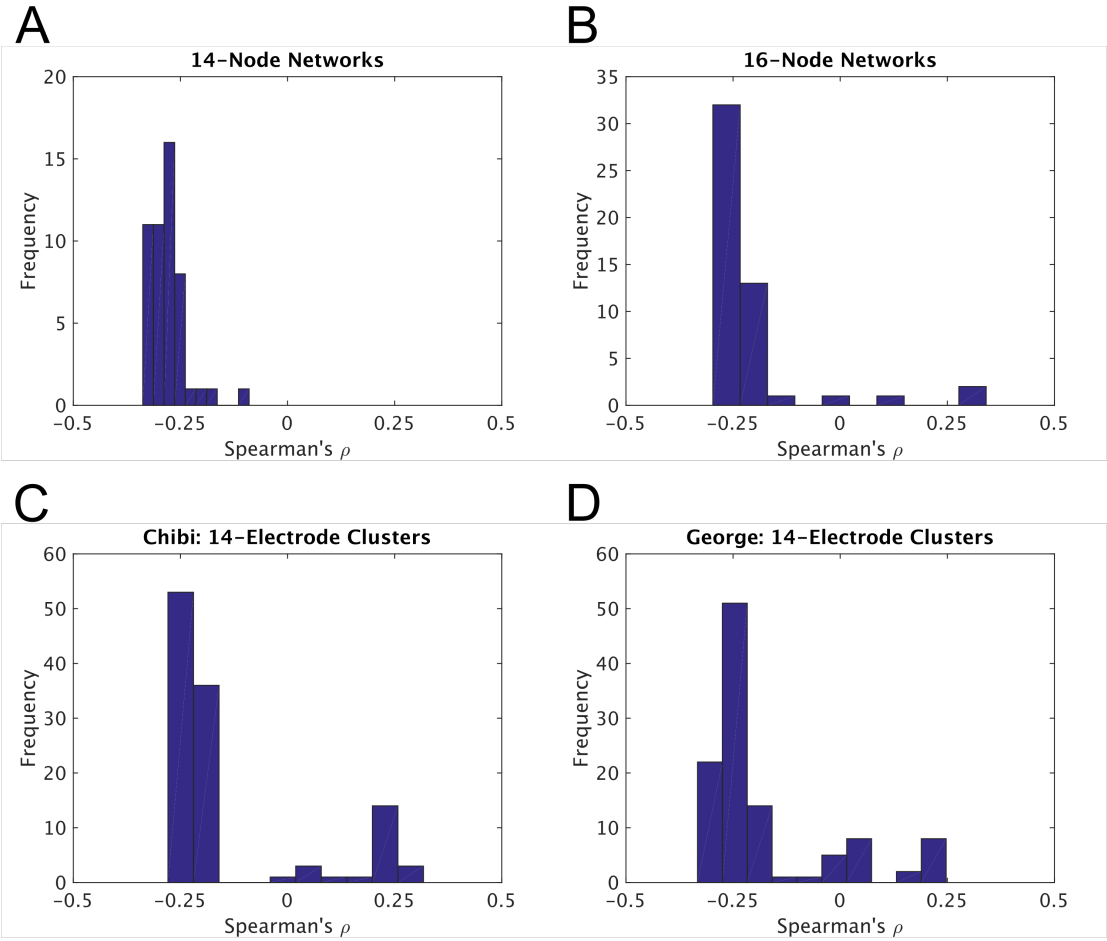


S2 Figure. Assessing data multivariate normality. All time-series data analyzed in this paper were approximately multivariate normal. This is important, because the estimator of geometric integrated information we used assumes multivariate normality. As a graphical test of multivariate normality, we used multivariate Q-Q plots: if data are multivariate normal, then there should be a linear relationship between the ordered Mahalanobis distances of the data from the mean vector and their corresponding chi-square quantiles. Here, we show sample multivariate Q-Q plots for our 14-node brain-like networks of coupled Rössler oscillators (**A**), our cut, 50-node brain-like networks of coupled Rössler oscillators (**B**), our cut, 300-node brain-like networks of coupled Rössler oscillators (**C**), the ECoG data from Chibi (**D**), the ECoG data from George (**E**), our 14-node regular lattice networks (rewiring probability p=0) of coupled Rössler oscillators (**F**), our 14-node small-world networks (rewiring probability p=0.5) of coupled Rössler oscillators (**G**), our 14-node random networks (rewiring probability p=1) of coupled Rössler oscillators (**H**), our 100-node regular lattice networks (rewiring probability p=0) of coupled Rössler oscillators (**I**), our 100-node small-world networks (rewiring probability p=0.5) of coupled Rössler oscillators (**J**), and our 100-node random networks (rewiring probability p=1) of coupled Rössler oscillators (**K**).

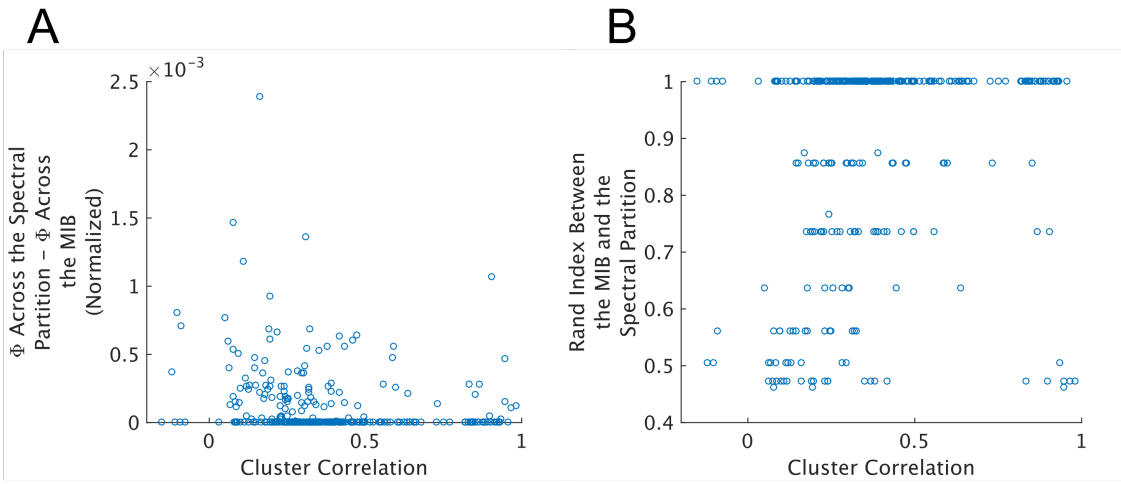


(Caption on next page.)

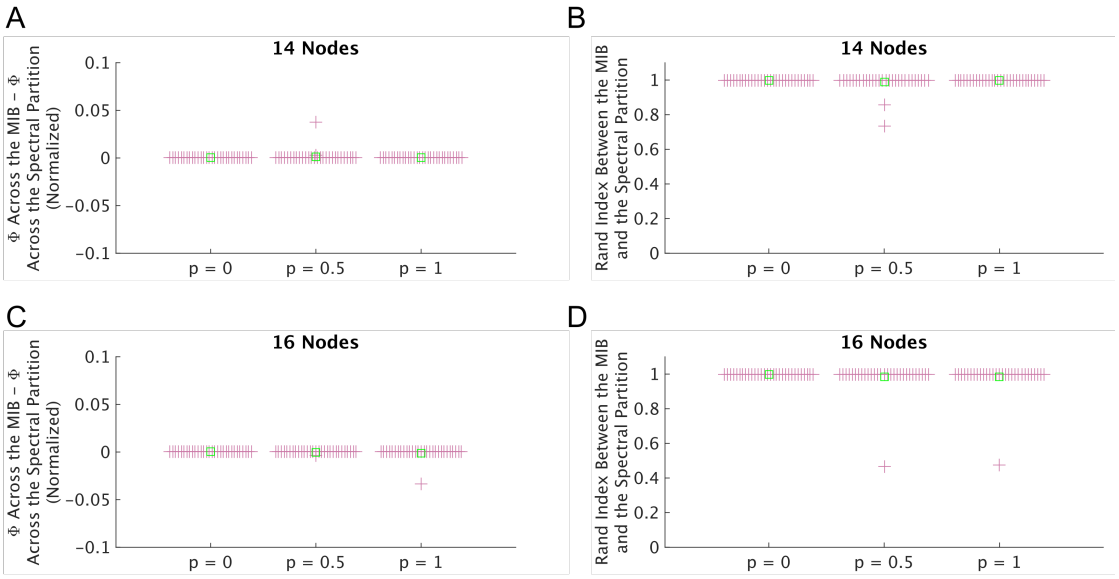
S3 Figure. Integrated information as a function of time-lag. Integrated information is measured over a time-lag τ. For all analyses in our paper, we picked a time-lag that, on average, maximized integrated information for the system at hand. To test a range of time-lag values, we measured integrated information across 17 different time-lags, logarithmically spaced between 1 and 100. For both our 14-node and 16-node brain-like networks of coupled Rössler oscillators, we were able to calculate integrated information across the MIB (estimated from finite data) for each candidate time-lag. Here we plotted the median values of integrated information across all simulated networks. We found that integrated information was typically maximized at a time-lag of 4 (**A**). For the 14-electrode monkey ECoG data, we picked a time-lag of 38, as the median integrated information as a function of time-lag across all 112 sets of electrodes per monkey seemed to peak or aymptote around 38 (**B**). For the whole monkey brains, we calculated integrated information using our spectral clustering-based approach, and found that our estimates peaked around a time-lag of 48, which is 96 ms. (**C**). We similarly calculated integrated information as a function of time-lag using our spectral clustering approach in our 100-node Watts-Strogatz networks (which ranged from regular lattice networks to small-world networks to random networks, depending on the value of rewiring probability p). Interestingly, we found quite different behaviors of integrated information as a function of time-lag for networks with a rewiring probability less than 0.1 than we did for networks with a rewiring probability greater than 0.1, but all networks peaked at a time-lag of 48, and so we chose a τ of 48 for the analyses in Fig. 2.5. (**D**). Finally, we calculated integrated information across the MIB in our 14-node Watts-Strogatz networks (**E**) and in our 16-node Watts-Strogatz networks (**F**), and found that the median integrated information peaked around a time-lag of 30. For the cut networks analyzed in Fig. 2.3, we used a time-lag of 1 beceause there is objectively no information integration in these networks, and thus the choice of time-lag should not matter.



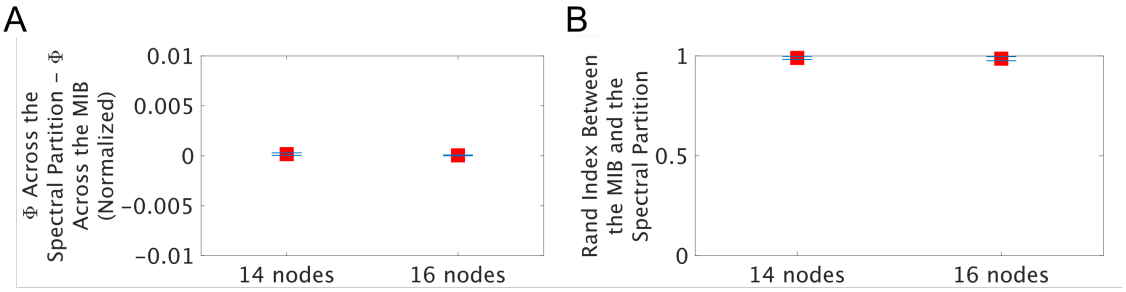
S4 Figure. Integrated information and partition similarity to the MIB. In general, but not always, the closer a partition is to a network's minimum information bipartition, the smaller integrated information (normalized) will be across that partition. Here, we calculated Φ^{G} (normalized) across every possible bipartition of every small dataset (up to 16 nodes) analyzed in the main body of this paper. We then computed the Rand Index between those bipartitions and the ground-truth minimum information bipartition of that dataset. If finding a partition close to the ground-truth MIB is important for accurate estimation of integrated information, then we should in general expect a negative correlation between Φ^{G} (normalized) across all possible bipartitions and the Rand Indices between those bipartitions and the ground-truth MIB (though in some networks there may be large local minima that are dissimilar to the MIB). In other words, the more dissimilar a partition is from the MIB, then the higher normalized integrated information across that partition should typically be (again, discounting the case of large local minima). Here, we find that this is generally the case in both real and simulated data. Because the distributions of Rand Indices were not normally distributed, we calculated the correlation between Rand Indices and normalized Φ^{G} values using a Spearman's rank correlation. We found negative Spearman's correlations, as expected, for most 14-node networks of coupled Rössler oscillators (**A**), 16-node networks of coupled Rössler oscillators (**B**), and 14-electrode clusters of ECoG data for both monkeys (**C, D**).



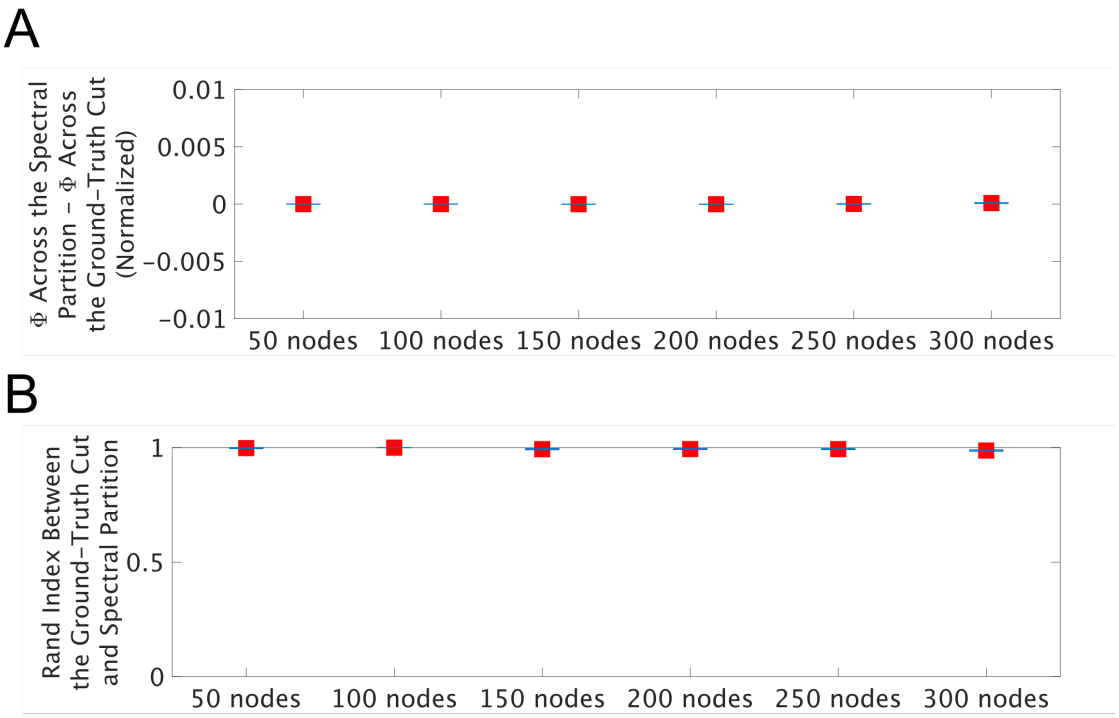
S5 Figure. Cluster correlation as a predictor of MIB estimation accuracy. Though our spectral clustering-based approach found the exact MIB in the majority of small datasets analyzed in this paper (for which the ground-truth can be computed through a brute-force search), it did not always find the exact solution. As we report in the main body of the paper, in the cases that spectral clustering did not find the exact MIB, it usually still found partitions close to the MIB, and yielded values of integrated information that were close to the ground-truth. But this presents an interesting problem: is there a way to predict how likely it is that spectral clustering found the right solution - or something near it - when the ground-truth is not known? Though we could not find a variable that was predictive of the success of spectral clustering in all cases, we did observe that the more correlated nodes were on either side of the partition, the more likely it was that spectral clustering had found the ground-truth MIB. Here, we capture that observation with a variable which we call "cluster correlation," which is the mean of all within-cluster correlations minus the mean of all cross-cluster correlations. In other words, if spectral clustering splits a network into clusters A and B, then the cluster correlation is the mean of the correlations between all nodes in A, minus the mean of correlations between nodes in A and nodes in B. In this analysis, we collapsed results across all datasets for which the ground-truth MIB could be observed (i.e., all 100 small simulated brain-like networks of coupled Rössler oscillators, and the 224 small sets of monkey ECoG electrodes). **A** We observed a general negative correlation (Spearman's ρ=-0.28, p$< 10^{-6}$) between the cluster correlation and the difference between integrated information (normalized) across the MIB and integrated information (normalized) across the spectral partition. In other words, the stronger the within-community correlations were relative to cross-community correlations, the closer Φ^{G} (normalized) across the spectral partition was to Φ^{G} (normalized) across the MIB. This is probably because community clustering is generally easier when there are strong within-community weights or correlations. In contrast, we did not observe a significant negative correlation (Spearman's ρ=-0.0265, p=0.64) between the modularity Q of correlation matrices split by spectral partitions and the difference between integrated information (normalized) across the MIB and integrated information (normalized) across spectral partitions, despite the intuitive similarity between modularity and our measure of cluster correlation. **B** Consistent with our results in **A**, there was a significant *positive* correlation (Spearman's ρ=0.29, p$< 10^{-7}$) between the cluster correlation and the Rand Index between the spectral partition and the MIB. Again, the modularity of the correlation matrices split by spectral partitions was not significantly correlated with the Rand indices (Spearman's ρ=0.06, p=0.28). These results suggest that spectral clustering provides the most accurate estimate of the MIB in networks whose time-series data produce a correlation structure in which within-module correlations are stronger than cross-module correlations. That said, we found that spectral clustering also usually finds the exact MIB in non-modular networks, such as regular lattice networks and random networks (S6 Fig).



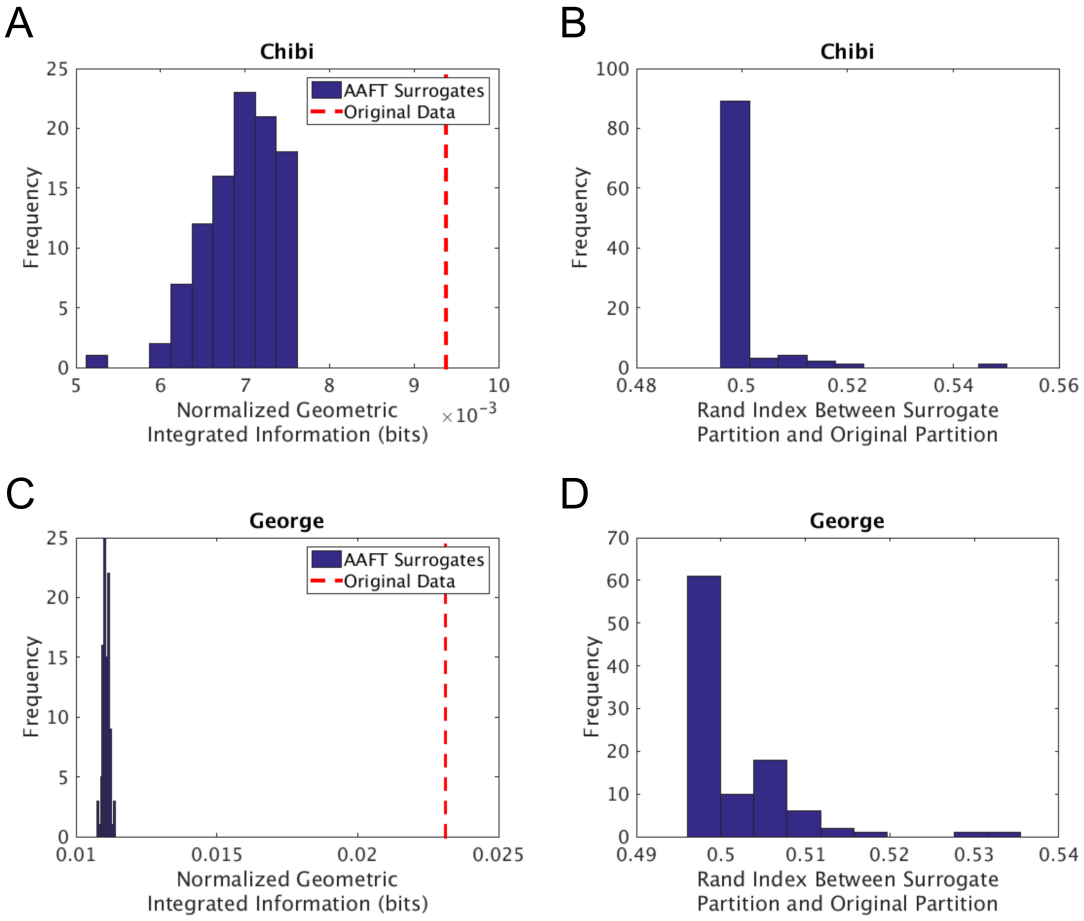
S6 Figure. Spectral clustering accuracy in Watts-Strogatz networks. Spectral clustering accurately identifies the MIB of a variety of network types. Here, as in Fig. 2.5 in the main body of the paper, we followed Watts and Strogaz's method for generating different network types by varying the rewiring probability p of lattice networks. A p of 0 yields a perfect lattice network, a p of 0.5 yields a small-world network, and a p of 1 yields a random network. Here, we generated 30 lattice ($p = 0$), 30 small-world ($p = 0.5$), and 30 random ($p = 1$) networks, of 14- and 16-nodes in size. With each of these networks, 25,000 time-points of oscillatory time-series data were generated using the stochastic coupled Rössler oscillator model, with the coupling coefficient set to 0.7 for all networks; these simulations yielded multivariate normal dynamics (Fig. 2.S2F-H). We then compared estimates of integrated information across the MIB and across the spectral clustering-based partition for each of the resulting 180 data sets. We found that spectral clustering performed perfectly in all regular lattice ($p = 0$), with perfect matches to the MIB (**B,D**) and no difference from the ground-truth Φ (normalized) values (**A,C**) in all tested networks. For the small-world networks ($p = 0.5$), spectral clustering performed perfectly in 28/30 14-node networks (**A,B**) and in 29/30 16-node networks (**C,D**). For the random networks ($p = 1$), spectral clustering performed perfectly in all 14-node networks (**A,B**) and in 29/30 16-node networks (**C,D**). Red squares are the mean across tested networks, and blue bars indicate standard error of the mean. Overall, these results show that spectral clustering can accurately identify the MIB in a variety of network types.



S7 Figure. Spectral clustering accuracy in small autogregressive systems. In the main body of the paper, we showed that our spectral clustering-based approach finds the exact MIB (determined through a brute-force search) in almost all small brain-like networks of coupled oscillators (Fig. 2.2). To show that our approach not only generalizes across a variety of network types (Fig. 2.S6), but also generalizes across a variety of network *dynamics*, we also used the same networks used in Fig. 2.2 to generate a set of autoregressive time-series data (as opposed to coupled Rössler oscillators). Autoregressive data are, by construction, multivariate Gaussian, and so the same estimator of geometric integrated information used for the Rössler oscillators can be used for autoregressive simulations. To ensure that the autoregressive simulations were linearly stable, we used `var_specrad.m` function of the MVGC Multivariate Granger Causality Matlab Toolbox [220] to decay the coefficients of the adjacency matrices of the brain-like networks of Fig. 2.2, so that their spectral radii were 0.8. Moreover, because the model order of our simulations was 1, we used a time-lag of 1 in our calculations of integrated information for these data. Our spectral clustering approach once again performed almost perfectly: the difference to the ground-truth values of Φ^{G} was 0 for 48/50 of both the 14- and 16-node networks (**A**), and the Rand Index between the spectral partitions and the ground-truth MIBs was 1 for the same 48/50 14- and 16-node networks (**B**). Just as was the case for the coupled oscillators, the Queyranne algorithm performed poorly for the autoregressive data, yielding no exact matches for the 14-node networks (mean difference to the ground-truth normalized Φ^{G} values = 0.04 bits, mean Rand index = 0.56) and only one exact match in the 16-node networks (mean difference to the ground-truth normalized Φ^{G} values = 0.03 bits, mean Rand index = 0.57).



S8 Figure. Spectral clustering accuracy in large autogregressive systems. In the main body of the paper, we showed that our spectral clustering-based approach well approximates the MIB of large-cut networks (Fig. 2.3). To further show that our approach generalizes across a variety of network dynamics, we also used the same large, cut networks used in Fig. 2.3 to generate a set of autoregressive time-series data, following the same procedure as in Fig. 2.S7. We found that our spectral clustering approach performed almost perfectly in large autoregressive systems: the difference between Φ^{G} across the spectral partitions and Φ^{G} across the ground-truth cuts was 0 for 191/240 networks tested, with the difference less than 10^{-6} bits (normalized) for all network sizes. The spectral partitions were perfect matches to the ground-truth cuts for those same 191/240 networks, with a mean Rand index greater than 0.98 for all network sizes.



S9 Figure. Surrogate analysis of monkey ECoG results. To test the robustness of our results for the full monkey brains (Fig. 2.4), we used surrogate statistical testing. We generated 100 surrogates for each ECoG electrode in both monkeys using the Amplitude Adjusted Fourier Transform (AAFT) algorithm [165], which creates a time-series with the same linear structure and amplitude distribution as the original data, but which is otherwise random. We created surrogates from each electrode independently so as to break any cross-electrode coupling or correlations. We thus had 100 surrogate multivariate datasets for each monkey. We then calculated integrated information in each of these datasets using our spectral clustering approach. If the results reported in Fig. 2.4 are meaningful, i.e. they are not an artifact of either the ECoG data or the spectral clustering algorithm, then we should expect two results here: 1) normalized integrated information in the original ECoG data (in which there is some information actually being integrated) should be significantly greater than the distribution of integrated information calculated from the surrogate datasets (in which there is objectively no information integration, despite sharing all linear features with the original data), and 2) the estimated MIBs of the surrogate datasets should be random, and thus dissimilar to the estimated MIBs of the monkey cortices, which we claim are not random (Fig 4). That is precisely what we found: normalized integrated information was significantly higher in the original data than in the surrogate data in both monkeys (**A,C**), and the Rand indices between the MIBs estimated from original data and the MIBs estimated from the surrogate data clustered around 0.5 in both monkeys (**B,D**), which is precisely what we would expect for random partitions.

Chapter 3

A Simple Method for Detecting Chaos in Nature

Daniel Toker, Friedrich T. Sommer, and Mark D'Esposito

3.1 Abstract

Chaos, or exponential sensitivity to small perturbations, appears everywhere in nature. Moreover, chaos is predicted to play diverse functional roles in living systems. A method for detecting chaos from empirical measurements should therefore be a key component of the biologist's toolkit. But, classic chaos-detection tools are highly sensitive to measurement noise and break down for common edge cases, making it difficult to detect chaos in domains, like biology, where measurements are noisy. However, newer tools promise to overcome these limitations. Here, we combine several such tools into an automated processing pipeline, and show that our pipeline can detect the presence (or absence) of chaos in noisy recordings, even for difficult edge cases. As a first-pass application of our pipeline, we show that heart rate variability is not chaotic as some have proposed, and instead reflects a stochastic process in both health and disease. Our tool is easy-to-use and freely available.

3.2 Introduction

A remarkable diversity of natural phenomena are thought to be chaotic. Formally, a system is chaotic if it is bounded (meaning that, like a planet circling a star, its dynamics stay inside an

orbit rather than escaping off to infinity), and if it is deterministic (meaning that, with the exact same initial conditions, it will always evolve over time in the same way), and if tiny perturbations to the system get exponentially amplified (Supplementary Glossary, Supplementary Figures 3.1-3.2). The meteorologist Edward Lorenz famously described this phenomenon as the butterfly effect: in a chaotic system, something as small as the flapping of a butterfly's wings can cause an effect as big as a tornado. This conceptually simple phenomenon - i.e., extreme sensitivity to small perturbations - is thought to appear everywhere in nature, from cosmic inflation [134], to the orbit of Hyperion [135], to the Belousov-Zhabotinskii chemical reaction [136], to the electrodynamics of the stimulated squid giant axon [137]. These are only a few examples of the many places in nature where chaos has been found.

It is relatively simple to determine if a simulated system is chaotic: just run the simulation a few times, with very slightly different initial conditions, and see how quickly the simulations diverge (Supplementary Figure 3.1). But, if all that is available are measurements of how a real, non-simulated system evolves over time - for e.g., how a neuron's membrane potential changes over time, or how the brightness of a star changes over time - how can it be determined if those observations come from a chaotic system? Or if they are just noise? Or if the system is in fact periodic (Supplementary Glossary, Supplementary Figures 3.1-3.2), meaning that, like a clock, small perturbations do not appreciably influence its dynamics?

While a reliable method for detecting chaos using empirical recordings should be an essential part of any scientist's toolbox, such a tool might be especially helpful to biologists, as chaos is predicted to play an important functional role in a wide variety of biological processes[1]. For example, following early speculations about the presence of chaos in the electrodynamics of both cardiac [139] and neural [140] tissue, the science writer Robert Pool posited in 1989 that "chaos may provide a healthy flexibility for the heart, brain, and other parts of the body." [141] Though this point has been intensely debated since the 1980s [138, 282], a range of more specific possible biological functions for chaos have since been proposed, including potentially maximizing the information processing capacity of both neural systems [362] and gene regulatory networks [144], enabling multistable perception [145], allowing neural systems to flexibly transition between different activity patterns [146], and boosting cellular survival rates through the promotion of heterogeneous gene expression [147]. And there is good reason to expect chaos to exist in biological systems, as a large range of simulations of biological processes [283], and in particular of neural systems [282], show clear evidence of chaos. Moreover, unambiguous evidence of biological chaos has been found in a very small number of real cases that were amenable to

[1]That said, we note that real biological systems cannot be purely chaotic in the strict mathematical since, since they certainly contain some level of dynamic noise (Supplementary Glossary), but that researchers have long speculated that many biological processes are still predominantly deterministic, but chaotic [138].

comparison to good theoretical models; these include periodically stimulated squid giant axons [137] and cardiac cells [149], as well as the discharges of the *Onchidium* pacemaker neuron [150] and the *Nitella flexillis* internodal cell [151]. But, beyond simulations and these select empirical cases, most attempts to test the presence or predicted functions of chaos in biology have fallen short due to high levels of measurement noise (Supplementary Glossary) in biological recordings. For this reason, it has long been recognized that biologists need a noise-robust tool for detecting the presence (or absence) of chaos in their noisy empirical data [282, 283].

Researchers also need a tool that can detect varying degrees of chaos (Supplementary Glossary) in noisy recordings. In strongly chaotic systems, initially similar system states diverge faster than they do in weakly chaotic systems. And such varying degrees of chaos are predicted to occur in biology, with functional consequences. For example, a model of white blood cell concentrations in chronic granulocytic leukemia can display varying levels of chaos, and knowing how chaotic those concentrations are in actual leukemia patients could have important implications for health outcomes [152]. As another example, models of the human cortex predict that macro-scale cortical electrodynamics should be weakly chaotic during waking states and should be strongly chaotic under propofol anesthesia [276]; if this prediction is true, then detecting changing levels of chaos in large-scale brain activity could be useful for monitoring depth of anesthesia and for basic anesthesia research. Thus, it is imperative to develop tools that can not only determine that an experimental system is chaotic, but also tools to assess changing levels of chaos in a system.

Although classic tools for detecting the presence and degree of chaos in data are slow, require large amounts of data, are highly sensitive to measurement noise, and break down for common edge cases, more recent mathematical research has provided new, more robust tools for detecting chaos or a lack thereof in noisy time-series recordings. Here, for the first time (to our knowledge), we combine several key mathematical tools into a single, fully automated Matlab processing pipeline, which we call the Chaos Decision Tree Algorithm (Figure 3.1). The Chaos Decision Tree Algorithm takes a single time-series of any type - be it recordings of neural spikes, time-varying transcription levels of a particular gene, fluctuating oil prices, or recordings of stellar flux - and classifies those recorded data as coming from a system that is predominantly (or "operationally" [154]) stochastic, periodic, or chaotic. The algorithm requires no input from the user other than a time-series recording, though we have structured our code such that users can also select from among a number of alternative subroutines (see Methods, Figure 3.1).

In this paper, we show that the Chaos Decision Tree Algorithm performs with very high accuracy across a wide variety of both real and simulated systems, even in the presence of relatively high levels of measurement noise. Moreover, our pipeline can accurately track changing degrees

of chaos (for e.g. transitions from weak to strong chaos). With an eye toward applications to biology, the simulated systems we tested included a high-dimensional mean field model of cortical electrodynamics, a model of a spiking neuron, a model of white blood cell concentrations in chronic granulocytic leukemia, and a model of the transcription of the NF-κB protein complex. We also tested the algorithm on a wide variety of non-biological simulations, including several difficult edge cases; these included strange non-chaotic systems, quasi-periodic systems, colored noise, and nonlinear stochastic systems (see Supplementary Glossary for definitions of these terms), which are all classically difficult to distinguish from chaotic systems [156–158, 299]. We also tested the algorithm on a hyperchaotic system (Supplementary Glossary), which can be difficult to distinguish from noise [157], as well as on several non-stationary processes (Supplementary Glossary) in order to test the robustness of the algorithm against non-stationarity. Finally, we tested the Chaos Decision Tree Algorithm on several empirical (i.e. non-simulated) datasets for which the ground-truth presence or absence of chaos has been previously established by other studies. These included an electronic circuit in periodic, strange non-chaotic, and chaotic states [322], a chaotic laser [160], the stellar flux of a strange non-chaotic star [161], the linear/stochastic North Atlantic Oscillation index [162], and nonlinear/stochastic Parkinson's and essential tremors [158]. Overall, our pipeline performed with near-perfect accuracy in classifying these data as stochastic, periodic, or chaotic, as well as in tracking changing degrees of chaos in both real and simulated systems. Finally, we applied our algorithm to electrocardiogram recordings from healthy subjects, patients with congestive heart failure, and patients with atrial fibrillation [163], and provide evidence that heart rate variability reflects a predominantly stochastic, rather than chaotic process.

We have made our Matlab code freely and publicly available at
`https://figshare.com/s/80891dfb34c6ee9c8b34`.

3.3 Results

The Chaos Decision Tree Algorithm is depicted graphically in Figure 3.1. The pipeline consists of four steps: 1) Determine if the data are stochastic using permutation entropy [164] and a combination of Amplitude Adjusted Fourier transform surrogates [165, 166] and Cyclic Phase Permutation surrogates [166, 167] (Supplementary Glossary), 2) De-noise the data using the Schreiber de-noising algorithm [168] (Supplementary Glossary), 3) Correct for possible signal oversampling, and 4) Test for chaos using a modified 0-1 test for chaos [295–299] (Supplementary Glossary). For each step of the processing pipeline, we compared the performance of different available tools (i.e. different surrogate-based tests for stochasticity, different de-noising methods,

different downsampling methods, and different chaos-detection methods), and chose the tools with the highest classification performance (Supplementary Tables 3.1-3.14). Note that with user input, the Chaos Decision Tree Algorithm can use any of the alternative tools tested here, and that with no user input other than a time-series recording, the algorithm will automatically use the tools we found maximized its performance. All results reported in the main body of this paper are for this automated set of high-performing tools. See Supplementary Figure 3.3 for example time-traces illustrating each step of the algorithm.

We tested the performance of the (automated) Chaos Decision Tree Algorithm in detecting the presence and degree of chaos in a wide range of simulated and empirical systems for which the ground-truth presence of chaos, periodicity, or stochasticity has already been established. Details about each dataset and how the ground-truth presence or absence of chaos in those systems was previously determined are included in the Methods. Note that some systems are labeled "SNA," which is an abbreviation for "strange non-chaotic attractor" (Supplementary Glossary). These are systems whose attractors in phase space (Supplementary Glossary) are fractal (like chaotic systems), but which are periodic (i.e. non-chaotic). Among these, we included the only known non-artificial strange non-chaotic system, the stellar flux of the so-called golden star KIC 5520878, as recorded by the Kepler space telescope [161]. All simulated datasets consisted of 10,000 time-points, and all initial conditions were randomized. For systems with more than one variable, we here report results for linear combinations of those variables (see Methods), under the assumption that in most real-life cases, empirical recordings will contain features of multiple components of the system of study; that said, we also confirmed that the Chaos Decision Tree Algorithm has very high performance for individual system variables (Supplementary Table 3.15).

Results for simulations of biological systems are reported in Table 3.1, and results for non-biological simulations are reported in Table 3.2. Note that no measurement noise was added to the colored noise signals in Table 3.2, as doing so would flatten their power spectra. Because the datasets in Tables 3.1 and 3.2 were used to choose between alternative methods for detecting stochasticity (Supplementary Tables 3.1-3.4), de-noising (Supplementary Table 3.5), downsampling (Supplementary Table 3.6), and alternative tests of chaos (Supplementary Tables 3.8-3.13), as well as to optimize the 0-1 test for chaos (Supplementary Figures 3.4-3.6), we further tested the Chaos Decision Tree Algorithm on held out datasets, which were not used to adjudicate between alternative tools. These held out datasets included both simulated systems (Table 3.3) and recordings from real (non-simulated) systems (Table 3.4). Several of these held out datasets were of direct biological relevance: the periodically stimulated Poincaré oscillator in Table 3.3 is thought to be a good model of cardiac cell electrodynamics [173], which, like the

Poincaré oscillator, are chaotic when periodically stimulated with certain delays between stimulation pulses [149]; the integrated circuit in Table 3.4 is a physical implementation of equations that are based on the Hodgkin-Huxley neuron model [174]; and the tremor signals in Table 3.4 are direct recordings from patients. The Chaos Decision Tree Algorithm classified the systems in Tables 3.1-3.4 as stochastic, periodic, or chaotic with near-perfect accuracy even at high levels of measurement noise, with the exception of the noise-driven sine map (Table 3.2) - see Discussion. Finally, we tested the performance of the Chaos Decision Tree Algorithm on sub-samples of all systems in Tables 3.1-3.3, and confirmed that it is still highly accurate for data with just 1,000 time-points (Supplementary Table 3.16) or 5,000 time-points (Supplementary Table 3.17), though we note that performance for some systems did go down with less data, which is to be expected [298].

Table 3.5 reports the accuracy of the Chaos Decision Tree Algorithm in detecting degree of chaos. Formally, a system's degree of chaos is quantified by the magnitude of its largest Lyapunov exponent (Supplementary Glossary). Unfortunately, largest Lyapunov exponents are very difficult to estimate from finite, noisy time-series recordings. But, directly estimating largest Lyapunov exponents may not be necessary for tracking changing degrees of chaos in real systems: following prior observations of a strong correlation between a quick-to-compute and noise-robust measure called permutation entropy (Supplementary Glossary) and the largest Lyapunov exponents of several systems [164, 175], the Chaos Decision Tree Algorithm approximates degree of chaos by calculating the permutation entropy of the inputted signal, after it has been de-noised and corrected for possible over-sampling. In agreement with prior findings, we found that permutation entropy tracked degree of chaos in the logistic map, the Hénon map, the Lorenz system, a high-dimensional mean field model of the cortex, and an electronic circuit. See Methods for details on the parameters that were used to generate dynamics with different degrees of chaos in these systems, and for details on how ground-truth largest Lyapunov exponents were calculated. Note that without downsampling, the correlation between largest Lyapunov exponents and permutation entropy breaks down in continuous systems (Supplementary Table 3.14), which is to be expected, as permutation entropy has only been analytically proven to track degree of chaos in discrete-time systems [164, 176] (see Supplementary Glossary).

Finally, as a first-pass implementation of our method, we applied the Chaos Decision Tree Algorithm to recordings of human heart rate variability, made available by Physionet [163]. There has been considerable debate over whether or not the irregularities of heart rate signals (in either health or disease) reflect a predominantly chaotic process. While many classic chaos-detection methods have identified heart rate variability as chaotic (see Glass [138] for a review), other studies have argued that this is an erroneous classification, suggesting that heart rate

variability is in fact a nonlinear stochastic process [177, 178], and that prior classifications of heart rate signals as chaotic simply reflect the shortcomings of classic chaos-detection methods. In agreement with this view, we here show that the Chaos Decision Tree Algorithm classified heart rate signals from healthy subjects, congestive heart failure patients, and atrial fibrillation patients as stochastic, rather than chaotic, with the exception of two congestive heart failure patients (Table 3.6).

3.4 Discussion

In this paper, we have introduced a processing pipeline, called the Chaos Decision Tree Algorithm, that can accurately detect whether a time-series signal is generated by a predominantly stochastic, periodic, or chaotic system, and can also accurately track changing levels of chaos within a system using permutation entropy. The pipeline makes no assumptions about the input data. The Chaos Decision Tree Algorithm consists of four broad steps: 1) testing for stochasticity using surrogate data methods, 2) de-noising, 3) downsampling if data are over-sampled, and 3) testing for chaos using the modified 0-1 test. We tested the performance of several different surrogate data generation algorithms, de-noising algorithms, downsampling algorithms, and parameters for the modified 0-1 test. Each alternative algorithm and parameter choice has its relative strengths and weaknesses, and we have structured our code such that a user can specify which algorithms and parameters to use for each step of the pipeline. If a user only inputs a time-series recording without specifying any sub-algorithms or parameters, then our pipeline will automatically use the methods and parameters we found yielded the most accurate results across a large and diverse set of data. All analyses reported in the main body of this paper are for this automated set of subroutines.

We tested the (automated) Chaos Decision Tree Algorithm on a diverse range of simulations of biological systems, non-biological simulations, and empirical (non-simulated) data recordings. Empirical data were recorded from an integrated circuit in a periodic, strange non-chaotic, and chaotic state, a chaotic laser, the stellar flux of a strange non-chaotic star, the North Atlantic Oscillation index, a Parkinson's tremor, an essential tremor, and heart rate variability from congestive heart failure patients, atrial fibrillation patients, and healthy controls. In the cases for which the ground-truth was known (i.e. all datasets other than heart rate variability), the Chaos Decision Tree Algorithm performed at very high accuracy even at relatively high levels of measurement noise. For heart rate variability, our results support the hypothesis that cardiac rhythm variability is stochastic [177, 178]. Overall, these findings make us confident that the

Chaos Decision Tree Algorithm can be fruitfully applied to biological and non-biological signals contaminated by measurement noise.

We note a few limitations/shortcomings of our algorithm. First, the 0-1 test used in our pipeline might classify some very weakly chaotic systems (i.e. systems whose largest Lyapunov exponent is positive but very near zero) as periodic if the length of the time-series provided is short; but, with longer time-series, the test is guaranteed to provide accurate results [298]. We also note that the algorithm performed poorly for the noise-driven sine map, which was consistently misclassified as chaotic (Table 3.2). It is possible that this system was not classified as stochastic because its level of intrinsic noise was very low; in support of this, we found that the Chaos Decision Tree Algorithm classified nonlinear dynamical systems with very low levels of intrinsic noise as deterministic, and that classifications of stochasticity became more frequent as the level of intrinsic noise was increased (Supplementary Table 3.18). It is also possible that this system is in fact an example of noise-induced chaos [310]. Finally, although the choice of system observables did not appreciably affect the performance of our method (Supplementary Table 3.15), we agree with Letellier and colleagues [180] that some system observables are better representations of a system's dynamics than others, and that this can have important consequences for the accuracy of nonlinear time-series analysis methods such as this one. In light of these potential limitations, it bears re-emphasizing that the absence, presence, and degree of chaos can only be determined with absolute certainty in a computer model that is free of measurement noise, by running multiple simulations and seeing how quickly initially similar states diverge. Thus, although the Chaos Decision Tree Algorithm pipeline performs at very high accuracy, it should, when possible, be used in conjunction with analyses of a computer model of the system at hand.

We hope that the Chaos Decision Tree Algorithm will help advance the decades-old effort to bring the insights of chaos theory to biology. While a diverse range of biological simulations and a small number of real biological cases have been shown to be chaotic, detecting the presence and degree of chaos in biological recordings has been difficult. The Chaos Decision Tree Algorithm overcomes the difficulties of prior tests, by being fast, highly robust to measurement noise, and, unless the user specifies specific alternative subroutines, fully automated. We welcome any efforts to identify edge cases for which our pipeline systematically breaks down; given that our pipeline is a modular decision tree, new subroutines can be added to accommodate such edge cases. We hope that our pipeline (and perhaps future iterations of it) will be useful to any of the domains of science - and in particular of biology - in which chaos has been invoked, but not tested.

We thank Seiji Uenohara and colleagues for kindly sending us data recorded from their chaotic neuron integrated circuit, and we thank John F. Lindner for sending us data on the stellar flux of KIC 5520878. We also thank Michael Jansson for his feedback on stationarity testing.

The authors declare that they have no competing interests.

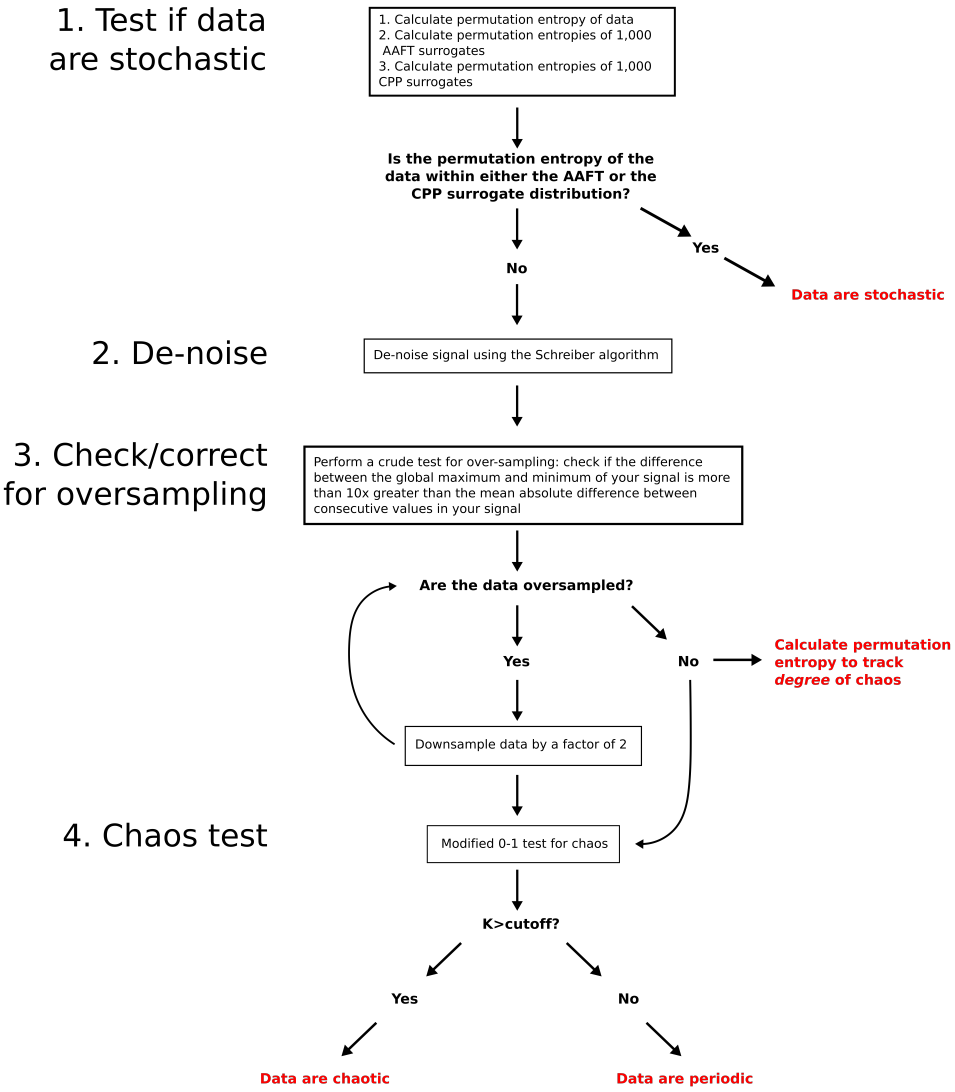


FIGURE 3.1: Caption on the following page.

Figure 3.1 (continued). The Chaos Decision Tree Algorithm. The first step of the algorithm is to test if data are stochastic. The Chaos Decision Tree Algorithm uses a surrogate-based approach to test for stochasticity, by comparing the permutation entropy of the original time-series to the permutation entropies of random surrogates of that time series. If the user does not specify which surrogate algorithms to use, the Chaos Decision Tree Algorithm automatically picks a combination of Amplitude Adjusted Fourier Transform [165] surrogates and Cyclic Phase Permutation [167] surrogates - see Supplementary Tables 3.2-3.3. If the permutation entropy of the original time-series falls within the distributions of permutation entropies of the random surrogates, then the time-series is classified as being generated by a stochastic system; if the permutation entropy falls outside the surrogate distributions, then the algorithm proceeds to denoise the inputted signal. Several de-noising subroutines are available, but if the user does not specify a de-noising algorithm, the pipeline will automatically use Schreiber's noise-reduction algorithm [168] (Supplementary Table 3.5). The pipeline then checks for signal oversampling; if data are oversampled, the pipeline iteratively downsamples the data until they are no longer oversampled (note that an alternative downsampling method proposed by Eyébé Fouda and colleagues [300] may be selected instead - see Supplementary Table 3.6). Finally, the Chaos Decision Tree Algorithm performs the 0-1 chaos test on the input data, which has been modified from the original 0-1 test to be less sensitive to noise, to suppresses correlations that arise from quasi-periodicity, and to normalize the standard deviation of the test signal (see Methods). The value for the parameter that suppresses signal correlations can be specified by the user, but is otherwise automatically chosen based on ROC analyses performed here (Supplementary Figure 3.4). The modified 0-1 test provides a single statistic, K, which approaches 1 for chaotic systems and approaches 0 for periodic systems. Any desired cutoff for K may be inputted to the pipeline, and if no cutoff is provided, the pipeline will use a cutoff based on the length of the time-series (Supplementary Figure 3.6). If K is greater than the cutoff, the data are classified as chaotic, and if they are less than or equal to the cutoff, they are classified as periodic.

TABLE 3.1: The Chaos Decision Tree Algorithm classified biological simulations as periodic or chaotic with near-perfect accuracy.

	Measurement noise level (% of std. dev.)				
System	0%	10%	20%	30%	40%
Cortical model [276] (chaotic)	100/100	100/100	100/100	97/100	83/100
Cortical model [276] (periodic)	100/100	100/100	100/100	100/100	100/100
Spiking neuron [181] (chaotic)	100/100	100/100	100/100	100/100	98/100
Granulocyte levels [152] (chaotic)	100/100	100/100	100/100	100/100	100/100
Granulocyte levels [152] (periodic)	100/100	100/100	100/100	100/100	83/100
NF-κB transcription [147] (chaotic)	99/100	99/100	100/100	100/100	100/100
NF-κB transcription [147] (periodic)	100/100	100/100	97/100	100/100	100/100

TABLE 3.2: The Chaos Decision Tree Algorithm classified non-biological simulations as stochastic, periodic, or chaotic with high accuracy. These simulated systems include strange non-chaotic attractors ("SNAs"), linear stochastic processes, and nonlinear stochastic processes, all of which are classically difficult to distinguish from chaos.

	Measurement noise level (% of std. dev.)				
System	0%	10%	20%	30%	40%
Cubic map [182] (chaotic)	100/100	100/100	100/100	100/100	100/100
Cubic map [182] (periodic)	100/100	100/100	100/100	100/100	100/100
Cubic map [182] (SNA HH)	100/100	100/100	100/100	100/100	100/100
Cubic map [182] (SNA S3)	100/100	100/100	100/100	100/100	0/100
GOPY map [183] (SNA)	100/100	100/100	100/100	99/100	14/100
Logistic map [184] (chaotic)	100/100	100/100	100/100	100/100	100/100
Logistic map [184] (periodic)	100/100	100/100	100/100	100/100	100/100
Lorenz system [185] (chaotic)	100/100	100/100	97/100	82/100	36/100
Generalized Hénon map [186] (hyperchaotic)	100/100	100/100	100/100	100/100	93/100
Freitas map [187] (nonlinear stochastic)	78/100	83/100	98/100	98/100	74/100
Noise-driven sine map [187] (nonlinear stochastic)	55/100	3/100	22/100	5/100	78/100
Bounded random walk [188] (nonlinear stochastic)	100/100	97/100	59/100	95/100	100/100
Cyclostationary process [189] (linear stochastic)	99/100	100/100	99/100	100/100	100/100
ARMA process (linear stochastic)	85/100	98/100	99/100	99/100	100/100
Trended random walk (linear stochastic)	100/100	89/100	90/100	98/100	100/100
Random walk (linear stochastic)	100/100	98/100	100/100	100/100	100/100
Violet noise [190] (linear stochastic)	99/100				
Blue noise [190] (linear stochastic)	100/100				
White noise [190] (linear stochastic)	100/100				
Pink noise [190] (linear stochastic)	100/100				
Red noise [190] (linear stochastic)	100/100				

TABLE 3.3: Classification accuracy in held out simulated systems. While the datasets in Tables 3.1-3.2 were used to optimize the Chaos Decision Tree Algorithm, these datasets were not. Performance was near-perfect.

	Measurement noise level (% of std. dev.)				
System	0%	10%	20%	30%	40%
Rössler system [191] (chaotic)	70/100	90/100	96/100	100/100	100/100
Ikeda map [192] (chaotic)	100/100	100/100	100/100	100/100	14/100
Hénon map [193] (periodic)	100/100	100/100	100/100	100/100	100/100
Cubic map [193] (period-doubled)	100/100	100/100	100/100	100/100	100/100
Poincaré oscillator [173] (periodic)	100/100	100/100	100/100	100/100	57/100
Poincaré oscillator [173] (quasi-periodic)	100/100	100/100	100/100	100/100	100/100
Poincaré oscillator [173] (chaotic)	100/100	100/100	100/100	100/100	100/100
Multivariate AR model (linear stochastic)	100/100	100/100	100/100	100/100	100/100

TABLE 3.4: Classification accuracy in empirical (non-simulated) data from stochastic, periodic, strange non-chaotic ("SNA"), and chaotic systems. As was the case for the data in Table 3.3, these datasets were not used to optimize algorithm performance, and so are also held out datasets. Algorithm performance was perfect.

System	
Neuron integrated circuit [322] (chaotic)	10/10
Neuron integrated circuit [322] (SNA)	10/10
Neuron integrated circuit [322] (periodic)	10/10
Laser [160] (chaotic)	1/1
Stellar flux [161] (SNA)	1/1
North Atlantic Oscillation index [162] (linear stochastic)	1/1
Essential tremor [158] (nonlinear stochastic)	1/1
Parkinson's tremor [158] (nonlinear stochastic)	1/1

TABLE 3.5: The Chaos Decision Tree Algorithm uses permutation entropy, calculated from data that have been de-noised and downsampled (if oversampled), to track the degree of chaos in a system, which might change as the state of the system changes. We here show Spearman correlations between permutation entropy and largest Lyapunov exponents, which measure degree of chaos but which are difficult to estimate from empirical data. Data include four simulated systems and recordings from an integrated circuit in different states. See Methods for how ground-truth largest Lyapunov exponents were calculated for these systems. *** $p<0.001$ (two-tailed) after Bonferroni-correcting for multiple comparisons to the same set of ground-truth largest Lyapunov exponents.

	Measurement noise level (% of std. dev.)				
System	0%	10%	20%	30%	40%
Logistic map [184]	0.93***	0.80***	0.93***	0.93***	0.91***
Hénon map [193]	0.92***	0.93***	0.94***	0.92***	0.88***
Lorenz system [185]	0.81***	0.71***	0.73***	0.69***	0.62***
Cortical model [276]	0.69***	0.55***	0.39***	0.33***	0.24***
Neuron integrated circuit [322]	0.94***				

TABLE 3.6: The Chaos Decision Tree Algorithm consistently classifies heart rate recordings, across conditions, as stochastic. The only exceptions were the heart rate signals recorded from two patients with congestive heart failure, which were classified as periodic.

	Classification		
Condition	Stochastic	Periodic	Chaotic
Healthy controls [163]	5/5	0/5	0/5
Congestive heart failure [163]	3/5	2/5	0/5
Atrial fibrillation [163]	5/5	0/5	0/5

3.5 Methods

3.5.1 The Chaos Decision Tree Algorithm

To understand the logic of the Chaos Decision Tree Algorithm, we begin with the final test in the decision tree. The crux of the Chaos Decision Tree Algorithm is the 0-1 test for chaos. The 0-1 test for chaos was originally developed by Gottwald and Melbourne [295], who later offered a slightly modified version of the test, which can cope with moderate amounts of measurement noise [296]. Several years later, Dawes and Freeland further modified the test, such that it could suppress correlations induced by quasi-periodic dynamics, and thus more effectively distinguish between chaotic and strange non-chaotic dynamics, which are difficult to distinguish given only a time-series recording [299]. The modified 0-1 test involves taking a one-dimensional time-series of interest $\boldsymbol{\phi}$, and using it to drive the following two-dimensional system:

$$\begin{aligned} p(n+1) &= p(n) + \phi(n)\cos cn \\ q(n+1) &= q(n) + \phi(n)\sin cn \end{aligned} \tag{3.1}$$

where c is a random value bounded between 0 and 2π. For a given c, the solution to Eq. 1 yields:

$$\begin{aligned} p_c(n) &= \sum_{j=1}^{n} \phi(j)\cos jc \\ q_c(n) &= \sum_{j=1}^{n} \phi(j)\sin jc \end{aligned} \tag{3.2}$$

Gottwald and Melbourne show that if the inputted time-series $\boldsymbol{\phi}$ is regular, the motion of $\mathbf{p}$ and $\mathbf{q}$ is bounded, while $\mathbf{p}$ and $\mathbf{q}$ display asymptotic Brownian motion if $\boldsymbol{\phi}$ is chaotic. The time-averaged mean square displacement of $\mathbf{p}$ and $\mathbf{q}$, plus the noise term proposed by Dawes and Freeland [299], is

$$M_c(n) = \frac{1}{N}\sum_{j=1}^{N}([p_c(j+n) - p_c(j)]^2 + [q_c(j+n) - q_c(j)]^2) + \sigma\eta_n. \tag{3.3}$$

where η_n is a uniformly distributed random variable between $[-\frac{1}{2}, \frac{1}{2}]$ and σ is the noise level. Finally, the outputted K-statistic of the 0-1 test uses a correlation coefficient to measure the growth rate of the mean squared displacement of the two-dimensional system in Eq. 1:

$$K_c = \mathrm{corr}(n, M_c(n)) \tag{3.4}$$

K is computed for 100 different values of c, randomly sampled between 0 and 2π, and the final output of the test is the median K across different values of c. For chaotic systems, this median K value will approach 1, and for periodic systems, K will approach 0 [295–299].

There are two parameters in this modified 0-1 test: the parameter σ that controls the level of added noise in Eq. 3, and the cutoff for what K-statistic values are classified as indicating chaos or periodicity in a finite time-series. We performed ROC-curve analyses for different values of σ and found that $\sigma = 0.5$ maximized classification performance across systems and noise levels (Supplementary Figure 3.4), and so our pipeline automatically sets σ to 0.5 if σ is not specified by the user. Note that for non-zero values of σ, K approaches zero as the standard deviation of a test signal approaches zero (Supplementary Figure 3.5), and so the Chaos Decision Tree Algorithm multiplies a test signal by a constant to fix its standard deviation at 0.5 before applying the 0-1 test. A cutoff for K can also be inputted to our Matlab script, such that data that yield a K value greater than that cutoff are classified as chaotic and data that yield a K value less than or equal to that cutoff are classified as periodic. If no cutoff is provided, a cutoff is chosen based on an analysis of optimal cutoffs as a function of time-series length (Supplementary Figure 3.6). If the automatically selected cutoff is greater than 0.99, the cutoff is set to $K = 0.99$, as K is upper-bounded by 1. We have confirmed that this automated cutoff selection yields highly accurate results for sub-samples of both test and held-out datasets (Supplementary Tables 3.16-3.17).

The 0-1 test described above only yields accurate results for data that are deterministic [156, 195, 298, 371]. A system is considered deterministic if, given the exact same initial conditions, it always evolves over time the same way, whereas a system is considered stochastic if there is appreciable randomness built in to its evolution over time (Supplementary Glossary, Supplementary Figures 3.1-3.2). Not only are all chaotic systems (predominantly) deterministic - and thus the possibility of chaos can be automatically rejected if a system is found to be stochastic (though we note that a mathematically rigorous definition of chaos has recently been extended to the domain of stochastic systems, under the framework of the Supersymmetric Theory of Stochastics [310]) - but the 0-1 test is also known to incorrectly classify stochastic dynamics as chaotic [156, 195, 371]. Thus, the Chaos Decision Tree Algorithm first rules out the possibility that data are predominantly stochastic before applying the modified 0-1 test. To do so, it uses a noise-robust method recently developed by Zunino and Kulp [196], which tests for determinism using surrogate statistics [165], with permutation entropy [164] as the test statistic. The calculation of permutation entropy relies on two parameters: permutation order and time-lag. We follow the recommendation from Bandt and Pompe [164] and set the time-lag to 1, and found that a permutation order of 8 maximized stochasticity detection accuracy

(Supplementary Tables 3.2-3.3). Moreover, we use a combination of amplitude adjusted Fourier transform surrogates [165] and Cyclic Phase Permutation surrogates [167], unlike Zunino and Kulp, who used iterative amplitude adjusted Fourier transform [165] surrogates, because we found that this combination led to far higher classification accuracy (Supplementary Tables 3.2-3.3). The Chaos Decision Tree Algorithm classifies data as stochastic (and thus does not proceed to subsequent steps) if the permutation entropy of the original data falls within either surrogate distribution. The algorithm uses the Toolboxes for Complex Systems implementation of the permutation entropy algorithm, written by Andreas Müller [197]. Surrogates are generated using the Matlab toolbox recently released by Lancaster and colleagues [166]. Note that because Fourier-based surrogates are strictly stationary, surrogate-based tests that use only Fourier-based algorithms are only valid if the test time-series is also stationary [166, 189]; that said, we found that non-stationarity did not affect the accuracy of a stochasticity test that uses a combination of amplitude adjusted Fourier transform and Cyclic Phase Permutation surrogates (Supplementary Tables 3.1-3.4). We also did not find that a normality transformation of the data improved the performance of our surrogate-based stochasticity test (Supplementary Table 3.2), counter to what has been suggested elsewhere [154].

If data "pass" the stochasticity test described above and are deemed operationally deterministic, then the Chaos Decision Tree Algorithm automatically denoises the inputted signal. We compared three de-noising algorithms: a moving average filter (using Matlab's smooth.m function), the Matlab Chaotic Systems Toolbox's [198] implementation of Schreiber's noise-reduction algorithm [168] (Box 1), and wavelet de-noising using an empirical Bayesian method with a Cauchy prior (using Matlab's wdenoise.m function). Although it is considerably slower to run, Schreiber denoising markedly outperforms the other two approaches in recovering the deterministic component of signals contaminated by measurement noise (Supplementary Table 3.5), and markedly improves the performance of the modified 0-1 test (Supplementary Table 3.6, Supplementary Figure 3.4). Thus, the Chaos Decision Tree Algorithm automatically uses Schreiber de-noising before testing for chaos, unless the user specifies one of the other two de-noising algorithms tested here to be used instead.

The final step of the Chaos Decision Tree Algorithm before applying the 0-1 test is to check if data are oversampled and to downsample them if they are. Gottwald and Melbourne have shown [297] that the 0-1 test can give inaccurate results for continuous (i.e. non-discrete-time) systems sampled at a very high frequency, but that it can accurately differentiate between periodic dynamics and chaotic dynamics in continuous deterministic systems when data are properly downsampled. In light of this, the Chaos Decision Tree Algorithm utilizes the (crude) test for oversampling used by Matthews [199], by calculating a measure η, which is the difference

between the global maximum and global minimum of the data divided by the mean absolute difference between consecutive time-points in the data. If $\eta > 10$, then the data are deemed to be oversampled, and the Chaos Decision Tree Algorithm iteratively downsamples the data by a factor of 2 until $\eta \leq 10$ or until there are fewer than 100 time-points left in the signal. We compared this approach both to no downsampling and to an alternative method, suggested by Eyébé Fouda and colleagues [300] to improve 0-1 test performance, which downsamples by taking just the local minima and maxima of oversampled signals. We found that downsampling after de-noising yields more accurate results than either alternative approach when oversampled signals are contaminated by measurement noise (Supplementary Table 3.6). We also note that recorded experimental data may be unlikely to be oversampled (Supplementary Table 3.7), and that this problem may be more likely to arise in simulated continuous systems. If the data are not oversampled, or if they have been downsampled, the Chaos Decision Tree Algorithm then applies the modified 0-1 test to the data, as described above.

Finally, the algorithm uses the permutation entropy of the inputted signal as a proxy for the degree of chaos in the system. Though the algorithm uses permutation entropy to establish whether or not a signal is predominantly deterministic (see above), permutation entropy has also been shown to tightly track the largest Lyapunov exponent (and therefore the degree of chaos) of the logistic map [164], the tent map [201], and the Duffing oscillator [175]. We should in general expect a close correspondence between permutation entropy and Lyapunov exponents, in light of the equivalence in discrete-time systems between permutation entropy and Kolmogorov-Sinai entropy [176, 202–204], which is upper-bounded by the sum of a system's positive Lyapunov exponents - a relationship known as the "Pesin identity" [205]. When calculating permutation entropy to track degree of chaos (rather than for determinism testing as above), we follow Bandt and Pompe's [164] recommendation and simply set the time-lag to 1 and the permutation order to 5, which we showed tracks degree of chaos in all systems tested (Table 3.5). Because this equivalence is only known to hold for discrete-time systems [176], permutation entropy is only calculated after the inputted signal has been de-noised and, if oversampled, downsampled; this considerably improves its ability to track degree of chaos in continuous systems (Table 3.5, Supplementary Table 3.14).

The full decision tree of our algorithm is depicted graphically in Figure 3.1.

3.5.2 Biological simulations

The following describes the simulations of biological systems analyzed in this paper. We only picked biological simulations for which the presence or absence of chaos has been established in

prior work. Initial conditions were randomized in all simulations. We also tested the effect of measurement noise on the accuracy of the Chaos Decision Tree Algorithm in classifying systems, by adding white noise to our simulated data, the amplitude of which was up to 40% the standard deviation of the original data. For each simulated system and level of measurement noise, we created 100 datasets with 10,000 time points.

Chaotic mean field cortical model. Steyn-Ross, Steyn-Ross, and Sleigh [276] describe a mean-field model of the cortex based on the equations first introduced by Liley and colleagues [340, 341], which includes electrical gap-junction synapses in addition to the standard chemical synapses used in the earlier models. The model contains both inhibitory and excitatory neural populations communicating locally through gap junctions and chemical synapses and communicating over long ranges via myelinated axons. The dynamics of each neural population in the model are determined by two first-order and six second-order partial differential equations, which is equivalent to 14 first-order differential equations. The primary output of the model is the mean excitatory firing rates of 120 neural populations, which approximates the large-scale cortical signals that might be measured through electrocortigraphy, magnetoencephalography, or electroencephalography. Steyn-Ross, Steyn-Ross, and Sleigh [276] show that just by varying the inhibitory gap-junction diffusive-coupling strength parameter in their model, they can produce dynamics ranging from periodicity to strong chaos. In their simulation of "waking" cortical dynamics, Turing (spatial) and Hopf (temporal) instabilities interact to produce chaotic, low-frequency spatiotemporal oscillations. For chaotic dynamics, we simulated 2,000,000 time-points of their "wake" simulation, with the inhibitory gap-junction diffusive-coupling strength parameter set to 0.4, and then downsampled the data to 10,000 time-points. We only applied our algorithm to the mean excitatory firing rate of one neural population, i.e. to just one out of 14 variables describing the dynamics of just one out of 120 interacting such 14-dimensional systems (though the variable is biologically well-defined). The Matlab code for the simulations is available in the Supplementary Material of Steyn-Ross, Steyn-Ross, and Sleigh [276].

Periodic mean field cortical model. Steyn-Ross, Steyn-Ross, and Sleigh show that their cortical mean field model enters a periodic, seizure-like state dominated by a Hopf instability when the inhibitory gap-junction diffusive-coupling strength parameter to 0.1. Just as in the chaotic case, we simulated 2,000,000 time-points and then down-sampled to 10,000 time-points. Note that Steyn-Ross, Steyn-Ross, and Sleigh estimate the largest Lyapunov exponent of their model to be around zero when the inhibitory gap-junction

diffusive-coupling strength parameter is 0.1, whereas our own estimate (using an automated version of their same method - see below) placed the largest Lyapunov exponent more clearly in the periodic regime, at -2.1.

Chaotic spiking neuron. Izhikevich [181, 208] describes a simple neuron model that can display both spiking and bursting behavior. The model consists of a neuron's membrane potential v, a membrane recovery variable u, an input current I, and parameters a, b, c, and d:

$$\begin{aligned} \frac{dv}{dt} &= 0.004v^2 + 5v + 140 - u + I \\ \frac{du}{dt} &= a(bv - u) \end{aligned} \tag{3.5}$$

with the auxiliary after-spike resetting:

$$\text{if } v \geq +30 \text{ mV}, \quad \text{then } \begin{cases} v \leftarrow c \\ u \leftarrow u + d. \end{cases} \tag{3.6}$$

When $a = 0.2$, $b = 2$, $c = -56$, $d = -16$, and $I = -99$, the neuron's membrane potential v (which is the variable we analyze) displays chaotic spikes [181, 208]. We simulated the Izhikevich neuron using a first-order Euler method, with an integration step of 0.25 ms. We generated 50,000 time points, and dowsampled by a factor of 5 (to avoid over-sampling).

Periodic white blood cell concentration. Inspired by the finding that chronic granulocytic leukemia involves apparently aperiodic oscillations in the concentration of circulating white blood cells [209], Mackey and Glass [152] study mathematical models of oscillating physiological control systems. They describe a simple mathematical model of the concentration of circulating white blood cells:

$$\frac{dx}{dt} = a\frac{x_\tau}{1 + x_\tau^c} - bx \tag{3.7}$$

where $a = 0.2$, $b = 0.1$, and $c = 10$. The parameter τ represents the delay between white blood cell production in bone marrow and the release of those cells into the blood stream. Since this cellular generation delay time is increased in some patients with chronic granulocytic leukemia, Mackey and Glass study the behavior of this system as a function of the delay time τ. They find that as τ increased, the resulting oscillations produced by this equation became aperiodic. Through formal analysis of Lyapunov exponents of this system, Farmer [210] later confirmed that for $\tau = 10$, the oscillations of this system are periodic. We simulated 100,000 time-points of the periodic Mackey-Glass system using a

first-order Euler method with an integration step of 1, and then downsampled by a factor of 10 (to avoid over-sampling).

Chaotic white blood cell concentration. For $\tau = 30$, Farmer confirmed [210] that the Mackey-Glass equation (Eq. 7) for the concentration of circulating white blood cells yields a chaotic oscillation. We simulated 100,000 time-points of the chaotic Mackey-Glass system using a first-order Euler method with an integration step of 1, and then downsampled by a factor of 10 (to avoid over-sampling).

Periodic NF-κB transcription. Heltberg and colleageus [147] recently described a five-dimensional mathematical model of oscillating concentrations of the transcription factor NF-κB, which regulates several genes involved in immune responses and is widely studied in immunity and cancer research. They show that the dynamics of NF-κB concentration are coupled to varying levels of a cytokin-like tumor necrosis factor (TNF). They show that when TNF oscillations have a low amplitude, NF-κB oscillations are periodic. We simulated periodic NF-κB oscillations using Heltberg and colleagues' Matlab code.

Chaotic NF-κB transcription. Heltberg and colleagues [147] show that by increasing the amplitude of the TNF signal, the oscillating number of NF-κB molecules in their model becomes chaotic. We simulated chaotic NF-κB oscillations using Heltberg and colleagues' Matlab code.

3.5.3 Non-biological simulations

Because there are only a limited number of biological simulations for which the presence of chaos has already been established, we also applied the Chaos Decision Tree Algorithm to a wide range of mathematical systems previously studied in the chaos theory and nonlinear time-series analysis literatures:

Chaotic cubic map. Venkatesan and Lakshmanan [182] describe a quasiperiodically forced cubic map, which can exhibit a large diversity of periodic, chaotic, and strange non-chaotic dynamics. In particular, the map exhibits many different routes to chaos. Their system is described by the following:

$$\begin{aligned} x_{i+1} &= Q + f\cos(2\pi\theta_i) - Ax_i + x_i^3 \\ \theta_{i+1} &= \theta_i + \omega(\text{mod } 1), \end{aligned} \tag{3.8}$$

where $\omega = \frac{\sqrt{5}-1}{2}$ (the golden ratio). We set $f = -0.8$, $Q = 0$, and $A = 1.5$, which Venkatesan and Lakshmanan have shown lead to chaotic dynamics [182]. For the results reported in Table 3.2, we followed Dawes and Freeland [299] in taking a linear combination of $\mathbf{x}$ and $\boldsymbol{\theta}$: $\phi_i = x_i/6 + \theta_i/10$. Results for $\mathbf{x}$ individually are reported in Supplementary Table 3.15 (results for $\boldsymbol{\theta}$ on its own are not informative, as $\boldsymbol{\theta}$ is an independent, quasi-periodic process).

Periodic cubic map. Venkatesan and Lakshmanan [182] show that the system in Eq. 8 exhibits periodic (one-frequency torus) dynamics when $f = 0$, $Q = 0$, and $A = 1$. We picked these parameters for periodic dynamics. To get a time-series $\boldsymbol{\phi}$ from the cubic map, we again took a linear combination of $\mathbf{x}$ and $\boldsymbol{\theta}$: $\phi_i = x_i/6 + \theta_i/10$. Results for $\mathbf{x}$ and $\boldsymbol{\theta}$ individually are reported in Supplementary Table 3.15.

Strange non-chaotic cubic map (HH). We set $f = 0.7$, $Q = 0$, and $A = 1.88697$ for one type of strange non-chaotic dynamics, which Venkatesan and Lakshmanan [182] have shown bring the forced cubic map into a strange non-chaotic regime via the Heagy-Hammel route (i.e., collision of a period-doubled quasi-periodic torus with its unstable parent). Results for $\mathbf{x}$ individually are reported in Supplementary Table 3.15.

Strange non-chaotic cubic map (S3). We set $f = 0.35$, $Q = 0$, and $A = 0.35$ for a second type of strange non-chaotic dynamics, which Venkatesan and Lakshmanan [182] have shown push the forced cubic map into a strange non-chaotic regime via Type-3 Intermittency (i.e., inverse period-doubling bifurcation). Results for $\mathbf{x}$ individually are reported in Supplementary Table 3.15.

Period-doubled cubic map. Venkatesan and Lakshmanan [182] show that the system in Eq. 8 exhibits period-doubled dynamics when $f = -0.18$, $Q = 0$, and $A = 1.1$. We picked these parameters for period-doubled dynamics. To get a time-series $\boldsymbol{\phi}$ from the cubic map, we again took a linear combination of $\mathbf{x}$ and $\boldsymbol{\theta}$: $\phi_i = x_i/6 + \theta_i/10$. Results for $\mathbf{x}$ individually are reported in Supplementary Table 3.15.

Strange non-chaotic GOPY model. The first known strange non-chaotic system was described by Grebogi, Ott, Pelikan, and Yorke, commonly referred to as the GOPY model [183]. The GOPY model is described by the following:

$$\begin{aligned} x_{i+1} &= 2\sigma(\tanh x_i)\cos(2\pi\theta_i) \\ \theta_{i+1} &= \theta_i + \omega (\mathrm{mod}\ 1), \end{aligned} \tag{3.9}$$

where $\sigma = 1.5$, $\theta = 0.5$, and $\omega = \frac{\sqrt{5}-1}{2}$ (the golden ratio). To get a time-series $\boldsymbol{\phi}$ from the GOPY model, we followed Dawes and Freeland [299] in taking a linear combination of $\mathbf{x}$ and $\boldsymbol{\theta}$: $\phi_i = x_i/6 + \theta_i/10$. Results for $\mathbf{x}$ individually are reported in Supplementary Table 3.15 (as is the case for the cubic map, results for $\boldsymbol{\theta}$ on its own are not informative, as $\boldsymbol{\theta}$ is an independent, quasi-periodic process).

Chaotic logistic map. The logistic map is one of the simplest known systems that can exhibit both periodic and chaotic behavior. It was originally introduced by biologist Robert May [184] as a discrete-time model of population growth. It is described by the following equation:

$$x_{i+1} = r x_i (1 - x_i) \tag{3.10}$$

where x_i represents the ratio of the population size at time i to the maximum possible population size. For chaotic dynamics [184], we set $r = 4$.

Periodic logistic map. For periodic dynamics [184] in the logistic map, we set $r = 3.5$ in Eq. 10.

Chaotic Lorenz system. Perhaps the most famous of all chaotic systems, the Lorenz model of atmospheric convection is described by the following system of equations [185]:

$$\begin{aligned} \frac{dx}{dt} &= \sigma(y - x) \\ \frac{dy}{dt} &= x(\rho - z) - y \\ \frac{dz}{dt} &= xy - \beta z \end{aligned} \tag{3.11}$$

where $\mathbf{x}$ is the rate of convection, $\mathbf{y}$ is the horizontal temperature variation, and $\mathbf{z}$ is the vertical temperature variation. Though the equations were initially meant to model atmospheric convection, identical equations have been found in models of a wide variety of physical systems, including lasers [211] and chemical reactions [212]. We set $\sigma = 10$, $\rho = 30$, and $\beta = \frac{8}{3}$, for which the Lorenz system exhibits chaos (determined by calculating the largest Lyapunov exponent of the system with these parameters, using Ramasubramanian's algorithm [213]). We integrated the equations for the Lorenz system using the Fourth Order Runge-Kutta method with an integration step of 0.01. To get a single time-series $\boldsymbol{\phi}$ from the Lorenz model, we took a linear combination of $\mathbf{x}$ and $\mathbf{y}$: $\boldsymbol{\phi}$=$\mathbf{x}$+$\mathbf{y}$. Results for $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ individually are reported in Supplementary Table 3.15.

Hyperchaotic generalized Henon map. Data from hyperchaotic systems, which contain more than one positive Lyapunov exponent, can be difficult to distinguish from noise

[157]. As such, hyperchaotic systems present a challenge to tests of determinism from time-series data, which might mistake hyperchaos for stochasticity. To demonstrate the robustness of the Chaos Decision Tree Algorithm's stochasticity test, we analyzed the Generalized Henon Map, which is described by the following equation:

$$x_{i+1} = a - x_{i-1}^2 - bx_{i-2} \tag{3.12}$$

We set $a = 1.76$ and $b = 0.1$, for which the Generalized Henon Map produces hyperchaos [186].

Noise-driven sine map. Freitas and colleagues [187] describe a non-chaotic, randomly driven system:

$$x_{i+1} = \mu \sin(x_i) + Y_i \eta_i \tag{3.13}$$

where $\mu = 2.4$, Y_i is a random Bernoulli process that equals 1 with probability 0.01 and 0 with probability 0.99, and η_i is a random variable uniformly distributed between -2 and 2. Freitas and colleagues show that a chaos-detection technique called "noise titration" [**?**] incorrectly classifies this system as chaotic.

Freitas map. Freitas and colleagues [187] also describe a nonlinear correlated noise process, which we here call the "Freitas map." The Freitas map contains dynamic noise added to a nonlinear moving average filter:

$$v_{i+1} = 3v_i + 4v_{i-1}(1 - v_i) \tag{3.14}$$

Freitas and colleagues show that the noise titration technique also incorrectly classifies this system as chaotic.

Bounded random walk. Nicolau [188] describes a bounded random walk (BRW), which is a globally stationary process with local unit roots (i.e. local non-stationarities):

$$X_t = X_{t-1} + e^k(e^{-\alpha_1(X_{t-1}-\tau)} - e^{\alpha_2(X_{t-1}-\tau)}) + \sigma_t \epsilon_t \tag{3.15}$$

where τ, k, α_1, α_2, and σ are parameters, and ϵ_t is a white noise error term. Note that the bounded random walk can be decomposed into a random walk, $X_t = X_{t-1} + \sigma_t \epsilon_t$, plus an adjustment function $e^k(e^{-\alpha_1(X_{t-1}-\tau)} - e^{\alpha_2(X_{t-1}-\tau)})$. The adjustment function serves to pull the random walk toward τ if the process deviates too far from τ. Though it is a stationary process (albeit with local non-stationarities), the bounded random walk is often mis-classified as non-stationary by stationarity tests [214]. Following Nicolau [188] and

Patterson [214], we set $\tau = 100$, $k = -15$, $\alpha_1 = 3$, $\alpha_2 = 3$, and $\sigma = 0.4$, which generates a random walk that remains roughly within the interval of 100±5.

Cyclostationary process. A cyclostationary autoregressive process is essentially a combination of a noise-driven linear damped oscillator and linear relaxators. Cyclostationary systems are non-stationary because their probability distributions vary cyclically with time. Following Timmer [189], we simulate a cyclostationary process described by the following:

$$X_t = a_1 X_{t-1} + a_2 X_{t-2} + \epsilon_t \tag{3.16}$$

where ϵ_t is a white noise error term and

$$\begin{aligned} a_1 &= 2\cos(2\pi/T)e^{-1/\tau} \\ a_2 &= -e^{-2/\tau} \end{aligned} \tag{3.17}$$

τ is the relaxation time and T is the oscillation period. We set $\tau = 50$ and $T = 10$, which Timmer has shown leads to incorrect classification of this system as nonlinear or deterministic by surrogate tests that only use Fourier-based surrogates, which are strictly stationary.

ARMA process. A general autoregressive moving-average (ARMA) process is described by the following:

$$X_t = c + \epsilon_t + \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} \tag{3.18}$$

where c is a constant, $\phi_1, ..., \phi_p$ and $\theta_1, ..., \theta_p$ are parameters, and $\epsilon_t, \epsilon_{t-1}, ...$ are white noise error terms. An ARMA process with a lag of 1, or an ARMA(1) process, is:

$$X_t = c + \epsilon_t + \phi X_{t-1} + \theta \epsilon_{t-1} \tag{3.19}$$

When $\phi < 1$, the ARMA process is (weakly) stationary. When ϕ is close to but less than 1 and $\theta \neq 0$, ARMA processes, though stationary, are often mis-classified as non-stationary by stationarity tests [215]. All ARMA processes simulated in this paper were lag 1, and we set $c = 0$ and $\phi = 0.99$. For the analyses in Supplementary Tables 3.1-3.4 and in Table 3.2, we drew $\boldsymbol{\theta}$ from a random, normal distribution with mean $\mu = 0$ and standard deviation $\sigma = 1$ for each simulation. We also tested ARMA(1) processes for $\boldsymbol{\theta}$ values fixed at -0.5, 0, 0.5, and 0.9 (Supplementary Table 3.19).

Random walk. A random walk is modeled by an autoregressive process with a unit root:

$$X_t = X_{t-1} + \epsilon_t \tag{3.20}$$

where ϵ is a white noise error term with mean $\mu = 0$ and standard deviation $\sigma = 1$. Random walks are non-stationary.

Trended random walk. A trended random walk introduces a secondary non-stationarity, namely, a linear trend, to the random walk:

$$X_t = X_{t-1} + b + \epsilon_t \tag{3.21}$$

where ϵ is a white noise error term with mean $\mu = 0$ and standard deviation $\sigma = 1$ and b is the slope of the linear trend. For all trended random walks simulated in this paper, b was randomly drawn from a Gaussian distribution with mean $\mu = 0$ and standard deviation $\sigma = 0.01$.

Colored noise. Colored noise refers to a stationary, stochastic process with a non-uniform power spectrum. White noise has a uniform power spectrum (meaning equal power at all frequencies); pink noise has a power spectral density proportional to $\frac{1}{f}$, where f is frequency; red noise has a power spectral density proportional to $\frac{1}{f^2}$; blue noise has a power spectral density proportional to f; and violet noise has a power spectral density proportional to f^2. All colored noise signals were simulated using Zhivomirov's algorithm [190], available at https://www.mathworks.com/matlabcentral/fileexchange/42919-pink-red-blue-and-violet-noise-generation-with-matlab.

Rössler system. The Rössler system is described by the following system of differential equations [191]:

$$\begin{aligned} \frac{dx}{dt} &= -wy - z \\ \frac{dy}{dt} &= wx + ay \\ \frac{dz}{dt} &= b + z(x - c) \end{aligned} \tag{3.22}$$

We set $a = 0.2$, $b = 0.2$, and $c = 5.7$, for which the Rössler system exhibits chaos [191]. w controls the frequency of the system's oscillations, and was set to 1. We integrated the equations for the Rössler system using the Fourth Order Runge-Kutta method with an integration step of 0.01. We generated 5,000,000 time-points, and then downsampled to 10,000 datapoints. To get a single time-series $\boldsymbol{\phi}$ from the Rössler model, we took a linear combination of $\mathbf{x}$ and $\mathbf{y}$: $\boldsymbol{\phi}=\mathbf{x}+\mathbf{y}$. Results for $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ individually are reported in Supplementary Table 3.15.

Ikeda map. Ikeda and colleagues described a chaotic model of light passing through a nonlinear optical resonator [216]. The model can be simplified into a two-dimensional map [217]:

$$\begin{aligned} x_{i+1} &= 1 + u(x_i \cos t_i - y_i \sin t_i) \\ y_{i+1} &= u(x_i \sin t_i - y_i \cos t_i) \end{aligned} \tag{3.23}$$

where u is a parameter and

$$t_n = 0.4 - \frac{6}{1 + x_i^2 + y_i^2} \tag{3.24}$$

We set $u = 0.9$, for which the Ikeda map exhibits chaos [217]. Table 3.3 reports results for a linear combination of the two variables, $\mathbf{\Phi}$=$\mathbf{x}$+$\mathbf{y}$. Results for $\mathbf{x}$ and $\mathbf{y}$ individually are reported in Supplementary Table 3.15.

Hénon map. The Hénon map [193] is a two-dimensional system of equations:

$$\begin{aligned} x_{i+1} &= 1 - a x_i^2 + y_i \\ y_{i+1} &= b x_i \end{aligned} \tag{3.25}$$

We set $a = 1.25$ and $b = 0.3$, for which the Hénon map is periodic [218]. Table 3.3 reports results for a linear combination of the two variables, $\mathbf{\Phi}$=$\mathbf{x}$+$\mathbf{y}$. Results for $\mathbf{x}$ and $\mathbf{y}$ individually are reported in Supplementary Table 3.15.

Periodic Poincaré oscillator. The Poincaré oscillator has been widely studied as a model of biological oscillations, particularly as a model of the effect of periodic stimulation on the dynamics of biological oscillators [219]. The oscillator is described by the following equations:

$$\begin{aligned} \frac{dr}{dt} &= kr(1 - r) \\ \frac{d\Phi}{dt} &= 2\Phi \end{aligned} \tag{3.26}$$

where k is a positive value that controls the oscillator's relaxation rate. The phase of this system is described by its angular coordinate ϕ in a unit cycle. Periodic stimulation of the system is modeled as a perturbation of magnitude b away from the unit cycle, which leads to an instantaneous resetting of the phase of the oscillator, as determined by the following phase resetting curve:

$$g(\phi) = \frac{1}{2\pi} \arccos \frac{\cos 2\pi\phi + b}{\sqrt{1 + b^2 + 2b \cos 2\pi\phi}} (\text{mod } 1) \tag{3.27}$$

The period of the stimulation is determined by a parameter τ. For periodic dynamics, we analyze the time-varying phase of the Poincaré oscillator with $b = 1.13$ and $\tau = 0.69$, which Guevara and Glass show leads to phase locking between the oscillator and the periodic perturbations [173].

Quasi-periodic Poincaré oscillator. For quasi-periodic dynamics [173] in the Poincaré oscillator, wet set $b = 0.95$ in Eq. 20, with an inter-stimulus delay $\tau = 0.75$.

Chaotic Poincaré oscillator. For chaotic dynamics [173] in the Poincaré oscillator, wet set $b = 1.13$ and $\tau = 0.65$.

Stochastic Lorenz system. To study the effect of dynamic noise on our algorithm's classification of stochastic chaotic systems, we took the Lorenz system described in Eq. 11, and added intrinsic/dynamic Gaussian noise to the **x** component of the system (we found that the system was far less sensitive to noise being injected into the **y** variable):

$$\begin{aligned} \frac{dx}{dt} &= \sigma(y - x) + A\eta \\ \frac{dy}{dt} &= x(\rho - z) - y \\ \frac{dz}{dt} &= xy - \beta z \end{aligned} \tag{3.28}$$

where η_i is a normally distributed random variable with mean 0 and standard deviation 1, and A is a parameter that controls the amplitude of the dynamic/intrinsic noise. As for the deterministic case, we set $\sigma = 10$, $\rho = 30$, and $\beta = \frac{8}{3}$. The stochastic Lorenz system was simulated using the Fourth Order Runge-Kutta method with an integration step of 0.01. Supplementary Table 3.18 reports results for different values of A, both for all system variables individually and for the linear combination **x**+**y**.

Stochastic Rössler system. We also took the Rössler system described in Eq. 22, and added dynamical Gaussian noise to the **x** component of the system:

$$\begin{aligned} \frac{dx}{dt} &= \sigma(y - x) + A\eta \\ \frac{dy}{dt} &= x(\rho - z) - y \\ \frac{dz}{dt} &= xy - \beta z \end{aligned} \tag{3.29}$$

where η_i is a normally distributed random variable with mean 0 and standard deviation 1, and A is a parameter that controls the amplitude of the dynamic/intrinsic noise. As for the deterministic case, we set $a = 0.2$, $b = 0.2$, and $c = 5.7$, and simulated the model using

the Fourth Order Runge-Kutta method with an integration step of 0.01. We generated 5,000,000 time-points, and then downsampled to 10,000 datapoints. Supplementary Table 3.18 reports results for different values of A, both for all system variables individually and for the linear combination of $\mathbf{x}$+$\mathbf{y}$.

Multivariate AR model. We generated random multivariate autoregressive (AR) models using the Multivariate Granger Causality (MVGC) toolbox [220]. To create random regression matrices, we created random 5-node dense positive definite matrices using Matlab's `sprandsym.m` function, with a graph density of 1. To ensure stationary dynamics, we used the MVGC toolbox's `var_specrad.m` function to decay the coefficients of the random dense positive definite matrices so that their spectral radii were 0.8. To ensure uncorrelated noise in the resulting AR model, we created error matrices with diagonal elements set to 1 and off-diagonal elements set to 0. We then inputted these regression and error matrices into the MVGC toolbox's `var_to_tsdata.m` function to create multivariate time-series with 5 nodes and 10,000 time-points. We then applied the Chaos Decision Tree Algorithm to just the univariate activity of the first node of the resulting multivariate signal.

3.5.4 Empirical Data

We here describe the real-world data analyzed in this paper, and how these data were previously classified as stochastic, periodic, or chaotic:

A chaotic neuron integrated circuit. Data recorded from an integrated circuit were kindly sent to us by Seiji Uenohara and colleagues. The circuit is a physical implementation of a chaotic neuron model that is based on the Hudgkin-Huxley equations [221]. The equations governing the circuit's behavior can be reduced to the following two-dimensional map:

$$\begin{aligned} \zeta(t+1) &= k_r\zeta(t) + af(\zeta(t) + b\cos(2\pi\theta(t))) + a \\ \theta(t+1) &= \theta(t) + \omega \ (\mathrm{mod} 1) \end{aligned} \tag{3.30}$$

where $f(\cdot)$ is a monotonically decreasing nonlinear output function, b controls the amplitude of the quasiperiodic forcing, and $\omega = \frac{\sqrt{5}-1}{2}$ (the golden ratio). The quasiperiodic external forcing was inputted to the circuit using an analog board PXI-6289, which was also used to record the circuit's output. Varying the parameter b can bring the circuit into periodic, strange non-chaotic, and chaotic states, which Uenohara and colleagues were

able to classify by analyzing the consistency of the circuit's response to an external input [322]. There are 10 datasets recorded from the circuit's chaotic state.

A strange non-chaotic neuron integrated circuit. There are 10 datasets recorded from the strange non-chaotic state of Uenohara and colleagues' circuit.

A periodic neuron integrated circuit. There are 10 datasets recorded from the periodic state of Uenohara and colleagues' circuit.

Chaotic laser. Hübner and colleagues [160] used phase portrait, correlation integral, and autocorrelation function analyses to detect chaos in the intensity pulsing of an unidirectional far-infrared NH3 ring laser. Laser data were downloaded from https://www.pdx.edu/biomedical-signal-processing-lab/chaotic-time-series.

Stellar flux of a strange non-chaotic star. We analyzed stellar flux from KIC 5520878, the only known non-artificial strange non-chaotic system [161]. Data were sent to us by John F. Lindner, who, together with colleagues, determined the status of KIC 5520878 as a strange non-chaotic system using a series of spectral scaling analyses [161]. Data were originally obtained from the Mikulski Archive for Space Telescopes. Because there are large shifts in the data due to the stellar flux being recorded in different pixels of the Kepler Space Telescope, we visually inspected the data to find a relatively stable period (i.e. a period in between large shifts) and then detrended the data. We thus exclusively analyzed time points 11620 to 14003 from the dataset analzed in Lindner and colleagues' paper [161].

North Atlantic Oscillation Index. We analyzed the monthly mean North Atlantic Oscillation (NAO) Index from January 1950 to December 2018. The NAO Index is the difference in atmospheric pressure at sea level between the Azores high and the Icelandic low, and has been shown by several groups of researchers, employing a range of techniques, to be stochastic [157, 222–227]. Data were downloaded from the Climate Prediction Center website (http://www.cpc.ncep.noaa.gov/).

Parkinson's tremor. We analyzed recordings of a Parkinson's patient's hand acceleration, measured for 30 seconds at a sampling rate of 1000 Hz by piezoresistive accelerometers. Through analyses of correlation integrals, Poincaré and return maps, Lyapunov exponents, and the δ-ϵ method, Timmer and colleageus [189] showed that this Parkinson's tremor was a nonlinear stochastic oscillator. Data were downloaded from http://jeti.uni-freiburg.de/path_tremor/readme.

Essential tremor. We analyzed recordings of hand acceleration from a patient with an essential tremor, also measured for 30 seconds at a sampling rate of 1000 Hz by piezoresistive accelerometers. As they did with the Parkinson's tremor, Timmer and colleagues used correlation integrals, Poincaré and return maps, Lyapunov exponents, and the δ-ϵ method to show that this essential tremor was a nonlinear stochastic oscillator. Data were downloaded from http://jeti.uni-freiburg.de/path_tremor/readme.

Heart rate (healthy subjects). Five heart beat (RR-interval) time-series recordings from healthy subjects were downloaded from Physionet [163]: https://www.physionet.org/challenge/chaos/. The signals were recorded using continuous ambulatory (Holter) electrocardiograms, and are in sinus rhythm. Outliers were filtered out of the data using Physionet's WFDB software package. Though a full 24 hours of data were available for each subject, we only took the first 2.78 hours of data, corresponding to 10,000 time-points. This was both to save on computation time and to be consistent with the length of other time-series analyzed in this paper.

Heart rate (congestive heart failure patients). Five heart beat (RR-interval) time-series recordings from patients with congestive heart failure were downloaded from Physionet [163]. Like the healthy rate signals, these data were recorded using continuous ambulatory (Holter) electrocardiograms, are in sinus rhythm, and were filtered for outliers. Though a full 24 hours of data were available for each subject, we only took the first 2.78 hours of data.

Heart rate (atrial fibrillation). Five heart beat (RR-interval) time-series recordings from patients with congestive heart failure were downloaded from Physionet [163]. Like the healthy rate signals, these data were recorded using continuous ambulatory (Holter) electrocardiograms and were filtered for outliers, but are not in sinus rhythm. We only took the first 2.78 hours of data, corresponding to 10,000 time-points.

3.5.5 Parameters and largest Lyapunov exponents for data in Table 3.5

We here describe the methods used to generate data with different degrees of chaos for the analyses reported in Table 3.5, as well as the methods used to calcualte largest Lyapunov exponents in these systems.

Logistic map. The logistic map has only a single parameter, r (see above). Following Bandt and Pompe [164], we varied r between 3.5 and 4, in intervals of 0.001, to generate

501 10,000 time-point signals with different levels of chaos. Ground-truth largest Lyapunov exponents were calculated using the derivative method, which does not involve generating time-series data.

Hénon map. To generate different degrees of chaos in the Hénon map, we varied its a parameter (see above) between 1 and 1.4, in intervals of 0.001, to generate 401 10,000 time-point signals with different degrees of chaos. Ground-truth largest Lyapunov exponents were calcualted using code provided in *Dynamical Systems with Applications using Matlab* [228].

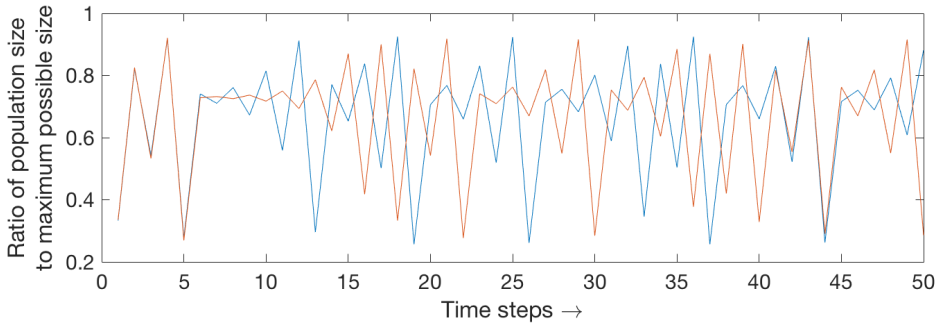
Lorenz system. For the Lorenz system, we varied its σ parameter between 5.75 and 15, in intervals of .05, and generated 10,000 time-points per simulation. Within this parameter range, the Lorenz system is chaotic, but displays varying degrees of chaos. To calculate largest Lyapunov exponents for each parameter, we used the algorithm provided by Ramasubramanian [213], which, like the algorithms used for the logistic and Hénon maps, does not involve generating time-series data.

Mean-field cortical model. Following Steyn-Ross, Steyn-Ross, and Sleigh [276], different levels of chaos in their mean-field cortical model were generated by varying two parameters: postsynaptic inhibitory response and inhibitory diffusion. The postsynaptic inhibitory response parameter (λ_i in their model) was varied between 0.98 and 1.018 in intervals of 0.001, and the inhibitory diffusion parameter (D_2 in their model) was varied between 0.1 and 0.8 in intervals of 0.05, producing a total of 585 parameter configurations. We simulatd 25,000 time-points in the model, with no downsampling, so we could again test the effect of the Chaos Decision Tree Algorithm's automated dowsampling on permutation entropy' ability to track level of chaos in continuous systems. Unfortunately, there are no tools for analytically estimating the largest Lyapunov exponent of the mean-field cortical model, and so largest Lyapunov exponents were approximated by running two noise-free simulations of the model for each parameter configuration, with very slightly different initial conditions, and fitting a line to the rate of divergence between the two simulations from the beginning of the simulations to the point when their divergence rate saturates and flattens out. The slope of the fitted line is taken as the estimate of the largest Lyapunov exponent [362]. While Steyn-Ross, Steyn-Ross, and Sleigh's code for estimating largest Lyapunov exponents using this method requires subjective evaluation of where to fit the line (i.e. finding the non-saturated part of the divergence rate plot), we automated this process by fitting a line from the beginning of the divergence rate plot to the point where its mean abruptly changes, reflecting saturation; this point was determined using Matlab's findchangepts.m function. Simulations with approximated largest

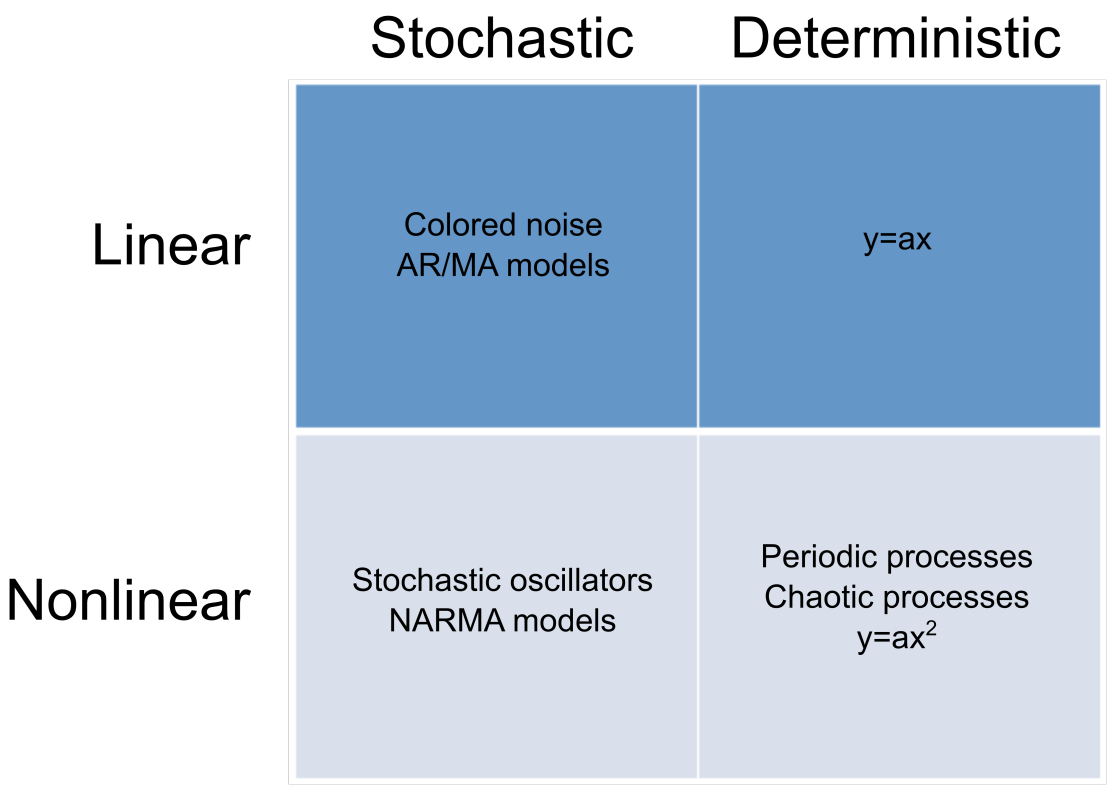
Lyapunov exponents less than -5 or greater than 5 were excluded from the analysis, as these are likely poor estimates; visual inspection confirmed that for these cases, there was often no clear point of saturation for the divergence rate between simulations, and so lines were often automatically fitted to particularly steep, short sub-segments of the plot. Visual inspection further confirmed that in most other cases, there was a clear, linear rate of divergence between simulations followed by saturation, and that the automatically fitted line was a good fit.

Neuron integrated circuit. The integrated circuit data analyzed in Table 3.5 are the same as those analyzed in Table 3.4. Because the circuit is a physical implementation of a known and simple two-dimensional system of equations, Uenuhara and colleagues used those equations to calculate the ground-truth largest Lyapunov exponents of the circuit in its three different states (periodic, strange non-chaotic, and chaotic), and report those largest Lyapunov exponents in their paper [322].

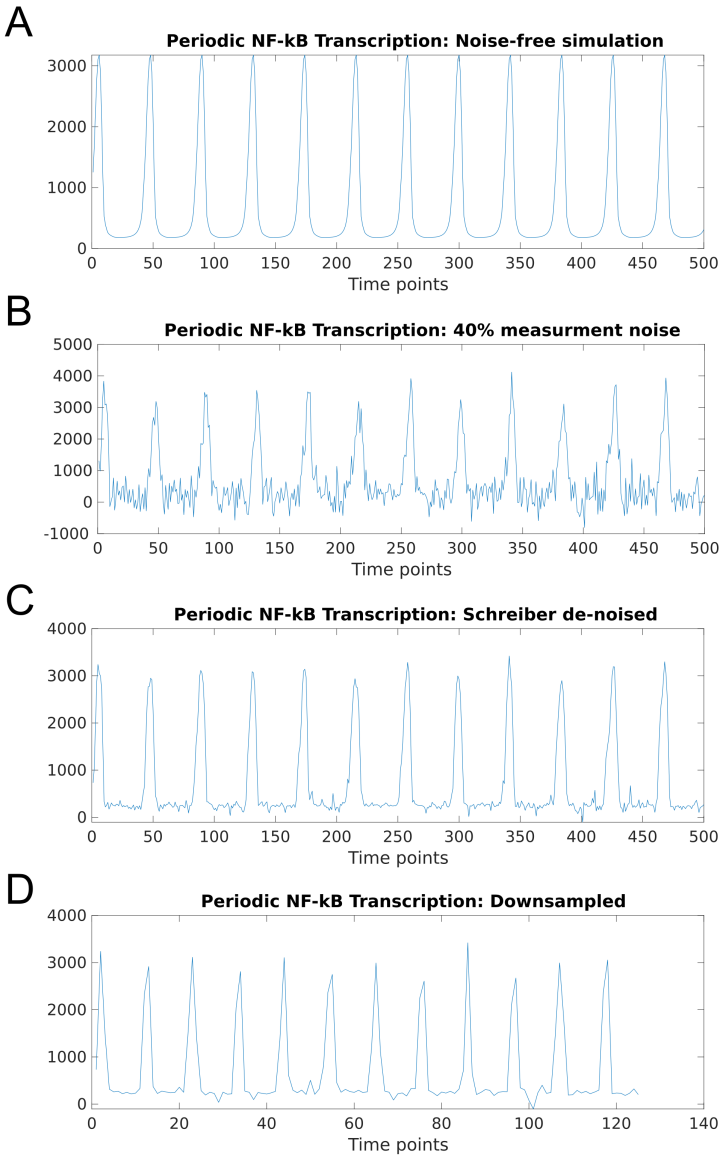
3.6 Supplementary Figures



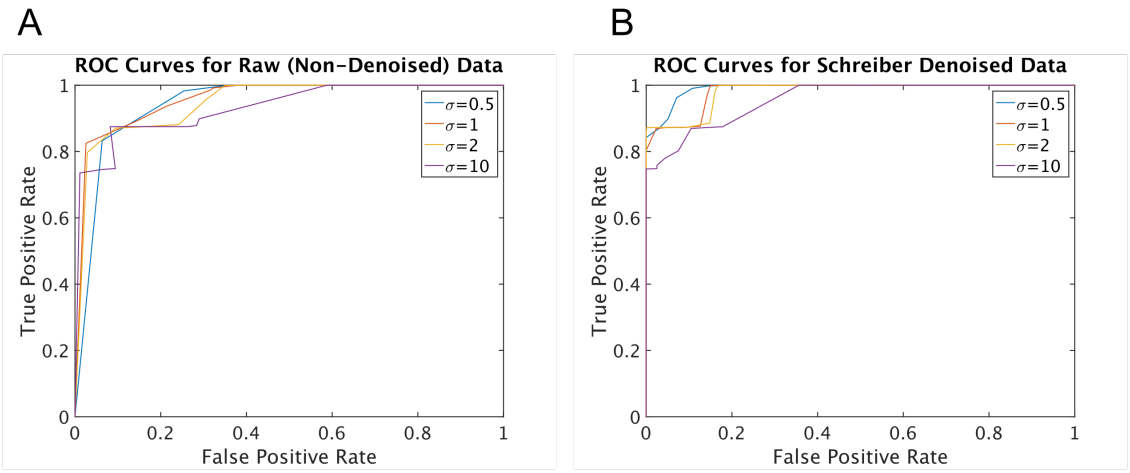
Supplementary Figure 3.2: A graphical depiction of chaos. Here, we show two simulations of the chaotic logistic map, which is a simple model of population dynamics (see Methods). The starting values of the two simulations are only very slightly different (they differ by a value of only 0.001), and yet by the fifth time-step, the two simulations show a clear divergence in their trajectories. This extreme sensitivity to small perturbations (within bounded, predominantly deterministic systems) is what it means to be chaotic.



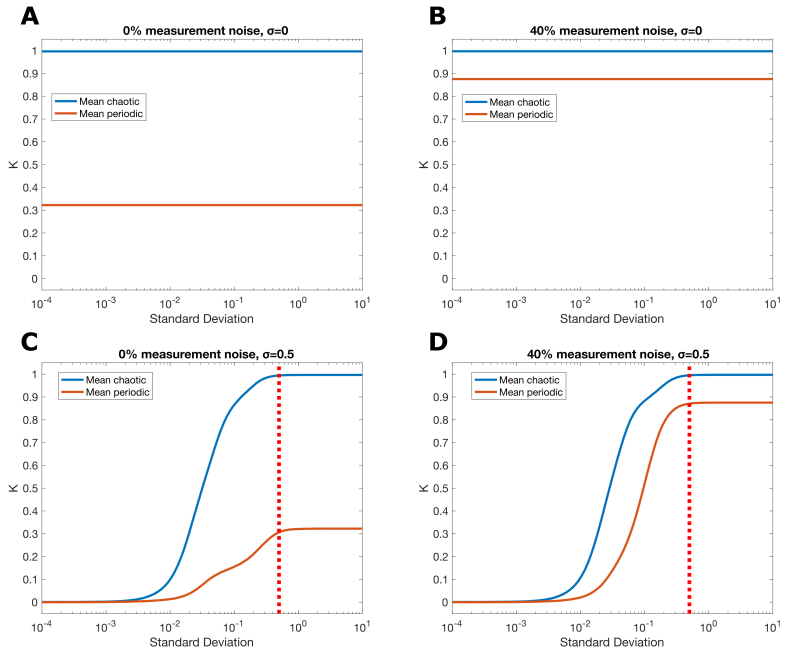
Supplementary Figure 3.3: Dynamical systems can be broadly categorized as linear or nonlinear, and as deterministic or stochastic. Both chaotic and periodic processes are nonlinear and deterministic, but chaotic processes have a positive largest Lyapunov exponent (meaning that initially similar system states diverge exponentially fast) and periodic processes have a negative largest Lyapunov exponent.



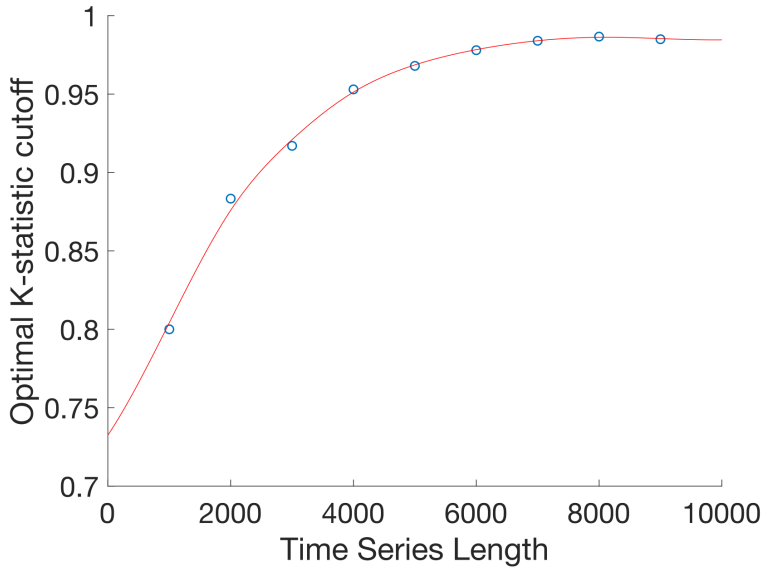
Supplementary Figure 3.4: Sample time-series to illustrate the steps of the Chaos Decision Tree Algorithm. **A** The first 500 time-points of a noise-free simulation of periodic NF-κB transcription (out of 10,000). **B** The same time-series as in **A**, but with added white noise, the amplitude of which is 40% the standard deviation of the original time-series. **C** The same time-series as in **B**, but after Schreiber de-noising. **D** The same time-series as in **C**, but after the de-noised signal has been iteratively downsampled until the difference between the global maximum and global minimum of the signal, divided by the mean absolute difference between consecutive time-points in the signal, is less than or equal to 10. The 0-1 test is applied after this final step.



Supplementary Figure 3.5: Receiver Operating Characteristic Curves for different parameters of the 0-1 chaos test, for both raw data (A) and Schreiber de-noised data (B). Here, we included all deterministic datasets from Tables 3.1 and 3.2. Systems were classified as periodic or chaotic based on different cutoffs of the K-statistic; cutoffs were K=0.1, K=0.5, K=0.55, K=0.6, K=0.65, K=0.7, K=0.75, K=0.8, K=0.85, K=0.9, K=0.95, K=0.96, K=0.97, K=0.98, and K=0.99. We also varied the parameter σ in the modified 0-1 test, which controls the amplitude of a noise term that is used to suppress correlations arising from quasi-periodicity (and thus improves classification of strange non-chaotic systems as periodic) - see Eq. 3 in the Methods. Across all datasets and noise levels, the 0-1 test with $\sigma = 0.5$ applied to Schreiber denoised data provided the highest classification accuracy, as can be seen by the large area under the curve in B.



Supplementary Figure 3.6: We found that inclusion of a noise term σ in the modified 0-1 test for chaos (see Eq. 3 in Methods) can lead to inaccurate results for signals with very low amplitude/standard deviation. To illustrate this effect, we ran the modified 0-1 test on a single simulation, with either 0% or 40% measurement noise, of each deterministic process in Tables 3.1-3.2. To vary standard deviation, we multiplied each simulation by a constant, such that its standard deviation was fixed at a given value, out of 100 values logarithmically spaced between 10^{-4} and 10^{1} . We then calculated the K-statistic (i.e. the output of the modified 0-1 test) for both σ=0 (**A**-**B**) or σ=0.5 (**C**-**D**). Here, we plot the mean K-statistic across all chaotic systems (blue), as well as the mean K-statistic across all periodic systems (red), as a function of signal standard deviation. For σ =0, K was invariant to the standard deviation of the signal, for either 0% measurement noise (**A**) or 40% measurement noise (**B**). For σ =0.5, on the other hand, K fell to zero for both periodic or chaotic signals as their standard deviation approached zero, for both 0% measurement noise (**C**) and 40% measurement noise (**D**). This is likely because the inclusion of a non-zero noise term in Eq. 3 of the 0-1 test (see Methods) overwhelms the mean squared displacement of signals with very small standard deviations. Note that for σ =0.5, K asymptotes at around a standard deviation of 0.5 for both levels of measurement noise. As such, we modified the 0-1 test to normalize the standard deviation of a test signal to 0.5 (see Methods).



Supplementary Figure 3.7: Though all analyses in the main body of the paper were performed on time-series with 10,000 time-points, we also considered how to optimize the algorithm's performance for shorter time-series. In particular, it is known that the K-statistic of the 0-1 test approaches 1 for chaotic systems and approaches 0 for periodic systems *as* the length of a time-series is increased, but that it can yield intermediate - and therefore ambiguous - results for shorter time-series. As such, we ran the chaos-testing portion of our algorithm's pipeline (which consists of signal de-noising, then normalizing the standard deviation of the signal, then applying the 0-1 test, and then classifying the system as either periodic or chaotic based on some cutoff of the outputted K-statistic) on sub-samples of all non-oversampled deterministic datasets in Tables 3.1 and 3.2, including all levels of measurement noise. We did not include signals classified as oversampled in this analysis, as the downsampling step of our algorithm would shorten the length of the time-series, which would warp the relationship between the optimal K-statistic cutoff and time-series length. We ran the pipeline on samples ranging from 1,000 to 9,000 time-points in length, in intervals of 1,000. For each time-series length, we calculated the F1 score of different K-statistic cutoffs, ranging from K=0.005 to K=0.995, in steps of 0.005. Here, we plot the smoothed vector of optimal K-statistic cutoffs for each time-series length. As expected, given the fact that the K-statistic *approaches* 1 for chaotic signals as time-series length is increased, the optimal cutoff also increased as a function of time-series length, asymptoting near K=0.985 for longer time-series. Thus, if no cutoff is provided to the algorithm, it will automatically pick a cut-off based on the smoothing spline fit plotted here in red. If the predicted optimal cutoff is greater than 0.99, the algorithm picks a cutoff of 0.99 (since the K-statistic is upper-bounded by 1). With this automated cutoff selection, we have confirmed that the Chaos Decision Tree Algorithm performs at very high accuracy for time-series with 1,000 points (Supplementary Table 3.16), 5,000 points (Supplementary Table 3.17), and 10,000 points (Tables 3.1-3.3).

3.7 Supplementary Tables

SUPPLEMENTARY TABLE 3.7: Our algorithm uses surrogate data methods to test for stochasticity (Fig. 3.1, Supplementary Tables 3.2-3.3), but other studies emphasize the importance of first testing for stationarity before using such methods [166, 189] (due to the inherent stationarity of Fourier-based surrogates). A stationary signal is one whose unconditional joint probability distribution is time-invariant. We here assess the performance of a number of stationarity tests (see Patterson [214, 215] for a thorough review of such tests and their relative strengths/weaknesses). Datasets analyzed include all stationary processes in Tables 3.1-3.2 (including bounded random walks, autoregressive processes with moving averages, and nonlinear stochastic processes, all of which are difficult edge cases for such tests [214, 215]), all unit root processes in Table 3.2 (including random walks and trended random walks), and a cyclostationary autoregressive process. See Methods for details on all datasets. The stationarity tests we assessed are the augmented Dickey-Fuller (ADF) test [97], the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test [98], and the Leybourne-McCabe (LMC) test [99], and the non-parametric Lo and MacKinlay Variance Ratio [100, 101] (LM-VR) and Breitung's Variance Ratio [102] (BVR) tests, all with and without first detrending the test signal. Shown: fraction of datasets classified as non-stationary, and F1 scores for stationary vs. unit root processes. Consistent with the analyses performed by Patterson [214, 215], we found that Breitung's Variance Ratio (BVR) test significantly outperformed other tests, though the poor performance for the cyclostationary process underscores the fact that a unit root is only one form of non-stationarity.

	Stationary	Unit Root	Cyclostationary	F1
ADF test	506/10500	242/1000	0/500	0.28
KPSS test	2850/10500	1000/1000	0/500	0.41
LMC Test	2172/10500	903/1000	0/500	0.44
LM-VR test	215/10500	187/1000	0/500	0.27
BVR test	161/10500	961/1000	0/500	0.91
ADF test (detrended)	6/10500	76/1000	0/500	0.14
KPSS test (detrended)	2850/10500	1000/1000	0/500	0.41
LMC Test (detrended)	2172/10500	903/1000	0/500	0.44
LM-VR test (detrended)	215/10500	187/1000	0/500	0.27
BVR test (detrended)	161/10500	935/1000	0/500	0.89

SUPPLEMENTARY TABLE 3.8: F1 scores for surrogate-based tests of stochasticity. For each test dataset, 500 surrogates were generated using Lancaster and colleagues' Matlab toolbox [166]. For all surrogate algorithms, the data are first pre-processed such that the start and end points of the data and their first derivatives are matched as closely as possible. We used permutation entropy as our test statistic, so that a signal was classified as stochastic if its permutation entropy fell within the distribution of permutation entropies calculated from the 500 surrogate time-series. We tested the effect of Schreiber denoising ("Denoised") vs. no denoising ("Raw") on test accuracy. We further tested the effect of normality transformation using the Box-Cox method ("Transformed (Box-Cox)"), following the recommendation of Chan and Tong [154], or using a rank-based inverse normal transformation ("Transformed (INT)"), vs. no normality transformation ("Non-transformed"). Moreover, we tested the benefit of excluding signals classified as non-stationary by Breitung's Variance Ratio test ("Stationary Data") vs. including all signals ("All data"). We assessed five different surrogate algorithms, namely Amplitude Adjusted Fourier Transform [165] (AAFT) surrogates, Fourier Transform [165] (FT) surrogates, iterative Amplitude Adjusted Fourier Transform [165] (iAAFT) surrogates, Cycle Shuffled Surrogates [105] (CSS), and Cyclic Phase Perutation [167] (CPP) surrogates, as well as a combination of Amplitude Adjusted Fourier Transform and Cyclic Phase Permutation (AAFT+CPP) surrogates (see Methods). Finally, we tested eight different permutation order values for the calculation of permutation entropy, which was the test statistic used to discriminate between the original and surrogate signals. Stochastic datasets consisted of all linear and nonlinear stochastic processes in Tables 3.1-3.2, and deterministic datasets consisted of all deterministic processes in Tables 3.1-3.2, which included all datasets with added measurement noise. Performance was highest for raw, non-transformed signals, with a permutation order of 8 and a combination of AAFT and CPP surrogates (bolded F1 score). In other words, if a signal's permutation entropy (with a permutation order of 8) fell within the distributions of permutation entropies of *either* AAFT surrogate signals *or* CPP surrogate signals, generated from the raw original time-series, then that signal was very likely stochastic (either linear or nonlinear stochastic). Counter to prior expectations [158, 166], no discernible benefit was gained by excluding signals classified as non-stationary by Breitung's Variance Ratio test. Table on next page.

Supplementary Table 3.2 (continued).

				n=3	n=4	n=5	n=6	n=7	n=8	n=9	n=10
Raw	Non-transformed	All data	AAFT	0.68	0.78	0.77	0.76	0.76	0.75	0.75	0.77
			FT	0.67	0.79	0.82	0.82	0.83	0.81	0.77	0.77
			CSS	0.01	0.01	0.02	0.04	0.08	0.07	0.19	0.29
			CPP	0.29	0.33	0.29	0.3	0.36	0.43	0.49	0.5
			IAAFT	0.67	0.73	0.77	0.77	0.77	0.74	0.71	0.7
			AAFT+CPP	0.68	0.86	0.86	0.86	0.87	**0.91**	0.89	0.89
		Stationary data	AAFT	0.59	0.72	0.71	0.7	0.7	0.69	0.69	0.72
			FT	0.58	0.72	0.76	0.76	0.77	0.75	0.71	0.7
			CSS	0.01	0.01	0.02	0.05	0.11	0.09	0.24	0.36
			CPP	0.28	0.32	0.28	0.28	0.36	0.46	0.53	0.55
			IAAFT	0.58	0.65	0.7	0.71	0.7	0.66	0.63	0.62
			AAFT+CPP	0.6	0.81	0.81	0.81	0.83	0.89	0.86	0.86
	Transformed (Box-Cox)	All data	AAFT	0.68	0.78	0.77	0.76	0.76	0.75	0.75	0.77
			FT	0.7	0.79	0.81	0.8	0.8	0.82	0.81	0.83
			CSS	0.03	0.03	0.03	0.04	0.07	0.04	0.1	0.12
			CPP	0.2	0.24	0.23	0.24	0.28	0.3	0.38	0.4
			IAAFT	0.65	0.7	0.72	0.73	0.74	0.75	0.76	0.75
			AAFT+CPP	0.59	0.74	0.76	0.76	0.77	0.78	0.77	0.78
		Stationary data	AAFT	0.59	0.72	0.71	0.7	0.7	0.69	0.69	0.72
			FT	0.61	0.72	0.74	0.74	0.74	0.77	0.76	0.77
			CSS	0.03	0.03	0.03	0.05	0.09	0.04	0.13	0.15
			CPP	0.16	0.21	0.2	0.21	0.25	0.27	0.37	0.38
			IAAFT	0.56	0.61	0.64	0.65	0.66	0.68	0.69	0.67
			AAFT+CPP	0.49	0.65	0.68	0.68	0.7	0.71	0.7	0.71
	Transformed (INT)	All data	AAFT	0.68	0.78	0.77	0.76	0.76	0.75	0.75	0.77
			FT	0.68	0.78	0.77	0.77	0.77	0.76	0.75	0.77
			CSS	0.06	0.1	0.15	0.27	0.13	0.02	0.1	0.09
			CPP	0.59	0.47	0.45	0.38	0.46	0.51	0.65	0.69
			IAAFT	0.67	0.8	0.8	0.79	0.8	0.79	0.79	0.8
			AAFT+CPP	0.67	0.73	0.76	0.75	0.77	0.78	0.78	0.78
		Stationary data	AAFT	0.59	0.72	0.71	0.7	0.7	0.69	0.69	0.72
			FT	0.59	0.72	0.71	0.7	0.7	0.69	0.69	0.72
			CSS	0.07	0.12	0.19	0.33	0.17	0.03	0.12	0.12
			CPP	0.51	0.39	0.39	0.27	0.33	0.37	0.54	0.57
			IAAFT	0.56	0.72	0.71	0.71	0.71	0.71	0.7	0.72
			AAFT+CPP	0.57	0.64	0.67	0.67	0.68	0.7	0.71	0.71
Denoised	Non-transformed	All data	AAFT	0.38	0.22	0.19	0.19	0.2	0.21	0.22	0.37
			FT	0.48	0.41	0.4	0.4	0.39	0.4	0.44	0.58
			CSS	0.05	0.15	0.12	0.12	0.14	0.17	0.11	0.3
			CPP	0.31	0.28	0.27	0.3	0.33	0.42	0.6	0.67
			IAAFT	0.47	0.47	0.47	0.47	0.48	0.5	0.54	0.59
			AAFT+CPP	0.51	0.44	0.42	0.44	0.47	0.55	0.7	0.78
		Stationary data	AAFT	0.43	0.25	0.23	0.24	0.24	0.26	0.26	0.43
			FT	0.48	0.42	0.42	0.42	0.42	0.43	0.46	0.59
			CSS	0.03	0.16	0.11	0.1	0.1	0.1	0.1	0.36
			CPP	0.32	0.31	0.32	0.35	0.38	0.48	0.61	0.72
			IAAFT	0.47	0.47	0.47	0.47	0.48	0.49	0.52	0.55
			AAFT+CPP	0.53	0.5	0.49	0.51	0.54	0.63	0.74	0.83
	Transformed (Box-Cox)	All data	AAFT	0.39	0.34	0.31	0.31	0.31	0.32	0.34	0.37
			FT	0.46	0.45	0.45	0.44	0.43	0.44	0.47	0.54
			CSS	0.02	0.13	0.1	0.13	0.17	0.1	0.16	0.18
			CPP	0.26	0.3	0.32	0.39	0.47	0.48	0.55	0.55
			IAAFT	0.5	0.49	0.51	0.52	0.52	0.54	0.56	0.6
			AAFT+CPP	0.45	0.5	0.49	0.55	0.61	0.63	0.67	0.66
		Stationary data	AAFT	0.44	0.41	0.38	0.38	0.39	0.4	0.41	0.43
			FT	0.49	0.5	0.51	0.49	0.49	0.5	0.53	0.57
			CSS	0.02	0.14	0.11	0.14	0.19	0.08	0.17	0.21
			CPP	0.19	0.24	0.27	0.35	0.44	0.42	0.48	0.5
			IAAFT	0.5	0.49	0.52	0.54	0.53	0.54	0.55	0.56
			AAFT+CPP	0.4	0.48	0.48	0.56	0.62	0.61	0.64	0.64
	Transformed (INT)	All data	AAFT	0.39	0.34	0.31	0.31	0.31	0.32	0.34	0.38
			FT	0.39	0.36	0.33	0.32	0.33	0.33	0.35	0.38
			CSS	0.04	0.08	0.13	0.1	0.03	0.04	0.05	0.09
			CPP	0.67	0.56	0.56	0.57	0.61	0.65	0.67	0.68
			IAAFT	0.43	0.43	0.42	0.41	0.42	0.44	0.47	0.51
			AAFT+CPP	0.71	0.65	0.66	0.7	0.75	0.78	0.78	0.78
		Stationary data	AAFT	0.43	0.4	0.38	0.38	0.38	0.4	0.41	0.43
			FT	0.44	0.42	0.4	0.39	0.4	0.41	0.42	0.43
			CSS	0.05	0.11	0.16	0.12	0.04	0.05	0.06	0.12
			CPP	0.55	0.47	0.46	0.45	0.47	0.5	0.54	0.55
			IAAFT	0.44	0.47	0.45	0.45	0.46	0.47	0.49	0.5
			AAFT+CPP	0.63	0.58	0.6	0.64	0.67	0.7	0.71	0.71

SUPPLEMENTARY TABLE 3.9: In Supplementary Table 3.2, we showed that the highest-performing surrogate-based method for detecting signal stochasticity was to use raw (non-denoised) signals, without normality transformation, and without excluding signals classified as non-stationary by Breitung's Variance Ratio test. For all analyses in Supplementary Table 3.2, we generated 500 surrogate time-series. Here, we tested whether generating 1,000 surrogate time-series from all raw time-series would yield higher performance in discriminating stochastic from deterministic processes. Though minor, we did observe higher F1 scores with a larger number of surrogates, and so the Chaos Decision Tree Algorithm uses 1,000 surrogates in its stochasticity test. We note that it is possible, and perhaps likely, that including even more surrogate signals in the stochasticity test will yield even more accurate results, and so the algorithm also allows the user to specify how many surrogates to generate.

	n=3	n=4	n=5	n=6	n=7	n=8	n=9	n=10
AAFT	0.67	0.79	0.78	0.77	0.76	0.76	0.75	0.78
FT	0.66	0.78	0.82	0.82	0.82	0.81	0.77	0.77
CSS	0.01	0.02	0.02	0.04	0.09	0.08	0.21	0.3
CPP	0.3	0.33	0.3	0.31	0.37	0.44	0.5	0.51
IAAFT	0.67	0.72	0.77	0.77	0.77	0.73	0.7	0.7
AAFT+CPP	0.67	0.86	0.87	0.87	0.88	**0.92**	0.89	0.89

Supplementary Table 3.10: To further test the relationship between stationarity and the accuracy of our surrogate-based stochasticity test, we broke down the performance of both Breitung's Variance Ratio test and our surrogate-based stochasticity test (which uses permutation entropy and a combination of Amplitude Adjusted Fourier Transform and Cyclic Phase Permutation surrogates) on all non-stationary processes analyzed in this paper, as well as stationary processes that are classically difficult to distinguish from non-stationary processes [214, 215]. These included bounded random walks (which, though globally stationary, have local unit roots) and autoregressive processes with a moving average component. We found that our surrogate-based stochasticity test performed with near-perfect accuracy for *linear* stochastic non-stationary processes, i.e. random walks, trended random walks, and a cyclostationary autoregressive process, as well as for stationary linear stochastic processes with moving averages (which can be difficult to distinguish from non-stationary processes [214, 215]). Where performance was slightly worse was for *nonlinear* stochastic processes, even though those processes were stationary (these processes were bounded random walks, the Freitas map, and the sine map). This suggests that non-stationarity may not affect the performance of surrogate-based stochaticity tests when such tests include non-Fourier based surrogates, such as the Cyclic Phase Permutation surrogates used here, though this possibility should be investigated more systematically in future work. In light of this result, and the results reported in Supplementary Table 3.2, the Chaos Decision Tree Algorithm does not automatically include a stationarity test, though the user can specify that the pipeline include a preliminary stationarity test and choose among any of the stationarity tests analyzed here.

	Fraction classified as stationary	Fraction classified as stochastic
Random walks (linear, stochastic, non-stationary)	23/500	498/500
Trended random walks (linear, stochastic, non-stationary)	16/500	477/500
Bounded random walks (nonlinear, stochastic, stationary)	487/500	451/500
ARMA(1) process (linear, stochastic, stationary)	488/500	481/500
Cyclostationary process (linear, stochastic, non-stationary)	500/500	498/500
Sine map (nonlinear, stochastic, stationary)	443/500	163/500
Freitas map (nonlinear, stochastic, stationary)	500/500	431/500

SUPPLEMENTARY TABLE 3.11: We tested the performance of three de-noising algorithms: a moving average filter (using Matlab's smooth.m function), Schreiber de-noising [168], and wavelet de-noising using an empirical Bayesian method with a Cauchy prior (using Matlab's wdenoise.m function). For each system, we created 100 noise-free simulations, and added white noise to those simulations, the amplitude of which was 40% the standard deviation of the original signals. We then applied the de-noising algorithms to the noise-contaminated signals, and calculated the Pearson correlation between the de-noised signals and the original noise-free signals. Shown: mean Pearson correlations across 100 datasets per system. Schreiber de-noising vastly outperformed the other two approaches, and so the Chaos Decision Tree Algorithm automatically uses Schreiber de-noising; the user can also specify that one of the other two de-noising methods be used.

System	Moving Average	Schreiber	Wavelet
Cortical model (chaotic)	0.81	0.96	0.90
Cortical model (periodic)	0.52	0.97	0.75
Spiking neuron (chaotic)	0.45	0.97	0.85
Granulocyte levels (chaotic)	0.76	0.94	0.87
Granulocyte levels (periodic)	-0.70	0.97	0
NF-κB transcription (chaotic)	0.97	0.99	0.99
NF-κB transcription (periodic)	0.97	0.98	0.97
Cubic map (chaotic)	0.090	0.95	0.020
Cubic map (periodic)	0.010	0.97	0
Cubic map (SNA HH)	-0.25	0.96	0
Cubic map (SNA S3)	0.65	0.97	0
GOPY map (SNA)	0.18	0.94	0
Logistic map (chaotic)	0.42	0.97	0.030
Logistic map (periodic)	0.67	0.98	0
Lorenz system (chaotic)	0.42	0.93	0.07
Generalized Hénon map (hyperchaotic)	0.060	0.94	0.010
Freitas map	0.55	0.94	0.51
Noise-driven sine map	0.97	0.98	0.97

SUPPLEMENTARY TABLE 3.12: De-noising and downsampling improves 0-1 test performance on over-sampled continuous data. Shown: mean K statistic (which is the output of the 0-1 test), with $\sigma = 0.5$ (see Methods, Supplementary Figure 3.4), with standard error, across 100 samples of three oversampled simulations (the oversampling statistic η, whose mean across 100 samples is reported on the left, is the difference between the global maximum and global minimum of the signal divided by the mean absolute difference between consecutive time-points in the data [199]. If $\eta > 10$, then the downsampling approach simply downsamples the data until $\eta \leq 10$ or until there are fewer than 100 time-points left in the downsampled signal). The 0-1 test was performed on each dataset without downsampling, with downsampling, and with only local minima and maxima of the signal (following an alternative approach suggested by Eyébé Fouda and colleagues [300] to improve 0-1 test performance, which a user can select to use instead of downsampling). Note that for noise-contaminated oversampled data, only Schreiber de-noising followed by downsampling brings the K statistic within ranges expected for periodic or chaotic systems (highlighted).

Noise-free			
	Raw	Downsampled	Local Minima/Maxima
Chaotic Lorenz system ($\bar{\eta}$=334)	-0.0008 +/- 0	0.998 +/- 0	0.962 +/- 0.002
Periodic transcription ($\bar{\eta}$=21)	0.21 +/- 0.026	0.088 +/- 0.01	-0.026 +/- 0
Chaotic transcription ($\bar{\eta}$=199)	0.997 +/- 0	0.996 +/- 0.001	0.914 +/- 0.01
40% noise, no denoising			
	Raw	Downsampled	Local Minima/Maxima
Chaotic Lorenz system ($\bar{\eta}$=17)	0.998 +/- 0	0.998 +/- 0	0.998 +/- 0.001
Periodic transcription ($\bar{\eta}$=12)	0.997 +/- 0	0.982 +/- 0.003	0.998 +/- 0
Chaotic transcription ($\bar{\eta}$=26)	0.998 +/- 0	0.997 +/- 0.001	0.998 +/- 0
40% noise, Schreiber denoising			
	Raw	Downsampled	Local Minima/Maxima
Chaotic Lorenz system ($\bar{\eta}$=46)	0.998 +/- 0	0.994 +/- 0	0.997 +/- 0
Periodic transcription ($\bar{\eta}$=22)	0.991 +/- 0	0.364 +/- 0.01	0.998 +/- 0
Chaotic transcription ($\bar{\eta}$=94)	0.998 +/- 0	0.996 +/- 0.001	0.998 +/- 0

SUPPLEMENTARY TABLE 3.13: Sampling statistic in empirical (i.e. non-simulated) deterministic datasets. Because $\eta < 10$ in these datasets, none of them would be downsampled by the Chaos Decision Tree Algorithm.

System	
Neuron integrated circuit (chaotic)	$\bar{\eta}$=1.4
Neuron integrated circuit (SNA)	$\bar{\eta}$=1.9
Neuron integrated circuit (periodic)	$\bar{\eta}$=1.7
Laser [160] (chaotic)	η=7.9
Stellar flux [161] (SNA)	η=7

SUPPLEMENTARY TABLE 3.14: Classification accuracy of BenSaïda's implementation of the Shintani-Linton neural network-based chaos-detection algorithm [113] in raw (non-denoised), deterministic, simulated datasets. Systems were classified as periodic if their estimated largest Lyapunov expoent was zero or negative, and otherwise were classified as chaotic. Note that performance seems to go up at higher noise levels for periodic systems. This is likely because the algorithm is mis-classifying these datasets as noise; pure noise has an infinite largest Lyapunov exponent, but neural network-based largest Lyapunov exponent estimators can often assign negative Lyapunov exponents to noise.

	Measurement noise level (% of std. dev.)				
System	0%	10%	20%	30%	40%
Cortical model (chaotic)	49/100	100/100	100/100	100/100	100/100
Cortical model (periodic)	0/100	0/100	0/100	0/100	0/100
Spiking neuron (chaotic)	100/100	100/100	100/100	100/100	100/100
Granulocyte levels (chaotic)	0/100	8/100	1/100	2/100	0/100
Granulocyte levels (periodic)	0/100	39/100	99/100	100/100	100/100
NF-κB transcription (chaotic)	0/100	0/100	0/100	0/100	0/100
NF-κB transcription (periodic)	100/100	100/100	100/100	100/100	100/100
Cubic map (chaotic)	86/100	77/100	30/100	0/100	0/100
Cubic map (periodic)	0/100	0/100	22/100	98/100	100/100
Cubic map (SNA HH)	0/100	0/100	0/100	44/100	88/100
Cubic map (SNA S3)	95/100	98/100	100/100	100/100	100/100
GOPY map (SNA)	0/100	0/100	0/100	76/100	100/100
Logistic map (chaotic)	100/100	99/100	100/100	95/100	23/100
Logistic map (periodic)	0/100	0/100	0/100	5/100	4/100
Lorenz system (chaotic)	0/100	0/100	0/100	0/100	0/100
Generalized Hénon map (hyperchaotic)	100/100	100/100	100/100	100/100	100/100

SUPPLEMENTARY TABLE 3.15: Classification accuracy of BenSaïda's implementation of the Shintani-Linton neural network-based algorithm [113] in deterministic, simulated datasets after Schreiber de-noising.

	Measurement noise level (% of std. dev.)				
System	0%	10%	20%	30%	40%
Cortical model (chaotic)	66/100	0/100	0/100	0/100	0/100
Cortical model (periodic)	0/100	0/100	0/100	0/100	0/100
Spiking neuron (chaotic)	100/100	100/100	100/100	100/100	100/100
Granulocyte levels (chaotic)	0/100	7/100	7/100	1/100	0/100
Granulocyte levels (periodic)	0/100	0/100	2/100	83/100	100/100
NF-κB transcription (chaotic)	0/100	0/100	0/100	0/100	0/100
NF-κB transcription (periodic)	100/100	100/100	100/100	100/100	100/100
Cubic map (chaotic)	80/100	66/100	47/100	15/100	4/100
Cubic map (periodic)	0/100	0/100	0/100	0/100	0/100
Cubic map (SNA HH)	0/100	0/100	0/100	0/100	2/100
Cubic map (SNA S3)	53/100	70/100	88/100	83/100	96/100
GOPY map (SNA)	0/100	0/100	0/100	2/100	100/100
Logistic map (chaotic)	27/100	14/100	93/100	100/100	100/100
Logistic map (periodic)	0/100	0/100	0/100	0/100	0/100
Lorenz system (chaotic)	0/100	0/100	0/100	0/100	0/100
Generalized Hénon map (hyperchaotic)	100/100	100/100	100/100	100/100	100/100

SUPPLEMENTARY TABLE 3.16: Classification accuracy of the NOLDS Python library implementation of the Rosenstein algorithm [114] in raw (non-denoised), deterministic, simulated datasets. Datasets were classified as chaotic if their estimated largest Lyapunov exponent was positive, and periodic if their estimated largest Lyapunov exponent was zero or negative.

	Measurement noise level (% of std. dev.)				
System	0%	10%	20%	30%	40%
Cortical model (chaotic)	100/100	100/100	100/100	100/100	100/100
Cortical model (periodic)	100/100	100/100	100/100	100/100	100/100
Spiking neuron (chaotic)	0/100	0/100	0/100	0/100	0/100
Granulocyte levels (chaotic)	4/100	7/100	10/100	6/100	9/100
Granulocyte levels (periodic)	100/100	90/100	90/100	90/100	90/100
NF-κB transcription (chaotic)	100/100	100/100	100/100	100/100	100/100
NF-κB transcription (periodic)	0/100	0/100	0/100	0/100	0/100
Cubic map (chaotic)	100/100	100/100	100/100	100/100	100/100
Cubic map (periodic)	6/100	78/100	72/100	99/100	100/100
Cubic map (SNA HH)	0/100	86/100	100/100	100/100	100/100
Cubic map (SNA S3)	0/100	0/100	0/100	0/100	0/100
GOPY map (SNA)	0/100	18/100	0/100	0/100	0/100
Logistic map (chaotic)	29/100	23/100	19/100	29/100	42/100
Logistic map (periodic)	100/100	90/100	92/100	97/100	90/100
Lorenz system (chaotic)	2/100	0/100	0/100	1/100	1/100
Generalized Hénon map (hyperchaotic)	0/100	0/100	0/100	0/100	0/100

SUPPLEMENTARY TABLE 3.17: Classification accuracy of the NOLDS Python library implementation of the Rosenstein algorithm [114] in deterministic, simulated datasets after Schreiber de-noising. Datasets were classified as chaotic if their estimated largest Lyapunov exponent was positive, and periodic if their estimated largest Lyapunov exponent was zero or negative.

	Measurement noise level (% of std. dev.)				
System	0%	10%	20%	30%	40%
Cortical model (chaotic)	80/100	100/100	100/100	100/100	100/100
Cortical model (periodic)	100/100	100/100	100/100	100/100	100/100
Spiking neuron (chaotic)	0/100	0/100	0/100	0/100	0/100
Granulocyte levels (chaotic)	87/100	85/100	84/100	56/100	22/100
Granulocyte levels (periodic)	100/100	85/100	90/100	87/100	85/100
NF-κB transcription (chaotic)	0/100	0/100	0/100	0/100	0/100
NF-κB transcription (periodic)	100/100	100/100	100/100	100/100	100/100
Cubic map (chaotic)	100/100	100/100	100/100	100/100	100/100
Cubic map (periodic)	38/100	3/100	12/100	70/100	66/100
Cubic map (SNA HH)	0/100	1/100	6/100	42/100	88/100
Cubic map (SNA S3)	0/100	0/100	0/100	0/100	0/100
GOPY map (SNA)	0/100	33/100	0/100	0/100	0/100
Logistic map (chaotic)	23/100	18/100	17/100	14/100	24/100
Logistic map (periodic)	100/100	84/100	86/100	86/100	89/100
Lorenz system (chaotic)	1/100	2/100	2/100	0/100	0/100
Generalized Hénon map (hyperchaotic)	0/100	0/100	0/100	0/100	0/100

SUPPLEMENTARY TABLE 3.18: Classification accuracy of the NOLDS Python library implementation of the Eckman algorithm [115] in raw (non-denoised), deterministic, simulated datasets. Datasets were classified as chaotic if their estimated largest Lyapunov exponent was positive, and periodic if their estimated largest Lyapunov exponent was zero or negative. With the exception of the noise-free simulation of periodic granulocyte levels, the algorithm yielded positive Lyapunov exponents for *all* datasets, hence the seemingly perfect classification accuracy for chaotic systems and the total inaccuracy for periodic systems.

	Measurement noise level (% of std. dev.)				
System	0%	10%	20%	30%	40%
Cortical model (chaotic)	100/100	100/100	100/100	100/100	100/100
Cortical model (periodic)	0/100	0/100	0/100	0/100	0/100
Spiking neuron (chaotic)	100/100	100/100	100/100	100/100	100/100
Granulocyte levels (chaotic)	100/100	100/100	100/100	100/100	100/100
Granulocyte levels (periodic)	100/100	0/100	0/100	0/100	0/100
NF-κB transcription (chaotic)	100/100	100/100	100/100	100/100	100/100
NF-κB transcription (periodic)	0/100	0/100	0/100	0/100	0/100
Cubic map (chaotic)	100/100	100/100	100/100	100/100	100/100
Cubic map (periodic)	0/100	0/100	0/100	0/100	0/100
Cubic map (SNA HH)	0/100	0/100	0/100	0/100	0/100
Cubic map (SNA S3)	0/100	0/100	0/100	0/100	0/100
GOPY map (SNA)	0/100	0/100	0/100	0/100	0/100
Logistic map (chaotic)	100/100	100/100	100/100	100/100	100/100
Logistic map (periodic)	0/100	0/100	0/100	0/100	0/100
Lorenz system (chaotic)	100/100	100/100	100/100	100/100	100/100
Generalized Hénon map (hyperchaotic)	100/100	100/100	100/100	100/100	100/100

SUPPLEMENTARY TABLE 3.19: Classification accuracy of the NOLDS Python library implementation of the Eckman algorithm [115] in deterministic, simulated datasets after Schreiber de-noising. Datasets were classified as chaotic if their estimated largest Lyapunov exponent was positive, and periodic if their estimated largest Lyapunov exponent was zero or negative.

	Measurement noise level (% of std. dev.)				
System	0%	10%	20%	30%	40%
Cortical model (chaotic)	100/100	100/100	100/100	100/100	100/100
Cortical model (periodic)	0/100	0/100	0/100	0/100	0/100
Spiking neuron (chaotic)	100/100	100/100	100/100	100/100	100/100
Granulocyte levels (chaotic)	100/100	100/100	100/100	100/100	100/100
Granulocyte levels (periodic)	0/100	0/100	0/100	0/100	0/100
NF-κB transcription (chaotic)	100/100	100/100	100/100	100/100	100/100
NF-κB transcription (periodic)	0/100	0/100	0/100	0/100	0/100
Cubic map (chaotic)	100/100	100/100	100/100	100/100	100/100
Cubic map (periodic)	0/100	0/100	0/100	0/100	0/100
Cubic map (SNA HH)	0/100	0/100	0/100	0/100	0/100
Cubic map (SNA S3)	0/100	0/100	0/100	0/100	0/100
GOPY map (SNA)	0/100	0/100	0/100	0/100	0/100
Logistic map (chaotic)	100/100	100/100	100/100	100/100	100/100
Logistic map (periodic)	0/100	0/100	0/100	0/100	0/100
Lorenz system (chaotic)	100/100	100/100	100/100	100/100	100/100
Generalized Hénon map (hyperchaotic)	100/100	100/100	100/100	100/100	100/100

Supplementary Table 3.20: Spearman correlations between largest Lyapunov exponents and permutation entropy calculated from raw data (i.e. data that were not de-noised and downsampled). While the correlations are still strong and significant for the discrete-time logistic and Hénon maps, performance is very poor for the continuous Lorenz system and mean-field cortical model. This is to be expected, because permutation entropy is equivalent to Kolmogorov-Sinai entropy (which is upper-bounded by a system's positive Lyapunov exponents) for *discrete* systems [176], and downsampling is effectively a discrete mapping of a continuous process. This is the same reason that de-noising and downsampling improves performance of the 0-1 test for continuous systems (Supplementary Table 3.6). Note that largest Lyapunov exponents in the cortical model are rough approximations (see Methods). *** p<0.001 after Bonferroni-correcting for multiple comparisons to the same set of ground-truth largest Lyapunov exponents.

	Measurement noise level (% of std. dev.)				
System	0%	10%	20%	30%	40%
Logistic map	0.97***	0.94***	0.94***	0.94***	0.94***
Hénon map	0.93***	0.92***	0.92***	0.88***	0.87***
Lorenz system	0.76***	-0.72***	-0.66***	-0.55***	-0.42***
Cortical model	0.15***	0.28***	0.01	0.01	0.04
Neuron integrated circuit	0.93***				

Supplementary Table 3.21: Accuracy of the (automated) Chaos Decision Tree Algorithm for different variables of the multi-dimensional systems analyzed in this paper. For the Lorenz system, the Rössler system, the Ikeda map, the Hénon map, the periodic cubic map, the strange non-chaotic cubic maps, and the period-doubled cubic maps, classification accuracy was high for any variable in the system, for relatively low levels of measurement noise. For higher levels of measurement noise, performance dropped for individual variables of the Lorenz, Rössler, GOPY map, and strange non-chaotic cubic map systems.

	Measurement noise level (% of std. dev.)				
System	0%	10%	20%	30%	40%
Lorenz system, **x** variable (chaotic)	100/100	100/100	98/100	77/100	34/100
Lorenz system, **y** variable (chaotic)	100/100	100/100	99/100	75/100	44/100
Lorenz system, **z** variable (chaotic)	100/100	100/100	99/100	75/100	44/100
Rössler system, **x** variable (chaotic)	37/100	57/100	88/100	100/100	100/100
Rössler system, **y** variable (chaotic)	97/100	98/100	99/100	100/100	100/100
Rössler system, **z** variable (chaotic)	100/100	51/100	92/100	33/100	16/100
GOPY system, **x** variable (SNA)	100/100	100/100	1000/100	83/100	21/100
Ikeda map, **x** variable (chaotic)	100/100	100/100	100/100	100/100	100/100
Ikeda map, **y** variable (chaotic)	100/100	100/100	100/100	100/100	100/100
Hénon map, **x** variable (periodic)	100/100	81/100	100/100	100/100	100/100
Hénon map, **y** variable (periodic)	100/100	100/100	100/100	100/100	100/100
Cubic map, **x** variable (chaotic)	90/100	93/100	89/100	88/100	99/100
Cubic map, **x** variable (periodic)	100/100	100/100	99/100	97/100	72/100
Cubic map, **x** variable (SNA HH)	100/100	100/100	100/100	100/100	41/100
Cubic map, **x** variable (SNA S3)	100/100	100/100	100/100	100/100	0/100
Cubic map, **x** variable (period-doubled)	100/100	100/100	100/100	100/100	32/100

SUPPLEMENTARY TABLE 3.22: Accuracy of the (automated) Chaos Decision Tree Algorithm for 1,000 time-points. Some datasets led to numerical errors during surrogate data generation; for these, we show classification accuracy out of the datasets that did not lead to errors.

	Measurement noise level (% of std. dev.)				
System	0%	10%	20%	30%	40%
Cortical model (chaotic)	100/100	100/100	68/100	24/100	51/100
Cortical model (periodic)	100/100	100/100	100/100	100/100	32/100
Spiking neuron (chaotic)	69/100	77/100	85/100	84/100	25/100
Granulocyte levels (chaotic)	100/100	100/100	100/100	100/100	100/100
Granulocyte levels (periodic)	100/100	100/100	100/99	99/100	94/100
NF-κB transcription (chaotic)	95/100	68/100	55/100	40/100	23/100
NF-κB transcription (periodic)	0/100	0/100	0/100	0/100	2/100
Cubic map (chaotic)	100/100	100/100	100/100	100/100	99/100
Cubic map (periodic)	100/100	100/100	100/100	97/100	52/100
Cubic map (SNA HH)	100/100	100/100	100/100	100/100	100/100
Cubic map (SNA S3)	100/100	100/100	100/100	100/100	55/100
GOPY map (SNA)	0/100	0/100	0/100	0/100	0/100
Logistic map (chaotic)	100/100	94/100	55/100	12/100	4/100
Logistic map (periodic)	100/100	100/100	100/100	100/100	98/99
Lorenz system (chaotic)	0/100	1/100	3/100	0/100	0/100
Generalized Hénon map (hyperchaotic)	100/100	100/100	99/100	34/100	1/100
Freitas map (nonlinear stochastic)	100/100	100/100	99/100	99/100	100/100
Noise-driven sine map (nonlinear stochastic)	9/83	65/85	79/90	98/100	100/100
Bounded random walk (nonlinear stochastic)	100/100	100/100	100/100	100/100	100/100
Cyclostationary process (linear stochastic)	93/100	97/100	98/100	99/100	100/100
ARMA(1) process (linear stochastic)	66/100	97/100	100/100	100/100	100/100
Trended random walk (linear stochastic)	98/100	97/100	100/100	100/100	100/100
Random walk (linear stochastic)	100/100	99/100	100/100	100/100	100/100
Rössler system (chaotic)	18/100	23/100	54/100	78/100	25/100
Ikeda map (chaotic)	100/100	100/100	78/100	11/100	24/100
Hénon map (periodic)	100/100	100/100	100/100	100/100	100/100
Cubic map (period-doubled)	99/100	95/100	60/100	12/100	0/100
Poincaré oscillator (periodic)	100/100	100/100	100/100	100/100	100/100
Poincaré oscillator (quasi-periodic)	100/100	100/100	100/100	100/99	73/100
Poincaré oscillator (chaotic)	100/100	100/100	100/100	96/100	67/100
Multivariate AR model (linear stochastic)	100/100	100/100	99/100	100/100	100/100
Violet noise (linear stochastic)	100/100				
Blue noise (linear stochastic)	100/100				
White noise (linear stochastic)	100/100				
Pink noise (linear stochastic)	100/100				
Red noise (linear stochastic)	99/100				

Supplementary Table 3.23: Accuracy of the (automated) Chaos Decision Tree Algorithm for 5,000 time-points.

	Measurement noise level (% of std. dev.)				
System	0%	10%	20%	30%	40%
Cortical model (chaotic)	100/100	100/100	100/100	62/100	91/100
Cortical model (periodic)	100/100	100/100	100/100	100/100	100/100
Spiking neuron (chaotic)	94/100	93/100	98/100	99/100	45/100
Granulocyte levels (chaotic)	100/100	100/100	100/100	100/100	100/100
Granulocyte levels (periodic)	100/100	100/100	100/100	100/100	100/100
NF-κB transcription (chaotic)	97/100	98/100	98/100	99/100	99/100
NF-κB transcription (periodic)	100/100	100/100	100/100	89/100	100/100
Cubic map (chaotic)	100/100	100/100	100/100	100/100	100/100
Cubic map (periodic)	100/100	100/100	100/100	100/100	100/100
Cubic map (SNA HH)	100/100	100/100	100/100	100/100	99/100
Cubic map (SNA S3)	100/100	100/100	100/100	100/100	0/100
GOPY map (SNA)	29/100	21/100	6/100	2/100	14/100
Logistic map (chaotic)	100/100	100/100	100/100	100/100	92/100
Logistic map (periodic)	100/100	100/100	100/100	100/100	100/100
Lorenz system (chaotic)	70/100	59/100	42/100	18/100	4/100
Generalized Hénon map (hyperchaotic)	100/100	100/100	100/100	100/100	26/100
Freitas map (nonlinear stochastic)	97/100	100/100	97/100	96/100	87/100
Noise-driven sine map (nonlinear stochastic)	38/100	22/100	56/100	50/100	96/100
Bounded random walk (nonlinear stochastic)	98/100	97/100	99/100	100/100	100/100
Cyclostationary process (linear stochastic)	97/100	98/100	100/100	100/100	100/100
ARMA(1) process (linear stochastic)	72/100	99/100	100/100	100/100	100/100
Trended random walk (linear stochastic)	98/100	89/100	98/100	100/100	100/100
Random walk (linear stochastic)	100/100	98/100	100/100	100/100	100/100
Rössler system (chaotic)	37/100	55/100	91/100	99/100	100/100
Ikeda map (chaotic)	100/100	100/100	100/100	93/100	23/100
Hénon map (periodic)	100/100	100/100	100/100	100/100	100/100
Cubic map (period-doubled)	100/100	100/100	100/100	100/100	100/100
Poincaré oscillator (periodic)	100/100	100/100	100/100	100/100	100/100
Poincaré oscillator (quasi-periodic)	100/100	100/100	100/100	100/100	100/100
Poincaré oscillator (chaotic)	100/100	100/100	100/100	100/100	100/100
Multivariate AR model (linear stochastic)	100/100	100/100	100/100	100/100	100/100
Violet noise (linear stochastic)	100/100				
Blue noise (linear stochastic)	100/100				
White noise (linear stochastic)	100/100				
Pink noise (linear stochastic)	100/100				
Red noise (linear stochastic)	98/100				

SUPPLEMENTARY TABLE 3.24: Fraction of datasets of noise-driven chaotic systems classified as stochastic, for different levels of intrinsic noise and different system observables. For both the stochastic Lorenz and stochastic Rössler systems, the parameter A controls the amplitude of intrinsic white noise injected into the **x** variable of the system (see Methods). The **x** (i.e. noise-driven) variable of the stochastic Lorenz system is classified as stochastic for any level of intrinsic noise, and classifications of stochasticity became more frequent for the **y**, **z**, and **x**+**y** variables with higher levels of intrinsic noise, particularly in the presence of additional measurement noise. For the stochastic Rössler system, increasing levels of intrinsic noise in the **x** variable led to more frequent classifications of stochasticity in all variables; interestingly, although noise was injected into the **x** variable, classifications of stochasticity became particularly frequent in the **y**, **z**, and **x**+**y** variables as the level of intrinsic noise was increased, especially (again) in the presence of additional measurement noise. Note that the level of noise injected into the Rössler system was lower than the level of noise injected into the Lorenz system, as we found that higher levels of dynamic noise led to numerical errors in the integration of the Rössler system.

	Measurement noise level (% of std. dev.)				
System	0%	10%	20%	30%	40%
Stochastic Lorenz, **x** variable (A=.5)	61/100	85/100	92/100	98/100	100/100
Stochastic Lorenz, **y** variable (A=.5)	8/100	48/100	65/100	94/100	100/100
Stochastic Lorenz, **z** variable (A=.5)	0/100	21/100	30/100	53/100	76/100
Stochastic Lorenz, **x**+**y** variable (A=.5)	45/100	60/100	73/100	97/100	99/100
Stochastic Lorenz, **x** variable (A=1.5)	98/100	82/100	93/100	99/100	100/100
Stochastic Lorenz, **y** variable (A=1.5)	1/100	0/100	4/100	62/100	99/100
Stochastic Lorenz, **z** variable (A=1.5)	0/100	81/100	84/100	95/100	95/100
Stochastic Lorenz, **x**+**y** variable (A=1.5)	35/100	9/100	29/100	86/100	99/100
Stochastic Lorenz, **x** variable (A=2.5)	97/100	99/100	99/100	100/100	100/100
Stochastic Lorenz, **y** variable (A=2.5)	0/100	0/100	0/100	34/100	90/100
Stochastic Lorenz, **z** variable (A=2.5)	0/100	91/100	66/100	87/100	99/100
Stochastic Lorenz, **x**+**y** variable (A=2.5)	40/100	28/100	46/100	92/100	99/100
Stochastic Rössler **x** variable (A=.05)	0/100	0/100	0/100	0/100	71/100
Stochastic Rössler **y** variable (A=.05)	0/100	0/100	78/100	0/100	0/100
Stochastic Rössler **z** variable (A=.05)	0/100	50/100	80/100	99/100	99/100
Stochastic Rössler **x**+**y** variable (A=.05)	0/100	0/100	0/100	72/100	0/100
Stochastic Rössler **x** variable (A=.15)	0/100	0/100	0/100	0/100	0/100
Stochastic Rössler **y** variable (A=.15)	0/100	0/100	0/100	0/100	0/100
Stochastic Rössler **z** variable (A=.15)	8/100	99/100	100/100	100/100	100/100
Stochastic Rössler **x**+**y** variable (A=.15)	0/100	0/100	0/100	0/100	0/100
Stochastic Rössler **x** variable (A=.25)	0/100	0/100	7/100	38/100	62/100
Stochastic Rössler **y** variable (A=.25)	87/100	96/100	98/100	97/100	99/100
Stochastic Rössler **z** variable (A=.25)	3/100	99/100	100/100	99/100	100/100
Stochastic Rössler **x**+**y** variable (A=.25)	40/100	47/100	65/100	84/100	92/100

SUPPLEMENTARY TABLE 3.25: The results for the autoregressive moving-average (ARMA) processes in Supplementary Tables 3.1-3.4 and in Table 3.2 are for random values of the moving average parameter θ. We sought to further test whether the moving average parameter θ had any systematic effect on our algorithm's performance. Here, we set the parameter ϕ to 0.99 as we did in Supplementary Tables 3.1-3.4 and in Table 3.2, and tested our algorithm on ARMA(1) processes with four different values of θ: -0.5, 0, 0.5, and 0.9. 10,000 time-points were generated for each simulation. Performance was high for all parameters.

	Measurement noise level (% of std. dev.)				
System	0%	10%	20%	30%	40%
ARMA(1), θ=-0.5 (linear stochastic)	88/100	96/100	100/100	100/100	100/100
ARMA(1), θ=0 (linear stochastic)	71/100	97/100	100/100	100/100	99/100
ARMA(1), θ=0.5 (linear stochastic)	74/100	99/100	98/100	100/100	100/100
ARMA(1), θ=0.9 (linear stochastic)	90/100	98/100	100/100	100/100	100/100

3.8 Supplementary Notes

SUPPLEMENTARY GLOSSARY

Deterministic. A process is deterministic if, given the exact same initial conditions, it will always evolve over time in the exact same way (i.e. there is no randomness built in to its evolution).

Measurement noise. All empirical recordings are contaminated by some level of measurement or "observational" noise, which is noise that is not intrinsic to a system. In other words, a system could be entirely deterministic, but because of measurement error, a signal recorded from that system could be noisy. To simulate such measurement error, we added random white noise of varying amplitudes to the datasets in Tables 3.1-3.5 and Supplementary Tables 3.1-3.4, 3.8-3.19.

Dynamic noise. Some systems have noise *built in* to the dynamics of the system. Such dynamic or "intrinsic" noise could be negligible, in that it is washed out by a system's deterministic components (and the system can therefore be modeled, in theory, by deterministic equations). Other systems, on the other hand, are significantly affected by intrinsic noise. Such systems are considered **stochastic**: no matter their initial conditions, they will always evolve over time differently, because their dynamics have some intrinsic randomness. In single-neuron dynamics, for example, there may be inherent stochasticity because of the probabilistic gating of voltage-dependent ion channels [117], though such stochastic events may be "washed out" by predominantly deterministic processes on larger scales. Several such systems were analyzed in this paper, including the noise-driven sine map, the Freitas map, bounded random walks, random walks, a cyclostationary autoregressive process, an autoregressive moving-average process, a random multivariate autoregressive process, colored noise, the stochastic Lorenz system, the stochastic Rössler system, the North Atlantic Oscillation index, and essential and Parkinson's tremors. In general, it is difficult to distinguish these stochastic (i.e. intrinsically noisy) processes from deterministic processes that are contaminated by measurement noise; it is also difficult to distinguish either case from deterministic chaotic processes (see below for definition).

Linear. The state of a linear process is directly proportional to its inputs or previous state (e.g. $y = ax$).

Nonlinear. The state of a nonlinear process is *not* directly proportional to its inputs or previous state (e.g. $y = ax^2$).

Largest Lyapunov exponent. The largest Lyapunov exponent of a system quantifies the largest rate of divergence of initially infinitesimally close trajectories through **phase space** (see below for definition).

Chaotic. A system is chaotic if it is bounded, deterministic, nonlinear, and has a positive largest Lyapunov exponent, meaning that initially similar phase space trajectories diverge exponentially fast.

Periodic. A system is periodic if it is deterministic, nonlinear, and has a negative largest Lyapunov exponent, meaning that initially similar phase space trajectories remain close.

Quasiperiodic. The dynamics of a quasiperiodic system exhibit regular cycles like those of a periodic system; but, unlike a purely periodic system that stably revisits the same system states, quasiperiodic systems return to states that are similar but not identical to previous states. Quasiperiodic systems are also not chaotic, because they are not sensitive to initial conditions (i.e. they have a negative largest Lyapunov exponent).

Period-doubled. In many dynamical systems, modulation of a system parameter can lead to an abrupt change in the system's dynamics, such that it oscillates at twice its original period. These systems are periodic (i.e. they have a negative largest Lyapunov exponent).

Strange non-chaotic. A strange non-chaotic system has a strange (i.e. fractal) phase space attractor like a chaotic system, but a negative or zero largest Lyapunov exponent. It is generally difficult to experimentally distinguish strange non-chaotic systems from chaotic systems [299].

Hyperchaotic. A hyperchaotic system is a deterministic, nonlinear system with more than one positive Lyapunov exponent. These systems are generally difficult to distinguish from noise [157].

Colored noise. Colored noise refers to stochastic processes with a non-uniform power spectrum (i.e. different levels of power at different frequencies). It is difficult to distinguish colored noise from chaos [156, 157].

Degree of chaos. The magnitude of a system's largest Lyapunov exponent quantifies its degree of chaos. Higher largest Lyapunov exponents indicate higher degrees of chaos, because they indicate faster rates of divergence in phase space.

Nonlinear stochastic. A nonlinear stochastic system is a nonlinear system with randomness built in to its evolution over time, making it difficult to distinguish from chaotic

systems [157].

Stationarity. A stationary process is one whose joint probability distribution is time-invariant; in other words, for a stationary process, statistical properties like mean and variance do not fluctuate over time.

Attractor. An attractor is the orbit in phase space toward which a deterministic system tends to evolve. The attractors of chaotic systems are called "strange attractors" because they have a fractal structure.

Phase space. A space representing all possible states of a system. A single point in phase space corresponds to a single state of the system. For example, for one particle, a single point in phase space determines that particle's location and momentum. A dynamical system produces, in general, trajectories in its phase space, i.e., the system's state changes with time.

Schreiber de-noising algorithm. Almost two decades ago, Schreiber introduced a simple nonlinear noise-reduction algorithm [168], which replaces each point in a time-series with the average value of that point's "neighborhood" in phase space (see above for definition of phase space). The algorithm first uses delay coordinate embedding to create a map that is topologically equivalent to a system's ground-truth phase space attractor (a very common procedure in nonlinear time-series analysis [121]). Each point's neighborhood in phase space is defined by the number of steps k in that point's past and the number of steps l in that point's future that are used to construct embedding vectors, as well as the radius r of that point's neighborhood in phase space. The parameters k and l are set to 1, and the radius r is set to the standard deviation of the time-series.

Surrogate testing. A common approach for testing if a given time-series reflects a deterministic process is to create *surrogates* of that time-series, which share some key features with the original time series, such as its power spectrum and amplitude distribution, but are otherwise stochastic [165, 166]. A "test statistic" is then calculated for both the original time-series and for the set of surrogate datasets, and if the value of the test statistic for the original time-series lies outside the distribution of values for the surrogate datasets, then the original time-series likely reflects a deterministic process [165, 166]. We follow Zunino and Kulp [196] and use permutation entropy as our test statistic (see below). We further tested a range of surrogate data generation algorithms, and picked a combination

of Amplitude Adjusted Fourier Transform (AAFT) surrogates [165] and Cyclic Phase Permutation (CPP) surrogates [167], which led to the highest performance in detecting signal stochasticity (Supplementary Tables 3.2-3.3).

0-1 test for chaos. Gottwald and Melbourne's 0-1 test for chaos uses a given signal to drive a simple 2-dimensional system, and calculates the growth rate K of the mean square displacement of that system. K will approach 0 for periodic systems and will approach 1 for chaotic systems [295–299]. See Methods for more details. While the test has been used for some physics and engineering applications, it has seen only very limited application to biology [195].

Receiver operating characteristic (ROC) curve. An ROC curve assesses the accuracy of a binary classifier by plotting its true positive rate vs. false positive rate for different discrimination thresholds (in the case of the 0-1 test, the threshold in question is the cutoff for what K-statistic values that are classified as chaotic or as periodic). The more accurate a classifier is across discrimination thresholds, the larger its area under the curve in an ROC plot will be.

Permutation entropy. Permutation entropy is an extremely quick-to-compute and noise-robust measure of a signal's complexity [164]. Following Zunino and Kulp [196], the Chaos Decision Tree Algorithm uses permutation entropy to test if a time-series is deterministic or stochastic, by comparing the permutation entropy of a signal to the permutation entropies of its surrogates (Figure 3.1, Methods): if a given time-series reflects a predominantly deterministic process, then it will have a lower permutation entropy than its surrogates, since surrogates are inherently stochastic and therefore higher entropy than a matching deterministic (even deterministic chaotic) process. It's also for this reason that permutation entropy tracks degree of chaos [164, 175], as stronger chaos means less predictability, and therefore more entropy. More formally, we should in general expect a close relationship between permutation entropy and systems' degree of chaos, because permutation entropy is equivalent to **Kolmogorov-Sinai entropy** for a broad class of discrete-time dynamical systems [176, 202–204]. Kolmogorov-Sinai entropy is a measure of the entropy *rate* of a system, i.e. how much entropy a system generates per unit time. Kolmogorov-Sinai entropy tracks systems' degree of chaos, since it is upper-bounded by the sum of a system's positive Lyapunov exponents. This relationship is known as the "Pesin identity" [205].

While Kolmogorov-Sinai entropy is very difficult to measure from time-series data, its fast approximation through permutation entropy is not, making permutation entropy an ideal measure for the practical estimation of a system's degree of chaos. Although permutation entropy, like the 0-1 test, tracks chaos in discrete-time systems, we show that, just as with the 0-1 test, permutation entropy can also track chaos in *continuous* systems if signals from those systems have been properly downsampled (Table 3.5, Supplementary Table 3.14).

Chapter 4

The Cortex Maintains a High Information Capacity during Conscious States by Operating Near Edge-of-Chaos Criticality

4.1 Abstract

During normal waking states, the human cortex is able to process massive amounts of information. But, for reasons that are still unknown, this information-carrying capacity is consistently disrupted during the loss of consciousness. Motivated by the fact that systems poised at the critical point in between chaotic and non-chaotic behavior tend to exhibit a high capacity for information processing, it has been proposed that the electrodynamics of the healthy cortex operate near this edge-of-chaos critical point during conscious states, and that unconscious brain states constitute excursions away from this critical point, thereby disrupting cortical information capacity. While this hypothesis was previously difficult to test, recent advances in both cortical modeling and nonlinear time-series analysis methods allow for renewed empirical assessment of this proposal. Here, we apply recently developed mathematical tools to cortical electrophysiology recordings from waking, anesthesia, seizure, and psychedelic states, and present evidence that low-frequency macro-scale cortical dynamics operate near edge-of-chaos criticality, and that this is indeed associated with a high information capacity. We further show that transitions of

low-frequency activity away from this critical point during the loss of consciousness may explain the disruption of cortical information-carrying capacity during unconscious states.

4.2 Introduction

Human consciousness is thought to rely on information processing in the cortex. This is why, it has been argued [229, 230], cortical information-holding capacity is disrupted whenever we lose consciousness. Empirically, it has been observed that the Lempel-Ziv complexity [231] (a measure of information capacity [232]) of macro-scale cortical electrodynamics, as recorded using electroencephalography (EEG), magnetoencephalography (MEG), or electrocorticography (ECoG), is high when subjects are awake [233–240], in a waking psychedelic state [238, 241], or are locked-in [235], and is low when subjects are experiencing seizures [234, 236], are in dreamless sleep [239], in a vegetative state [235], in a minimally conscious state [235], or under propofol [233, 235, 237, 240, 242], midazolam [235], xenon [235], isoflourane [240], desflurane [240], or sevoflurane anesthesia [233, 240]. These findings are already of direct clinical impact, as they are prompting the use of Lempel-Ziv complexity and other information capacity measures in the diagnosis of pathological brain states [235, 243], but the mechanisms underlying these observed reductions of cortical information-carrying capacity during the loss of consciousness are currently unknown.

Because many systems exhibit their highest information-carrying capacity when their dynamics are poised near critical points or phase transitions (for e.g. the phase transition from liquid to gas [244] or paramagnetism to ferromagnetism [245]), some have pointed [230] to the so-called "critical brain hypothesis" [246, 247] to explain this loss of cortical information-carrying capacity during states of unconsciousness. Tell-tale signs of criticality (for e.g. power-law statistics [248]) have been identified in the neural electrodynamics of several species [249–257], and, most important for the discussion at hand, these signatures of neural criticality have been associated with high information capacity in both real and simulated neural activity [246, 258–260] - a finding which underpins the proposal that criticality might explain the high information-carrying capacity of cortical dynamics during conscious states [230]. Moreover, signatures of neural criticality seem to disappear under anesthesia [259, 261] and seizures [262–266] (though others find evidence of criticality during seizures [267, 268]), which may explain [230] why both states are associated with a *loss* of information capacity [233–237, 240, 242]. Finally, there is evidence that cortical dynamics are even closer to criticality during psychedelic states than in normal waking states [269], which has been taken to explain [230, 269] the puzzling but consistent observation of *higher* cortical information entropy during psychedelic states than during normal

waking states [238, 241, 270]. In summary, it is possible that cortical dynamics exhibit a high information-carrying capacity during conscious states (like normal waking states and psychedelic states) because dynamics in these states are are critical or near-critical. Moreover, it is possible that unconscious states (like generalized seizures and anesthesia) constitute deviations away from criticality, which disrupts cortical information-carrying capacity.

While the criticality hypothesis is a potentially promising explanation for the observed variance of information capacity across brain states, it leaves open a fundamental question: what, precisely, is this critical point? And where do unconscious brain states like anesthesia, seizures, and disorders of consciousness fall relative to that critical point? One long-standing but difficult-to-test proposal is that cortical dynamics operate near a type of criticality called edge-of-chaos criticality or the "onset of chaos" [271–277], a dynamical regime poised in between chaos (i.e. dynamics that are exponentially sensitive to inputs) and periodicity (i.e. dynamics that are insensitive to inputs). At this critical point, systems tend to exhibit their highest information-carrying capacity, by balancing the stability of periodic systems with the sensitivity to inputs of chaotic systems [272, 273, 278–281].

Despite early enthusiasm for the proposal that cortical dynamics operate near the onset of chaos, the empirical use of chaos theory in neuroscience has proven difficult [282]. As has been pointed out before [282, 283], strong evidence of biological chaos or edge-of-chaos criticality requires: 1) a good theoretical model of a system's dynamics, 2) a demonstration that the theoretical model can display periodic, edge-of-chaos, or chaotic behavior as its biologically realistic parameters are varied, 3) the development of time-series analysis tools capable of distinguishing periodicity, chaos, and randomness from noisy recordings, and 4) the use of those tools to empirically confirm the model's predictions of periodicity or chaos in different system states. While early tools and neural models were unable to fulfill these criteria [282], significant progress has been made on all these fronts. Specifically, current models of macro-scale cortical electrodynamics are now able to recapitulate many signal properties present in neural electrophysiology recordings, such as the 1/f power spectrum with peaks at well-studied frequency bands, as well as changes to synchrony, the power spectrum, and traveling wave dynamics across different brain states that match empirical observations [276, 284–286]. Moreover, periodicity, edge-of-chaos criticality, and chaos have been thoroughly studied in these models as their biologically realistic parameters are varied [276], and recent developments in nonlinear time-series analysis now allow for fairly robust comparison of these model results to noisy empirical data. Thus, we are now in a better position to ask whether the cortex does indeed maintain a high-information carrying capacity during conscious states by operating near edge-of-chaos criticality.

Here, we present preliminary evidence that cortical electrodynamics during normal waking states do in fact operate in the regime of weak (stochastic) chaos, near the edge-of-chaos critical point. Moreover, our analyses suggest that GABAergic anesthetics push cortical dynamics into the strongly chaotic regime and that generalized seizures induce phase transitions into the periodic regime, both of which constitute excursions away from edge-of-chaos criticality, and thereby disrupt the information-carrying capacity of the cortex. Finally, we present evidence that psychedelics may increase the information-carrying capacity of the cortex by tuning cortical dynamis even closer to edge-of-chaos criticality.

4.3 Results

As mentioned in the Introduction, strong evidence of biological chaos requires comparison of real data to an accurate model of the system of interest (because chaos can only be detected with absolute certainty in simulations). Although there are many available models of macro-scale cortical electrodynamics (none of which can yet be considered a well-verified "standard model") [284–286], we here choose to study the mean-field model of Steyn-Ross, Steyn-Ross, and Sleigh [276] because it has been used successfully to simulate waking [276, 287], seizure [276, 288–291], and GABAergic anesthesia [276, 287, 292–294] states, and so putative model parameters for each of these states are available from prior literature. See Methods for details on the model equations and parameters. By tuning cortical excitation, inhibition, and gap junction coupling, the model's deterministic component can display periodic, critical, and chaotic behaviors (Figure 4.1). To test whether the model exhibits the predicted relationship between edge-of-chaos criticality and information-carrying capacity, we calculated both the Lempel-Ziv complexity of different model states, as well as the largest Lyapunov exponent of different model states (see Methods). The largest Lyapunov exponent measures degree of chaos, and a largest Lyapunov exponent of zero corresponds to edge-of-chaos criticality. Consistent with the prediction that the cortex maintains a high-information carrying capacity during waking by operating near the edge-of-chaos critical point, we found that the Lempel-Ziv complexity of the model's simulated electrodynamics was maximal when those dynamics were poised near the onset of chaos, and that the model's simulation of waking states was near this critical, high-information-capacity regime (Figure 4.1a). Moreover, the model's GABAergic anesthesia simulation was in the strong chaos regime, which, consistent with the criticality hypothesis, corresponded to a reduction in Lempel-Ziv complexity (Figure 4.1a). Finally, the model's generalized seizure state was in the periodic regime, which likewise corresponded to a reduction in Lempel-Ziv complexity, consistent with

the hypothesis that excursions away from edge-of-chaos criticality disrupt cortical information-carrying capacity (Figure 4.1a).

These simulation-based findings furnish us with several specific and counter-intuitive predictions: 1) propofol anesthesia cortical dynamics should be higher-chaos but lower-entropy than normal waking states, 2) generalized seizure dynamics should be lower-chaos and lower-entropy than normal waking states, and 3) states that increase cortical information entropy relative to normal waking states should correspond to moderate reductions in chaoticity, by tuning cortical dynamics even closer to edge-of-chaos criticality. As mentioned in the Introduction, such predictions of varying degrees of chaoticity in real biological systems have historically been difficult to test due to a lack of noise-robust chaos-detection tools [282, 283], but the development of novel time-series analysis methods allow us to investigate this relationship between the chaoticity of cortical dynamics and their information-carrying capacity in real data. To illustrate that we can do so, we applied the modified 0-1 test for chaos [295–300] (see Methods), which can track chaos in real systems with exceptionally high accuracy, to the mean-field model's simulated dynamics, and found that the test's outputted K-statistic can robustly track the degree of chaos of the mean-field model: the Pearson correlation between ground-truth largest Lyapunov exponents (which measure degree of chaos, and which can only be calculated in simulations - see Methods) and the median K-statistic (across the model's simulated macrocolumns) was 0.84 ($p< 10^{-205}$). Moreover, by using the K-statistic as opposed to the ground-truth largest Lyapunov exponents, we were able to recapitulate the inverse-U relationship between chaoticity and Lempel-Ziv complexity in the model (Figure 4.1b). This means that we can test this inverse-U relationship between chaos and information-carrying capacity in real cortical recordings. Thus, we specifically predict: 1) propofol anesthesia should increase K and decrease Lempel-Ziv complexity relative to normal waking states, 2) generalized seizure dynamics should decrease both K and Lempel-Ziv complexity relative to normal waking states, and 3) states that increase Lempel-Ziv complexity relative to normal waking states should correspond to moderate reductions in K, reflecting moderate reductions in chaoticity.

To test these predictions, we first applied these same methods - namely, the modified 0-1 test for chaos and the Lempel-Ziv complexity algorithm - to electrocorticography recordings from the left cortices of two macaques, from both waking (eyes open) and propofol anesthesia states. Data were split into 10-second trials. For Macaque 1, 106 10-second trials were available from the awake, eyes open state and 121 trials were available from the propofol anesthesia state. For Macaque 2, 102 10-second trials were available from the awake, eyes open state and 101 trials were available from the propofol anesthesia state. See Methods for details on data preprocessing and analysis. We replicated prior findings [233, 235, 237, 240, 242] of a reduction in

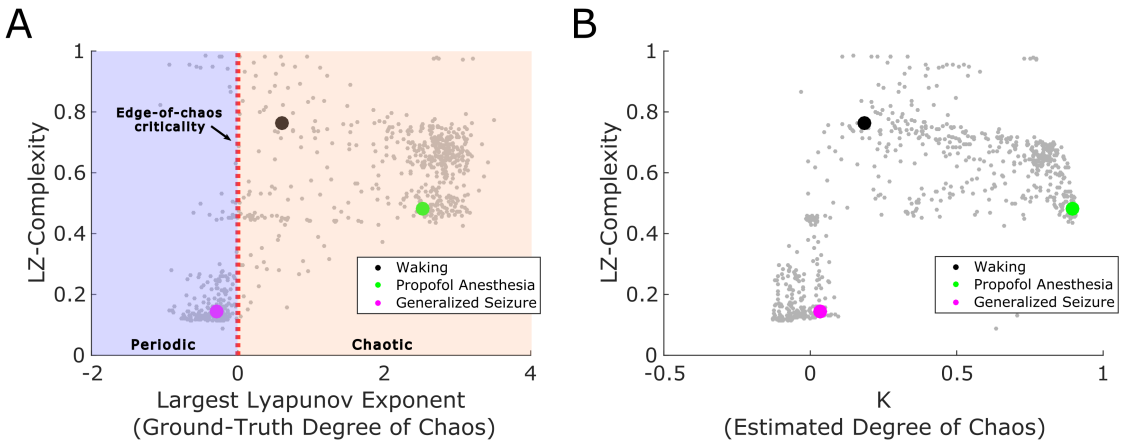


FIGURE 4.1: **A**. Like many simulated systems [272, 273, 278–280], a mean-field model of cortical electrodynamics exhibits its highest information-carrying capacity when its dynamics are poised near edge-of-chaos criticality (red dashed vertical line). Here, each gray dot corresponds to one simulation of the model, with different parameter configurations. The parameters we varied are postsynaptic excitation, postsynaptic inhibition, and gap junction coupling between inhibitory interneurons. The parameters corresponding to waking, GABAergic anesthesia, and generalized seizure states were taken from prior literature. We estimate the information-carrying capacity of the simulated cortical sheet using multivariate Lempel-Ziv complexity [231, 237], which, as a measure of signal compressibility, is bounded by the information (Shannon) entropy of the system, and therefore quantifies how much information the system can carry [232]. Degree of chaos is quantified by the largest Lyapunov exponent of the system, which is the rate of divergence between initially similar system states - see Methods. These simulation-based findings lead to three predictions: generalized seizures should correspond to a loss of chaoticity and a reduction in Lempel-Ziv complexity, propofol anesthesia should correspond to an increase in chaoticity and a reduction in Lempel-Ziv complexity, and states which increase Lempel-Ziv complexity relative to normal waking states should correspond, counter-intuitively, to moderate reductions in chaoticity (by tuning dynamics closer to edge-of-chaos criticality) **B**. While ground-truth degrees of chaos can only be determined in simulations (see Methods), we here show that we can test these three predictions using recently developed time-series analysis tools. Specifically, by applying the recently developed modified 0-1 test for chaos to the low-frequency component of the model's simulated dynamics (see Methods), we here show that we can recapitulate the inverse-U relationship between chaoticity and information-carrying capacity observed in **A**. The Spearman correlation between ground-truth largest Lyapunov exponents and K, which is the output of the 0-1 test, was 0.84 ($p < 10^{-205}$). This leads to the following testable predictions: generalized seizures should correspond to a significant drop in K and a reduction in Lempel-Ziv complexity, propofol anesthesia should correspond to an increase in K and a reduction in Lempel-Ziv complexity, and states which increase Lempel-Ziv complexity relative to normal waking states should correspond to moderate reductions of K. These predictions are tested in Figures 4.2-4.5.

Lempel-Ziv complexity during propofol anesthesia (Figure 4.2a) (two-sample, two-tailed t-tests comparing the Lempel-Ziv complexity of all waking trials to all propofol trials of Monkey 1: $p < 10^{-35}$, CI: [0.0314, 0.0411], of Monkey 2: $p < 10^{-23}$, CI: [0.025, 0.0358]), which indicates a loss of information-carrying capacity of cortical electrodynamics during the anesthetized state. We also confirmed the prediction that this decrease in Lempel-Ziv complexity during propofol anesthesia should correspond, counter-intuitively, to an *increase* in the chaoticity of cortical electrodynamics (Figure 4.2b) (two-sample, two-tailed t-tests comparing the median K-statistic of all waking trials to all propofol trials of Monkey 1: $p < 10^{-72}$, CI: [-0.0715, -0.0617], of Monkey 2: $p < 10^{-74}$, CI: [-0.0896, -0.0783]), suggesting that the loss of information-carrying capacity during this state might be mediated by an excursion away from edge-of-chaos criticality into the strongly chaotic regime. Finally, the change in the median power spectrum (across all electrodes, trials, and macaques) during propofol anesthesia relative to the waking baseline recapitulated the spectral changes in the GABAergic anesthesia state of the mean-field model, namely an increase in low-frequency power (Figures 4.2C-D), reaffirming that the model successfully recapitulates key features of the GABAergic anesthesia state (and thus lends credence to the model prediction that GABAergic anesthesia constitutes an excursion away from edge-of-chaos criticality into the strongly chaotic regime).

To then test the secondary prediction of this framework, which is that generalized seizures should correspond to a reduction in *both* Lempel-Ziv complexity and chaoticity, we next applied these same methods to ECoG recordings from two human epilepsy patients, during both waking baseline and generalized seizure states. Subject 1 experienced 3 seizures, and Subject 2 experienced 10 seizures. For both patients, seizures were focal with secondary generalizations. We isolated 10-second trials in which the seizures were fully generalized. For Subject 1, 77 10-second trials were available from the awake baseline state and 14 trials were available from the generalized seizure state. For Subject 2, 16 10-second trials were available from the awake baseline state and 18 trials were available from the generalized seizure state. See Methods for details on data preprocessing and analysis. We replicated prior findings of a significant reduction in Lempel-Ziv complexity during seizure states [234, 236] (Figure 4.3a) (two-sample, two-tailed t-tests comparing the Lempel-Ziv complexity of all waking trials to all seizure trials of subject 1: $p < 10^{-15}$, CI: [0.0468, 0.0713], of subject 2: $p = 0.0015$, CI: [0.0285, 0.109]). We also confirmed the mean-field model prediction of reduced chaoticity of cortical electrodynamics during these states (Figure 4.3b) (two-sample, two-tailed t-tests comparing the median K-statistic of all waking trials to all seizure trials of subject 1: $p < 10^{-37}$, CI: [0.3595, 0.4330], of subject 2: $p < 10^{-7}$, CI: [0.2848, 0.5507]). These results support the hypothesis that the loss of cortical information-carrying capacity during generalized seizures is mediated by a phase transition into the periodic regime, away from edge-of-chaos criticality. Finally, the change in the median

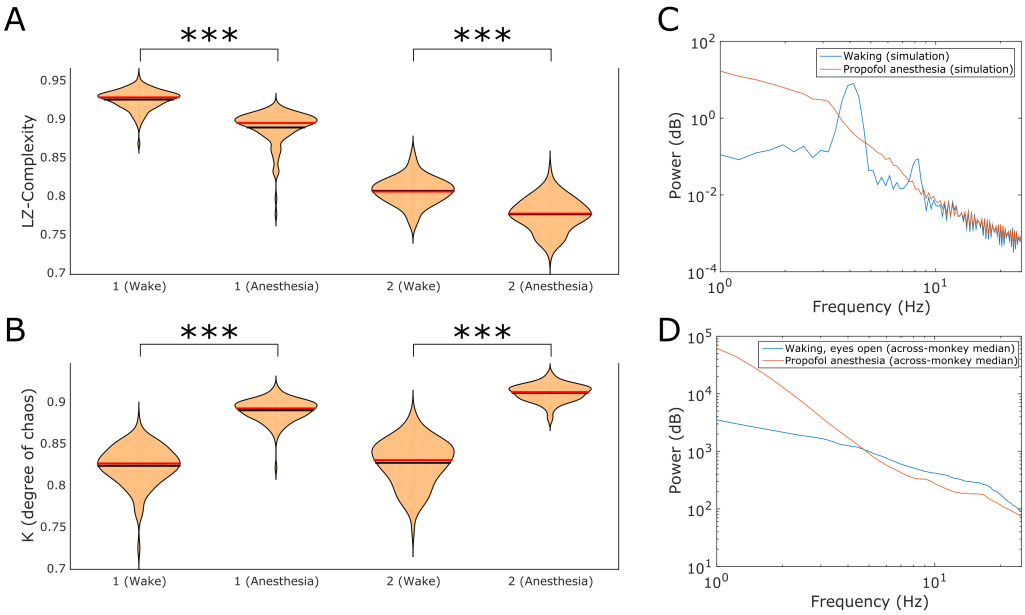


FIGURE 4.2: **A**. Replicating previously reported empirical results, we observed a significant reduction in the Lempel-Ziv complexity of 120 surface ECoG electrodes (spanning the left cortex) of two macaques during propofol anesthesia. From an information-theoretic perspective, this indicates a substantial loss of cortical information-carrying capacity during the anesthetized state. **B**. Consistent with the counter-intuitive model prediction that anesthesia reduces the information-carrying capacity of the cortex by increasing the chaoticity of its dynamics, we observed a significant increase in the chaoticity of cortical signals during propofol anesthesia, as measured by the K-statistic outputted by the modified 0-1 test for chaos. **C**. To further compare the mean-field model to our data, we here plot the median power spectrum across 120 simulated macrolumns in the model's anesthesia state (red), which shows a clear increase in low-frequency (<4hz) power relative to the median power spectrum of the waking simulation (blue). **D**. As was the case for the mean-field model, the median (across all electrodes and monkeys) power spectrum of cortical recordings during anesthesia (red) shows a clear increase in low-frequency (<4hz) power relative to waking states (blue).

power spectrum (across all recording channels, trials, and subjects) during generalized seizures relative to the waking baseline recapitulated the spectral changes in the generalized seizure state of the mean-field model (Figures 4.3C-D), namely a broadband power increase, as well as clear harmonics of the lowest-frequency oscillatory component (such harmonic behavior is a hallmark of periodic systems). This lends further credence to the model-based prediction that generalized seizures constitute excursions of cortical dynamics away from edge-of-chaos criticality into the periodic regime.

To test our third prediction, which is that increases in Lempel-Ziv complexity relative to normal waking states should, counter-intuitively, correspond to moderate reductions in the chaoticity of

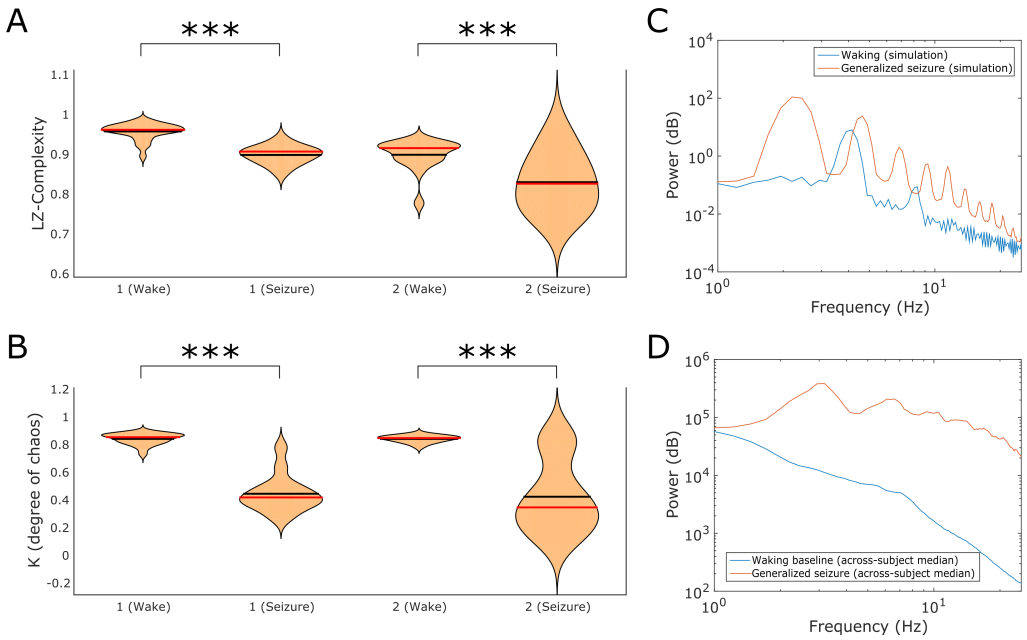


FIGURE 4.3: **A**. Confirming both model predictions and replicating prior results, we observed a significant reduction in Lempel-Ziv complexity across ECoG channels of two human epilepsy patients experiencing generalized seizures, suggesting a loss of cortical information-carrying capacity during this unconscious state. **B**. Confirming the model prediction that this loss of cortical information-carrying capacity during seizures is due to a loss of chaoticity, we also observed a significant decrease in the median K-statistic (outputted by the 0-1 test for chaos) of cortical electrodynamics during generalized seizures relative to conscious baselines. **C**. To further compare the mean-field model to our empirical data, we here plot the median power spectrum across 120 simulated macrolumns in the model's generalized seizure state (red), which shows a clear broadband power increase relative to the median power spectrum of the waking simulation (blue). Moreover, as is typical of periodic systems, the generalized seizure simulation power spectrum shows clear harmonics of its lowest frequency oscillation. **D**. As was the case for the mean-field model, the median (across channels, trials, and subjects) power spectrum of cortical recordings during generalized seizures (red) shows a increase in broadband power relative to baseline states (blue). Moreover, as was the case for the mean-field model, the seizure power spectrum shows clear harmonics of its lowest frequency oscillation, further supporting the conjecture that generalized seizures constitute an excursion away from edge-of-chaos criticality into the periodic regime.

cortical dynamics, we next applied the same analysis to MEG recordings from four human subjects following intraveneous administration of either a saline placebo or 2mg of the psychedelic compound lysergic acid diethylamide (LSD). The psychedelic state is the only known state under which cortical electrodynamics exhibit higher entropy than in the normal waking state [238, 241, 270], and prior studies offer preliminary evidence that this increase in cortical entropy may be mediated by a transition even closer to criticality [269]. If this prediction is true, then we should expect, counter-intuitively, a moderate reduction in the chaoticity of cortical electrodynamics during psychedelic states relative to normal waking states, even as Lempel-Ziv complexity is increased. To test this prediction, we re-analyzed previously published [301] MEG recordings of human subjects following administration of a placebo or LSD in different recording sessions, which we split into 10-second trials. For the placebo condition, 53 trials were available for Subject 1, 60 trials for Subject 2, 67 trials for Subject 3, and 59 trials for Subject 4. For the LSD condition, 63 trials were available for Subject 1, 42 trials for Subject 2, 67 trials for Subject 3, and 63 trials for Subject 4. We observed precisely the predicted result, with significant increases in Lempel-Ziv complexity in three out of four subjects (Figure 4.3a) (two-sample, two-tailed t-tests comparing the Lempel-Ziv complexity of all normal waking trials to all LSD trials of subject 1: $p < 10^{-18}$, CI: [-0.0280, -0.0189], of subject 2: $p < 10^{-4}$, CI: [-0.0189, -0.0158], of subject 3: $p = 0.5758$, CI: [-0.0058, 0.0033], of subject 4: $p < 10^{-9}$, CI: [-0.0190, -0.0099]) and significant decreases in the chaoticity of cortical dynamics in all subjects (Figure 4.3b) (two-sample, two-tailed t-tests comparing the median K-statistic of all waking trials to all LSD trials of subject 1: $p < 10^{-11}$, CI: [0.0653, 0.1155], of subject 2: $p < 10^{-5}$, CI: [0.0132, 0.0370], of subject 3: $p < 10^{-5}$, CI: [0.0210, 0.0557], of subject 4: $p = 0.0037$, CI: [0.0071, 0.0360]). Finally, although the psychedelic state has not been previously modeled using the mean-field model analyzed here, we note that the reduction in low-frequency power associated with the highest-entropy, nearest-to-criticality states of the model (Figure 4.4C) is consistent with the spectral changes observed in the psychedelic state relative to the placebo condition (Figure 4.4D), lending further support to the possibility that psychedelics tune cortical dynamics closer to edge-of-chaos criticality. Moreover, the critical, high-entropy states of the model corresponded to increases in cortical excitability coupled with slight decreases in the strength of gap junction coupling between inhibitory interneurons. While it is already known that psychedelics increase the frequency and amplitude of spontaneous excitatory postsynaptic potentials in cortical pyramidal cells via agonism of 5-HT2a receptors [302–305], it is currently unknown whether psychedelics have any effect on gap junction coupling. Based on these simulation results, together with prior empirical literature (see Discussion), it is therefore possible that psychedelics moderately inhibit cortical gap junction coupling - a possibility that will need to be investigated in future work.

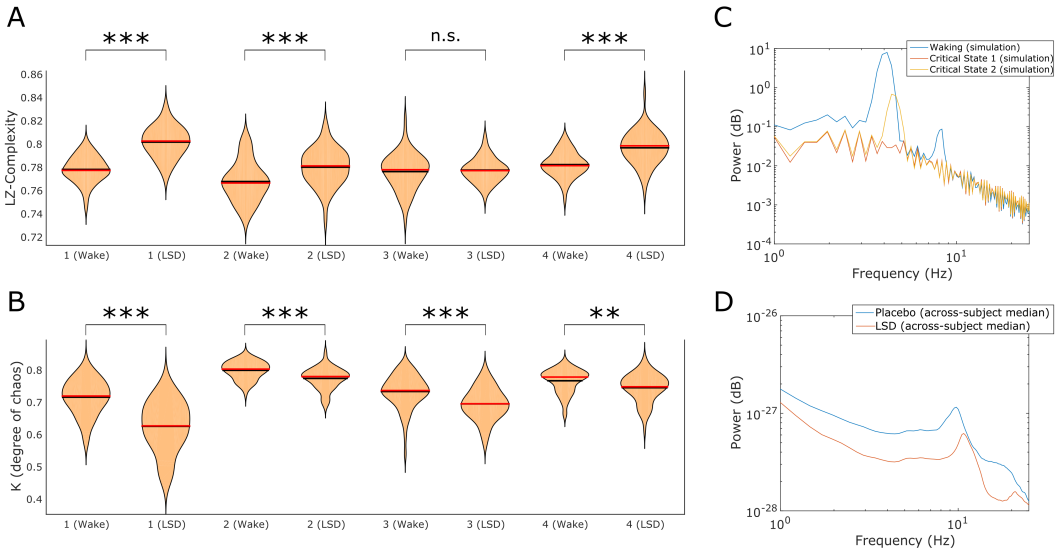


FIGURE 4.4: **A**. To test the model-based prediction that increases in cortical information-carrying capacity relative to normal waking states should correspond to a slight *reduction* in chaoticity, reflecting a transition closer to edge-of-chaos criticality, we analyzed MEG recordings from four human subjects following intravenous administration of either LSD or a saline placebo. The psychedelic state is the only known state in which cortical information-carrying capacity is higher than the normal waking state [238, 241, 270]. Replicating prior results [238], we observed a significant increase in Lempel-Ziv complexity across MEG channels of three out of four human subjects following administration of LSD, suggesting an increase in cortical information-carrying capacity during the psychedelic state. **B**. Confirming the counter-intuitive prediction that increases in the Lempel-Ziv complexity of cortical electrodynamics should be mediated by a slight decrease in chaoticity, we observed a significant decrease in the K-statistic (outputted by the modified 0-1 test for chaos) of cortical electrodynamics during the psychedelic state across all subjects. This lends support to the proposal that psychedelics increase cortical information capacity by tuning cortical electrodynamics closer to criticality [230, 269]. **C**. While the psychedelic state has not previously been formally studied in the mean-field model analyzed here, which means that we cannot easily pick out model parameters corresponding to the psychedelic state, we can examine the model parameters that tune its dynamics closest to critical, high-entropy states. Here, we plot the median (across simulated macrocolumns) power spectra of the only two parameter configurations whose resulting dynamics possessed a largest Lyapunov exponent of ± 0.1 (i.e., whose dynamics were close to edge-of-chaos criticality, which corresponds to a largest Lyapunov exponent of 0) and whose Lempel-Ziv complexity was greater than 0.9. In terms of parameter changes from the waking simulation, both critical states corresponded to an increase in cortical excitation as well as a slight decrease in gap junction coupling between inhibitory interneurons - see Discussion. Importantly, both critical states corresponded to reduced low-frequency power relative to the waking simulation. **D**. Consistent with the results for the critical, high-entropy states of the mean-field model, the median, cross-channel, cross-subject power spectrum of cortical recordings during the psychedelic state (red) shows a decrease in low-frequency power relative to placebo states (blue).

Finally, to visualize the relationship between chaos and information-carrying capacity across all datasets, we normalized each subject's Lempel-Ziv complexity and 0-1 chaos test results to their individual median normal waking baseline values. In Figure 4.5, we plot each subject's mean value, per condition. The observed inverse-U relationship between chaoticity and information-carrying capacity recapitulated the predicted relationship between these two features of cortical electrodynamics (Figure 4.1), and supports the hypothesis that cortical electrodynamics maintain a high information-carrying capacity during conscious states (i.e., normal waking and psychedelic states) by operating near edge-of-chaos criticality.

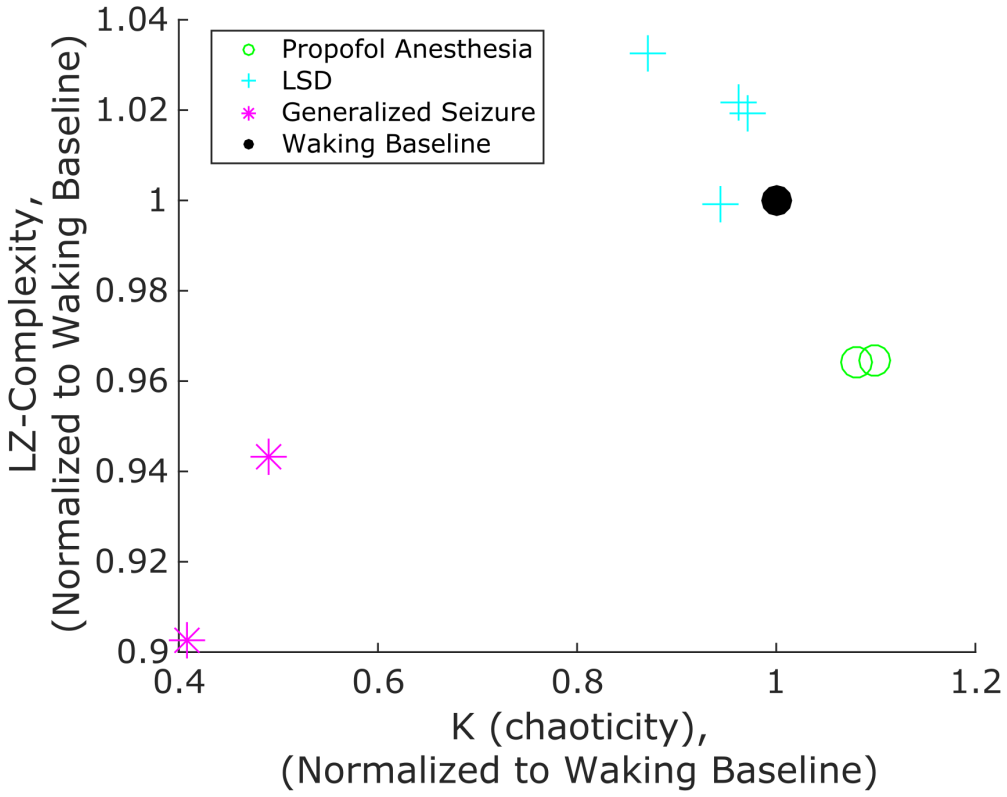


FIGURE 4.5: Normalizing Lempel-Ziv complexity and the K-statistic (the output of the modified 0-1 test for chaos) against each subject's individual normal waking baseline, and then averaging across all trials, per condition for each individual subject reveals precisely the predicted inverse-U relationship between chaoticity and information-carrying capacity. These results suggest that, as predicted by the mean-field model, cortical electrodynamics during normal waking states likely operate in the regime of weak chaos, close to edge-of-chaos criticality, and that this leads to a high information-carrying capacity. These results also support the mean-field model prediction that GABAergic anesthesia constitutes a transition into the strongly chaotic regime, which reduces the information-carrying capacity of the cortex, and that generalized seizures constitute a phase transition into periodicity, which also strongly reduces the information-carrying capacity of the cortex. Moreover, these results support the prediction that cortical electrodynamics during psychedelic states are even closer to criticality than they are during normal waking states, which may underpin the increaesd information-carrying capacity of cortical electrodynamics during these states.

4.4 Discussion

In this paper, we present evidence in favor of the hypothesis [230] that cortical electrodynamics exhibit a high information-carrying capacity during conscious states by operating near a critical point or phase transition. We specifically present evidence that the waking cortex operates near a type of phase transition called edge-of-chaos criticality or the "onset of chaos" - a dynamical regime in which many systems exhibit their highest information-carrying capacity [272, 273, 278–280], by balancing dynamical stability with responsiveness to their inputs [281]. We also present evidence that transitions away from this critical point - either into the strongly chaotic regime, as our evidence suggests is the case for GABAergic anesthesia, or into the periodic regime, as our evidence suggests is the case for generalized seizures - precipitate a loss of information-carrying capacity. Finally, we present evidence that transitions even closer to this critical point - as may be the case for the psychedelic state - precipitate an increase in the information-carrying capacity of macro-scale cortical electrodynamics.

It is worth pointing out how this work relates to the existing literature on neural criticality, considering that a central goal of our work was to elucidate what, precisely, is meant by criticality in the brain. First, it should be noted that one of the key hallmarks of neural criticality - namely, neural avalanches, which are bursts of neural activity whose size distributions obey power laws [249, 250, 253–255, 257–259, 306–308] - may not necessarily co-occur with the edge-of-chaos critical point studied here. In simulations of networks of neurons, it has been shown that the transition from sub-critical to avalanche to super-critical dynamics all occur within the chaotic phase of the system, which indicates that the onset of chaos and the appearance of avalanches are two distinct critical points [277]. Another question is how the periodic/chaotic transition maps onto the "ordered"/"disordered" transition more commonly invoked in the neural criticality literature. The finding reported here that cortical dynamics during conscious states operate on the chaotic side of edge-of-chaos criticality might seem at odds with the proposal that waking cortical dynamics operate on the "ordered" side of criticality [229]. While "chaos" and "disorder" have often been used interchangeably in this literature, this is in fact misleading: counter-intuitively, *chaos* is the "ordered" phase of a dynamical system with respect to edge-of-chaos criticality [309, 310], because a phase is considered "ordered" with respect to a critical point if it is characterized by whatever symmetry is broken at that critical point, and topological or de-Rahm supersymmetry of dynamical systems is broken at the periodic-to-chaotic transition [309–316] (a more intuitive example of this is ice water, which is "ordered" because it lacks the translational symmetry that makes liquid water "disordered," and that translational symmetry is spontaneously broken at the critical temperature of 0 degrees Celsius). Thus, our finding that cortical electrodynamics are weakly chaotic during conscious states is fully consistent (if

counter-intuitively so) with the proposal that waking cortical dynamics operate on the ordered side of criticality.

We acknowledge that the reported results are at odds with studies which have concluded that waking large-scale dynamics in the healthy human brain are approximately linear and stochastic [317–320], a possibility further supported by the empirical successes of neural time-series analysis algorithms that assume stochasticity and linearity [321]. While evidence of nonlinearity in macro-scale cortical electrodyamics has been reported for seizures [320], this has been taken to suggest that the transition from healthy to seizure states can be understood as a transition from linear to nonlinear dynamics [265, 266]. This would seem at odds with the proposal that waking cortical dynamics operate in a weakly chaotic state, close to edge-of-chaos criticality, which is a clearly nonlinear state. This apparent discrepancy, however, may simply be due to the erroneous classifications of EEG recordings as stochastic, owing to high levels of measurement noise in EEG data. We note that in the surrogate analyses conduced here, relatively few channels were classified as linear or stochastic (Tables 1-2).

Finally, it is necessary to emphasize that the results reported here are preliminary, and that significant work remains to be done to more rigorously test the proposed framework relating brain states, edge-of-chaos criticality, and cortical information-carrying capacity. We here outline potential follow-up studies. First, we point out a large methodological shortcoming of the work reported here, which is that analyses were carried out on cortical recordings from different species, imaging modalities, and laboratories. Because of this shortcoming, we designed our analyses as within-subject comparisons to normal, waking baselines (Figures 4.2-4.4). Where we report results across subjects, species, brain states, and imaging modalities, we normalized to each subject's individual normal waking baseline (Figure 4.5). A more rigorous test of this framework would be to eliminate these confounding differences by, for example, administering GABAergic anesthetics, psychedelic drugs, and seizure-inducers like pentylenetetrazole to the same cortical tissue, record from that tissue using the same equipment, and re-do the analyses reported here. Similarly, in a more controlled experimental setting, the degree of chaoticity of neural tissue in these different states might be assessed by measuring the consistency of the tissue's response to a repeated quasiperiodic input, as has been done to detect chaos in electronic circuits [322].

We suggest another follow-up to a result reported here, which is that the mean-field model of cortical electrodynamics exhibited its highest-entropy, nearest-to-critcality behavior when there was a moderate reduction in the strength of gap junction coupling between inhibitory interneurons as well as an increase in postsynaptic excitability - a parameter change which led to reduced spectral power at low frequencies (Figure 4.3C), recapitulating results observed for

the MEG recordings from humans following aministration of LSD (Figures 4.3A, 4.3B, 4.3D). This potentially points to an as-yet-unstudied molecular effect of psychedelics. It is already known that psychedelics induce a dramatic increase in the frequency and amplitude of spontaneous glutamatergic excitatory postsynaptic potentials in cortical layer V pyramidal cells [302–305], which leads to reduced broadband LFP [323] and MEG power [324], and that an increase in cortical excitability in mean-field models can recreate many of the effects of psychdedelics on macro-scale cortical dynamics [324, 325]. But, there is currently no published research on the effect of psychedelics on cortical gap junction coupling [326]. That said, in light of both our simulation-based results and prior empirical findings, we predict an inhibitory effect of psychedelics on gap junction coupling. To begin, it is already known that serotonin and other 5-HT2 agonists suppress gap junction coupling, and that 5-HT2 antagonists attenuate this effect [327]. Additionally, the antipsychotic drugs clozapine and haloperidol, which both diminish or block the effects of psychedelics [328–330], have been shown to increase the strength of gap junction coupling [331]. Finally, psychedelics are known to bear a number of striking resemblances to the antimalarial drug mefloquine, which, like psychedelics, is a potent agonist of the 5-HT2a receptor [326], reduces broadband EEG power [332], and can induce hallucinations and other psychiatric events [333]; importantly, mefloquine is also known to block several connexins [334–336] and is used to block gap junction coupling in experimental settings [337]. Thus, based on both our mean-field modeling results and prior empirical literature, we predict that psychedelics moderately block cortical gap junction coupling - a prediction that will need to be tested in future work.

Finally, we reiterate that the presence and degree of chaos in a system can only be determined with certainty in a simulation, by running the simulation multiple times with slightly different initial conditions and calculating the rate of divergence between different runs - see Methods. Thus, the only results here that are absolutely certain are those regarding the different degrees of chaos in the mean-field model we studied (Figure 4.1A). While our empirical results are consistent with the model's predictions, those predictions must themselves be treated with some caution in light of the shortcomings of the mean-field model we analyzed. To begin, the model lacks anatomical specificity: it treats the cortex as a continuum of locally connected tissue with sparse long-range myelinated connections, and lacks network properties like modularity that are characteristic of real brain networks, and which are known to influence network dynamics [338]. Second, the model treats the cortex as a two-dimensional sheet, when in reality the cortex and the white matter tracts that bridge cortical areas are extraordinarily complex three-dimensional structures. Third, the mean-field model ignores the effect of the thalamus on cortical dynamics, even though the thalamus is known to play a key (if not fully understood) role in cortical oscillations [339]. Thus, while our simulation and empirical results support the long-standing

proposal that waking cortical electrodynamics maintain a high information capacity by operating near edge-of-chaos criticality, the proposed framework will need to be revisited as our mean-field neural models become more biologically realistic, and as we develop more rigorous mappings between model parameters and real brain states.

4.5 Methods

4.5.1 Mean-Field Model Equations of Cortical Electrodynamics

The mean-field model of Steyn-Ross, Steyn-Ross, and Sleigh [276] is a more biologically detailed variant of the well-studied mean-field model first developed by Liley and colleagues [340, 341]. The model allows for straightforward manipulation of both the strength and balance of inhibitory and excitatory conductance, which have long been thought to be key in tuning neural dynamics to chaotic [342–347], critical [252, 253, 258, 267, 306–308, 344, 348, 349], and high-entropy [258, 345, 350] states. Furthermore, the model is unique in its inclusion of gap junction coupling between inhibitory interneurons, which recent empirical work has shown are also likely important for tuning neural dynamics toward and away from criticality [257].

The model simulates GABAergic anesthesia (e.g. propofol) as an increase in cortical inhibition coupled with a mild decrease in gap junction coupling between inhibitory interneurons, based on findings that GABAa agonists [351], and GABAergic anesthetics more specifically [352], inhibit gap junction communication [351, 352], and that these compounds increase postsynaptic inhibition by prolonging inhibitory postsynaptic potentials [353]. The model treats waking states as a balance between excitation and inhibition, with relatively strong gap junction coupling between inhibitory interneurons, which yields weak chaos (near edge-of-chaos criticality) in the model's deterministic component (Figure 4.1). Finally, the model has been shown to exhibit dynamics reminiscent of whole-of-cortex seizures when the coupling between its inhibitory neuron gap junctions is strongly reduced [276, 291]. This is consistent with observations of increased seizure frequency following either genetic ablation [354, 355] or drug-induced reduction [289, 355, 356] of gap junction coupling between inhibitory interneurons, though others note that drug-induced closing of gap junctions can decrease seizure frequency [354, 357–360] - a discrepancy that is likely due to differences in drug concentrations across studies [289, 358]. See Steyn-Ross, Sten-Ross, and Sleigh [276] for full details on model parameters.

The mean excitatory and inhibitory potentials V_e and V_i of each simulated macrocolumn in the mean-field model, positioned at a location $\overrightarrow{r} = (x, y)$, are described by:

$$\begin{aligned}\tau_b \frac{\delta V_b(\overrightarrow{r},t)}{\delta t} =& V_b^{\text{rest}} + \Delta V_b^{\text{rest}} - V_b(\overrightarrow{r},t) \\ &+ [\rho_e \psi_{eb}(\overrightarrow{r},t)\Phi_{eb}(\overrightarrow{r},t) + \rho_i \psi_{ib}(\overrightarrow{r},t)\Phi_{ib}(\overrightarrow{r},t)] \\ &+ D_{bb}\nabla^2 V_b(\overrightarrow{r},t)\end{aligned} \tag{4.1}$$

where presynaptic to postsynaptic directionality is indicated by the right arrow, the subscript e indicates a presynaptic excitatory neural population, the subscript i indicates a presynaptic inhibitory neural population, and the subscript b indicates either a postsynaptic excitatory or postsynaptic inhibitory neural population. The bracketed term in Eq. 1 represents voltage inputs via chemical synapses, and the final term in Eq. 1 represents voltage inputs from diffusive gap junction coupling. ∇^2 is the 2D Laplacian operator. D_{bb} represents the strength of diffusive gap junction coupling between adjacent neurons, such that D_{ee} is gap junction coupling between excitatory populations and D_{ii} is gap junction coupling between inhibitory populations. Because there is far more abundant gap-junction coupling between inhibitory interneurons than excitatory neurons [361], $D_{ee} = D_{ii}/100$. D_{ii} is one of the key biological parameters we vary. For a given excitatory or inhibitory neural population, V_b^{rest} is the mean resting potential, τ_b is the soma time constant, and ρ_b is the strength of chemical synapse coupling, which is scaled by the following reversal-potential function ψ_{ab}:

$$\psi_{ab}(\overrightarrow{r},t) = \frac{V_a^{\text{rev}} - V_b(\overrightarrow{r},t)}{V_a^{\text{rev}} - V_b^{\text{rest}}} \tag{4.2}$$

which equals one when a neuron is at its resting potential and equals zero when the membrane potential equals the reversal potential. For excitatory AMPA receptors, $V_e^{\text{rev}} = 0$ mV, and for inhibitory GABA receptors, $V_i^{\text{rev}} =$ -70 mV. The Φ_{ab} functions in Eq. 1 describe postsynaptic fluxes:

$$\begin{aligned}\left(\frac{\delta}{\delta t} + \gamma_e\right)^2 \Phi_{eb}(\overrightarrow{r},t) =& \gamma_e^2 [N_{eb}^{\alpha}\phi_{eb}^{\alpha}(\overrightarrow{r},t) + N_{eb}^{\beta} Q_e(\overrightarrow{r},t) \\ &+ \phi_{eb}^{\text{sc}}(\overrightarrow{r},t) + \phi_{eb}^{\alpha,\text{het}}(\overrightarrow{r},t)],\end{aligned} \tag{4.3}$$

$$\left(\frac{\delta}{\delta t} + \gamma_i\right)^2 \Phi_{ib}(\overrightarrow{r},t) = \gamma_i^2 N_{ib}^{\beta} Q_i(\overrightarrow{r},t) \tag{4.4}$$

where the α superscript corresponds to inputs from long-range myelinated axons: N_{eb}^{α} is the number of axonal inputs to a population and ϕ_{eb}^{α} is long-range flux. The β superscript corresponds to inputs from short-range chemical synapses, such that N_{eb}^{β} is the number of local

chemical synapses in a macrocolumn. $Q_{e,i}$ is the local spike-rate flux, and $\phi_{eb}^{\alpha,\text{het}}$ is a heterogeneous flux input. ϕ_{eb}^{sc} is white noise, taken to represent random inputs to the cortex from subcortical sources (e.g. sensory inputs); note that the inclusion of a noise term means that the above equations are stochastic differential equations, and that analyses of the chaoticity of the model pertain exclusively to the non-stochastic components of the model equations. γ_i is the inhibitory rate constant and γ_e is the excitatory rate constant, which we vary so as to the study the effect of excitation and inhibition on chaos in the model. See Steyn-Ross, Steyn-Ross, and Sleigh [276] for more details on the model equations. Other than the inhibitory gap-junction coupling strength D_{ii}, the excitatory rate constant γ_e, and the inhibitory rate constant γ_i (all of which we vary in our parameter sweep), all parameters in our simulations are unchanged from the original model, and are taken from the empirical literature [276]. D_{ii} was varied from 0.1 to 0.7 in steps of 0.2, and both γ_e and γ_i were varied from 0.945 to 1.05 in steps of 0.005. Of the 1,936 resulting simulations, 1,160 yielded suppressed, non-oscillatory activity. These were excluded from all analyses, leaving 776 model simulations with unique parameter configurations. The waking simulation corresponded to $\gamma_e = 1$, $\gamma_i = 1$, and $D_{ii} = 0.7$. The anesthesia simulation corresponded to $\gamma_e = 1$, $\gamma_i = 1.015$, and $D_{ii} = 0.5$, and the seizure simulation corresponded to $\gamma_e = 1$, $\gamma_i = 1$, and $D_{ii} = 0.1$. Critical State 1 in Figure 4.4 corresponded to $\gamma_e = 1.025$, $\gamma_i = 0.975$, and $D_{ii} = 0.3$, and critical State 2 in Figure 4.4 corresponded to $\gamma_e = 1.04$, $\gamma_i = 1$, and $D_{ii} = 0.5$. The model equations were integrated using a forward time center spaced first-order Euler method, with an integration step of 0.2 ms. Simulated electrodynamics were then downsampled to a sampling frequency of 500 Hz, and 10 seconds (i.e. 5,000 time-points) were extracted from the downsampled data, so as to perfectly match the length and sampling frequency of all empirical datasets analyzed.

4.5.2 Calculating largest Lyapunov exponents in the mean-field model

The ground-truth chaoticity of a system is determined by its largest Lyapunov exponent, which is the rate of divergence between initially similar trajectories in a system's phase space. A positive largest Lyapunov exponent means that a system is chaotic, because it indicates exponential divergence of initially similar system states. A negative largest Lyapunov exponent indicates periodicity, because it indicates exponentially fast convergence of initially similar states. A largest Lyapunov exponent near zero corresponds to edge-of-chaos criticality. The larger the largest Lyapunov exponent, the more strongly chaotic the system is. Following Destexhe [362] and Steyn-Ross, Steyn-Ross, and Sleigh [276], we estimate the largest Lyapunov exponent of the mean-field model by simulating two runs of its deterministic component, with slightly different initial conditions. The divergence between the excitatory firing rate of run 1 $Q_e^{(1)}$ and run 2

$Q_e^{(2)}$ is estimated as their summed squared-difference $\epsilon(t)$ down the midline of the simulated cortical grid:

$$\epsilon(t) = \sum_{i=1}^{N_x} (Q_e^{(1)}(x_i, t) - Q_e^{(2)}(x_i, t))^2 / \epsilon^{\max} \tag{4.5}$$

where $\epsilon^{\max}$ is a normalization parameter, which equals the maximum possible difference between the two runs:

$$\epsilon^{\max} = N_x \Big(\max(Q_e^{(1)}) - \min(Q_e^{(2)}) \Big)^2 \tag{4.6}$$

where N_x=120, i.e. the number of simulated macrocolumns. The rate of divergence between the two runs $\epsilon(t)$ is directly related to the largest Lyapunov exponent Λ of the system:

$$\epsilon(t) = \epsilon(0) \exp(\Lambda t) \tag{4.7}$$

where $\epsilon(0)$ is the distance between the two runs at $t = 0$. The largest Lyapunov exponent can therefore be estimated by measuring the slope of $\ln\epsilon(t)$-versus-t. A positive slope indicates a positive largest Lyapunov exponent, (and therefore chaotic dynamics), a negative slope indicates periodicity, and a flat slope indicates edge-of-chaos criticality. The largest Lyapunov exponent values in Figure 4.1a are the slopes of all model configurations that yielded ocsillatory (i.e. non-suppressed) dynamics.

4.5.3 Extracting low-frequency cortical activity

The mean-field model studied specifically simulates the low-frequency component of cortical electric oscillations. To compare the model results against real data, we therefore extracted the low-frequency component of both our simulated and real cortical signals. Although different frequencies of cortical electrodynamics have historically been studied at fixed, canonical frequency bands, with choices of oscillation center frequencies and bandwidths varying considerably across studies, there is growing evidence that these center frequencies and bandwidths vary considerably as a function of age [363, 364], brain state [365–367], subject [365], and species [368]. Given that our analyses span diverse brain states, species, and imaging modalities, it was important to identify subject- and channel-specific neural oscillation frequency ranges. We therefore identified low-frequency neural activity using the recently developed "Fitting Oscillations and One Over F" or "FOOOF" algorithm, which automatically parameterizes neural signals' power

spectra [369]. The algorithm fits a neural power spectrum as a linear combination of the background 1/f component with oscillations at specific frequencies that rise above this background 1/f component as peaks in the power spectrum. The algorithm fits the spectral power P as:

$$P = L + \sum_{n=0}^{N} G_N \tag{4.8}$$

where L is the background 1/f power spectrum, and each G_n is a Gaussian fit to a peak rising above the 1/f background:

$$G_n = a * exp(\frac{-(F-c)^2}{2w^2}) \tag{4.9}$$

where a is a given oscillation's amplitude, c is its center frequency, w is its bandwidth, and F is a vector of input frequencies. The 1/f background component L is modeled as an exponential in semilog-power space (i.e. with log power values as a function of linear frequencies):

$$L = b - log(k + F^{\chi}) \tag{4.10}$$

where b is a broadband power offset, χ is the spectral slope, k controls the "knee" at which the 1/f power spectrum bends [370], and F is a vector of input frequencies.

To specifically extract the low-frequency component of neural oscillations, we set the input frequency range F to 1-15 Hz. The FOOOF algorithm then identifies the center frequencies and bandwidths of putative oscillations within this frequency range. For all channels in our data, we extracted the lowest frequency component identified by the algorithm by low-pass filtering at the high-frequency end of the bandwidth of the slowest oscillation identified. By applying the modified 0-1 test for chaos (see below) to these extracted low-frequency components of our simulated mean-field dynamics, we confirmed that we could reliably detect changing degrees of chaos in the deterministic, low-frequency component of the model (Figure 4.1).

4.5.4 The modified 0-1 test for chaos

The 0-1 test was developed by Gottwald and Melbourne [295] as a simple tool for testing whether a deterministic, discrete-time system is chaotic, using only a single time-series recorded from that system. Gottwald and Melbourne provided an early modification to the test, which made it more robust against measurement noise [296]. Dawes and Freeland added additional modifications to the test, improving its ability to distinguish between chaotic dynamics and

strange non-chaotic or quasiperiodic dynamics [299]. Their modified 0-1 test takes a univariate time-series $\boldsymbol{\phi}$, and uses that time-series to drive the following two-dimensional system:

$$\begin{aligned} p(n+1) &= p(n) + \phi(n)\cos cn \\ q(n+1) &= q(n) + \phi(n)\sin cn \end{aligned} \tag{4.11}$$

where c is a random value bounded between 0 and 2π. For a given c, the solution to Eq. 1 yields:

$$\begin{aligned} p_c(n) &= \sum_{j=1}^{n} \phi(j)\cos jc \\ q_c(n) &= \sum_{j=1}^{n} \phi(j)\sin jc \end{aligned} \tag{4.12}$$

If the time-series $\boldsymbol{\phi}$ is periodic, the motion of $\mathbf{p}$ and $\mathbf{q}$ is bounded, while if the time-series $\boldsymbol{\phi}$ is chaotic, $\mathbf{p}$ and $\mathbf{q}$ display asymptotic Brownian motion. The time-averaged mean square displacement of $\mathbf{p}$ and $\mathbf{q}$ is

$$M_c(n) = \frac{1}{N}\sum_{j=1}^{N}([p_c(j+n) - p_c(j)]^2 + [q_c(j+n) - q_c(j)]^2) + \sigma\eta_n. \tag{4.13}$$

where η_n is a uniformly distributed random variable between $[-\frac{1}{2}, \frac{1}{2}]$ and σ controls the amplitude of the added random variable η_n. We set σ to 0.5, based on our previously published analyses of the effect of different parameter values for 0-1 test performance across diverse datasets. To compute degree of chaos using a single statistic K, the 0-1 test calculates the growth rate of the mean squared displacement of the two-dimensional system in Eq. 5 using a correlation coefficient:

$$K_c = \mathrm{corr}(n, M_c(n)) \tag{4.14}$$

K is computed for 100 different values of c, randomly sampled between 0 and 2π, and the output of the test is the median K across different values of c. As the length of a time-series is increased, this median K value will approach 1 for chaotic systems, and will approach 0 for periodic systems, and will track degree of chaos for finite-length time-series [295–299].

The 0-1 test is designed to detect and track chaos in deterministic, discrete-time systems. Thus, a recorded signal must first be tested for stochasticity [371], and, if deemed to reflect a predominantly deterministic process, the signal must be downsampled if oversampled from a continuous system [297]. In our previous work, we have shown that the possibility of stochasticity can be ruled out with extremely high accuracy by comparing the permutation entropy of a signal

to the permutation entropy of 1,000 Amplitude Adjusted Fourier Transform surrogates and 1,000 Cyclic Phase Permutation surrogates. If the permutation entropy of the test signal falls within either distribution of surrogate permutation entropies, the test signal very likely reflects a stochastic process (either linear or nonlinear stochsatic). Very few of the signals tested here were classified as stochastic by this relatively strict test (Tables 1-2). In terms of transforming continuous signals into discrete-time signals, two approaches have been proposed: downsampling [297], or taking all local minima and maxima of a continuous signal [300]. While there are no theoretically rigorous recommendations regarding how much to downsample a signal if this first approach is chosen, or even how to determine if a continuous signal is oversampled, it has been shown that iteratively downsampling a signal until it passes a crude test of oversampling - namely, until the difference between the global maximum and global minimum of a signal divided by the mean absolute difference between consecutive time-points in the data is less than 10 - can significantly improve 0-1 test performance in oversampled data. We found that the second approach (namely, applying the modified 0-1 test to the series of local minima and maxima of a test signal) tracked chaoticity in the low-frequency component of the mean-field model dynamics more accurately (Spearman correlation between K and ground-truth largest Lyapunov exponents was 0.84, $p< 10^{-205}$) than did the downsampling approach (Spearman correlation between K and ground-truth largest Lyapunov exponents was 0.71, $p< 10^{-120}$), and so we used the local minima and maxima approach on our real cortical recordings.

4.5.5 Epilepsy data

Surface ECoG data from nine epilepsy patients were downloaded from the European Epilepsy Database [372]. Of these, only two subjects experienced fully generalized seizures (in both cases, seizures were focal with secondary generalizations). Signals were recorded at a sampling rate of 1024 Hz. Data were demeaned, detrended, and bandstop filtered at 50 Hz and harmonics (the line noise frequency in Europe). Data were resampled to 500 Hz and divided into 10-second trials. For the seizure state, we only included trials for which seizures were fully generalized for the entire trial duration. The data were then visually inspected for artifacts. Data from electrodes with consistent artifacts were removed, and 10-second trials with artifacts spanning multiple electrodes were removed.

4.5.6 Anesthesia data

Open-source ECoG recordings spanning the left cortices of two macaques were downloaded from Neurotycho.org [373]. Data were collected during resting, propofol anesthesia, and recovery states. Macaques were seated with head and arm movement restricted. Macaque 1 was intravenously administered 5.2 mg/kg of propofol, and Macaque 2 was intravenously administered 5 mg/kg of propofol. Loss of consciousness was determined by the emergence of slow wave oscillations and the cessation of responses to physical stimuli. All data for the propofol condition are from the macaques' unconscious state. Signals were recorded at a sampling rate of 1,000 Hz. Data were split into 10-second trials, demeaned, band-stop filtered at 50 Hz and harmonics (the line noise frequency in Japan, where data were collected), and detrended. Data were then visually inspected for artifacts. Data from electrodes with consistent artifacts were removed, and 10-second trials with artifacts spanning multiple electrodes were removed. Data were then re-referenced to the common average, and downsampled to 500 Hz.

4.5.7 LSD data

Previously published [301] MEG recordings of four humans following intravenous administration of either 2 mg of LSD or a saline placebo were re-analyzed. Exclusion criteria are described in Carhart-Harris et al [301]. Due to the slow pharmacodynamics of LSD, data were recorded four hours after drug administration. Subjects lay in a supine position during data acquisition. MEG signals were recorded using a CTF 275-channel radial gradiometer system with a sampling frequency of 1200 Hz. Data were split into 10-second trials, demeaned, detrended, and bandstop filtered at 50 Hz and harmonics (the line noise frequency in the United Kingdom, where data were collected). Data were then visually inspected, and consistently artifact-affected channels were removed, and trials with significant artifacts across channels were removed. Data were then downsampled to 500 Hz. We then ran an independent components analysis on the data, and removed components that corresponded to ocular, cardiac, or muscle artifacts.

TABLE 4.1: The number of channels, across all 10-second trials in every condition, for which low-frequency activity (as identified by the FOOOF algorithm) was classified as stochastic. A channel was classified as stochastic if its permutation entropy fell within the distribution of permutation entropies calculated from 1,000 Amplitude Adjusted Fourier Transform surrogate time-series *or* from 1,000 Cyclic Phase Permutation surrogate time-series. This relatively strict test has been shown to reliably detect whether a signal is generated by a predominantly stochastic process (either linear stochastic or nonlinear stochastic). Relatively few channels were classified as stochastic for any condition, species, or imaging modality, which is consistent with the prediction that low-frequency cortical electrodynamics are predominantly deterministic, with some low levels of intrinsic noise.

Macaque ECoG: Awake, Eyes Open	2,595/ 26,246
Macaque ECoG: Propofol Anesthesia	25/ 28,030
Human MEG: Placebo, Eyes Open	0/ 61,697
Human MEG: LSD, Eyes Open	0/ 62,530
Human MEG: Epileptic, Waking Baseline	13/3,684
Human MEG: Epileptic, Generalized Seizure	0/2,092

TABLE 4.2: The number of channels, across all 10-second trials in every condition, for which activity under 30 Hz was classified as stochastic. A channel was classified as stochastic if its permutation entropy fell within the distribution of permutation entropies calculated from 1,000 Amplitude Adjusted Fourier Transform surrogate time-series *or* from 1,000 Cyclic Phase Permutation surrogate time-series. As was the case for low-frequency activity, relatively few channels were classified as stochastic for any condition, species, or imaging modality, consistent with the prediction that cortical electrodynamics are predominantly deterministic, with some low levels of intrinsic noise.

Macaque ECoG: Awake, Eyes Open	2,091/ 26,246
Macaque ECoG: Propofol Anesthesia	12/ 28,030
Human MEG: Placebo, Eyes Open	0/ 61,697
Human MEG: LSD, Eyes Open	0/ 62,530
Human MEG: Epileptic, Waking Baseline	8/3,684
Human MEG: Epileptic, Generalized Seizure	0/2,092

Bibliography

[1] Sporns, O. *Networks of the Brain* (MIT press, 2010).

[2] Buzsaki, G. *Rhythms of the Brain* (Oxford University Press, 2006).

[3] Bertolero, M. A., Yeo, B. T. & D'Esposito, M. The modular and integrative functional architecture of the human brain. *Proceedings of the National Academy of Sciences* **112**, E6798–E6807 (2015).

[4] Minshew, N. J., Goldstein, G. & Siegel, D. J. Neuropsychologic functioning in autism: Profile of a complex information processing disorder. *Journal of the International Neuropsychological Society* **3**, 303–316 (1997).

[5] Koch, C., Massimini, M., Boly, M. & Tononi, G. Neural correlates of consciousness: progress and problems. *Nature Reviews Neuroscience* **17**, 307 (2016).

[6] Churchland, P. S. & Sejnowski, T. J. *The computational brain* (MIT press, 2016).

[7] Shannon, C. E. A mathematical theory of communication. *Bell system technical journal* **27**, 379–423 (1948).

[8] Gratton, C., Nomura, E. M., Pérez, F. & D'Esposito, M. Focal brain lesions to critical locations cause widespread disruption of the modular organization of the brain. *Journal of cognitive neuroscience* **24**, 1275–1285 (2012).

[9] Alkire, M. T., Hudetz, A. G. & Tononi, G. Consciousness and anesthesia. *Science* **322**, 876–880 (2008).

[10] Tononi, G. & Massimini, M. Why does consciousness fade in early sleep? *Annals of the New York Academy of Sciences* **1129**, 330–334 (2008).

[11] Blumenfeld, H. & Taylor, J. Why do seizures cause loss of consciousness? *The Neuroscientist* **9**, 301–310 (2003).

[12] Rudie, J. D. *et al.* Altered functional and structural brain network organization in autism. *NeuroImage: clinical* **2**, 79–94 (2013).

[13] Coben, R., Clarke, A. R., Hudspeth, W. & Barry, R. J. Eeg power and coherence in autistic spectrum disorder. *Clinical neurophysiology* **119**, 1002–1009 (2008).

[14] Achard, S. *et al.* Hubs of brain functional networks are radically reorganized in comatose patients. *Proceedings of the National Academy of Sciences* **109**, 20608–20613 (2012).

[15] Kaada, B., Harkmark, W. & Stokke, O. Deep coma associated with desynchronization in eeg. *Electroencephalography and Clinical Neurophysiology* **13**, 785–789 (1961).

[16] Hartikainen, P., Soininen, H., Partanen, J., Helkala, E. & Riekkinen, P. Aging and spectral analysis of eeg in normal subjects: a link to memory and csf ache. *Acta Neurologica Scandinavica* **86**, 148–155 (1992).

[17] Wu, K. *et al.* A longitudinal study of structural brain network changes with normal aging. *Frontiers in human neuroscience* **7**, 113 (2013).

[18] Oizumi, M., Tsuchiya, N. & Amari, S.-i. Unified framework for information integration based on information geometry. *Proceedings of the National Academy of Sciences* **113**, 14817–14822 (2016).

[19] Skarda, C. A. & Freeman, W. J. How brains make chaos in order to make sense of the world. *Behavioral and brain sciences* **10**, 161–173 (1987).

[20] Korn, H. & Faure, P. Is there chaos in the brain? ii. experimental evidence and related models. *Comptes rendus biologies* **326**, 787–840 (2003).

[21] Destexhe, A. Oscillations, complex spatiotemporal behavior, and information transport in networks of excitatory and inhibitory neurons. *Physical Review E* **50**, 1594 (1994).

[22] Nagao, N., Nishimura, H. & Matsui, N. A neural chaos model of multistable perception. *Neural Processing Letters* **12**, 267–276 (2000).

[23] Rabinovich, M. & Abarbanel, H. The role of chaos in neural systems. *Neuroscience* **87**, 5–14 (1998).

[24] Alexander, D. M. & Globus, G. G. Edge-of-chaos dynamics in recursively organized neural systems. *Fractals of brain, fractals of mind* 31–73 (1996).

[25] Bertschinger, N. & Natschläger, T. Real-time computation at the edge of chaos in recurrent neural networks. *Neural computation* **16**, 1413–1436 (2004).

[26] Legenstein, R. & Maass, W. Edge of chaos and prediction of computational performance for neural circuit models. *Neural Networks* **20**, 323–334 (2007).

[27] Kadmon, J. & Sompolinsky, H. Transition to chaos in random neuronal networks. *Physical Review X* **5**, 041030 (2015).

[28] Aljadeff, J., Stern, M. & Sharpee, T. Transition to chaos in random networks with cell-type-specific connectivity. *Physical review letters* **114**, 088101 (2015).

[29] Kanders, K., Lorimer, T. & Stoop, R. Avalanche and edge-of-chaos criticality do not necessarily co-occur in neural networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **27**, 047408 (2017).

[30] Shannon CE. A mathematical theory of communication. ACM SIGMOBILE Mobile Computing and Communications Review. 2001;5(1):3–55.

[31] Borst A, Theunissen FE. Information theory and neural coding. Nature neuroscience. 1999;2(11):947–957.

[32] Laughlin S. A simple coding procedure enhances a neuron's information capacity. Zeitschrift für Naturforschung c. 1981;36(9-10):910–912.

[33] Lewicki MS. Efficient coding of natural sounds. Nature neuroscience. 2002;5(4):356–363.

[34] Olshausen BA, Field DJ. Sparse coding with an overcomplete basis set: A strategy employed by V1? Vision research. 1997;37(23):3311–3325.

[35] Brenner N, Strong SP, Koberle R, Bialek W, Van Steveninck RRDR. Synergy in a neural code. Neural computation. 2000;12(7):1531–1552.

[36] Dayan P, Abbott LF. Theoretical neuroscience. vol. 806. Cambridge, MA: MIT Press; 2001.

[37] Barlow HB. Possible principles underlying the transformations of sensory messages. 1961;.

[38] Ay N, et al. Information geometry on complexity and stochastic interaction. 2001;.

[39] Ay N. Information geometry on complexity and stochastic interaction. Entropy. 2015;17(4):2432–2458.

[40] Tegmark M. Improved measures of integrated information. PLOS Computational Biology. 2016;12(11):e1005123.

[41] Seth AK, Izhikevich E, Reeke GN, Edelman GM. Theories and measures of consciousness: an extended framework. Proceedings of the National Academy of Sciences. 2006;103(28):10799–10804.

[42] Oizumi M, Albantakis L, Tononi G. From the phenomenology to the mechanisms of consciousness: integrated information theory 3.0. PLoS Comput Biol. 2014;10(5):e1003588.

[43] Oizumi M, Amari Si, Yanagawa T, Fujii N, Tsuchiya N. Measuring integrated information from the decoding perspective. PLoS Comput Biol. 2016;12(1):e1004654.

[44] Barrett AB, Seth AK. Practical measures of integrated information for time-series data. PLoS Comput Biol. 2011;7(1):e1001052.

[45] Seth AK, Barrett AB, Barnett L. Causal density and integrated information as measures of conscious level. Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences. 2011;369(1952):3748–3767.

[46] Oizumi M, Tsuchiya N, Amari Si. Unified framework for information integration based on information geometry. Proceedings of the National Academy of Sciences. 2016;113(51):14817–14822.

[47] Amari Si. Information geometry and its applications. vol. 194. Springer; 2016.

[48] Balduzzi D, Tononi G. Integrated information in discrete dynamical systems: motivation and theoretical framework. PLoS Comput Biol. 2008;4(6):e1000091.

[49] Toker D, Sommer F. Moving Past the Minimum Information Partition: How To Quickly and Accurately Calculate Integrated Information. arXiv preprint arXiv:160501096. 2016;.

[50] Kitazono J, Kanai R, Oizumi M. Efficient Algorithms for Searching the Minimum Information Partition in Integrated Information Theory. Entropy. 2018;20(3):173.

[51] Sporns O, Betzel RF. Modular brain networks. Annual review of psychology. 2016;67:613–640.

[52] Shi J, Malik J. Normalized cuts and image segmentation. IEEE Transactions on pattern analysis and machine intelligence. 2000;22(8):888–905.

[53] Ng AY, Jordan MI, Weiss Y. On spectral clustering: Analysis and an algorithm. In: Advances in neural information processing systems; 2002. p. 849–856.

[54] Aicher C, Jacobs AZ, Clauset A. Adapting the stochastic block model to edge-weighted networks. arXiv preprint arXiv:13055782. 2013;.

[55] Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. Journal of statistical mechanics: theory and experiment. 2008;2008(10):P10008.

[56] Kitazono J, Oizumi M. phi_toolbox.zip; 2017. Available from: https://figshare.com/articles/phi_toolbox_zip/3203326/6.

[57] Hoel EP, Albantakis L, Marshall W, Tononi G. Can the macro beat the micro? Integrated information across spatiotemporal scales. Neuroscience of Consciousness. 2016;2016(1):niw012.

[58] Wibral M, Pampu N, Priesemann V, Siebenhühner F, Seiwert H, Lindner M, et al. Measuring information-transfer delays. PloS one. 2013;8(2):e55809.

[59] Sporns O, Tononi G, Edelman GM. Connectivity and complexity: the relationship between neuroanatomy and brain dynamics. Neural networks. 2000;13(8):909–922.

[60] Sporns O. Networks of the Brain. MIT press; 2010.

[61] Bassett DS, Bullmore E. Small-world brain networks. The neuroscientist. 2006;12(6):512–523.

[62] Bertolero MA, Yeo BT, D'Esposito M. The modular and integrative functional architecture of the human brain. Proceedings of the National Academy of Sciences. 2015;112(49):E6798–E6807.

[63] Bertolero M, Yeo B, D'Esposito M. The diverse club. Nature communications. 2017;8(1):1277.

[64] Craddock RC, James GA, Holtzheimer PE, Hu XP, Mayberg HS. A whole brain fMRI atlas generated via spatially constrained spectral clustering. Human brain mapping. 2012;33(8):1914–1928.

[65] Schwarz AJ, McGonigle J. Negative edges and soft thresholding in complex network analysis of resting state functional connectivity data. Neuroimage. 2011;55(3):1132–1146.

[66] Achard S, Bullmore E. Efficiency and cost of economical brain functional networks. PLoS computational biology. 2007;3(2):e17.

[67] Van Den Heuvel M, Mandl R, Pol HH. Normalized cut group clustering of resting-state FMRI data. PloS one. 2008;3(4):e2001.

[68] van den Heuvel MP, Stam CJ, Boersma M, Pol HH. Small-world and scale-free organization of voxel-based resting-state functional connectivity in the human brain. Neuroimage. 2008;43(3):528–539.

[69] Nakamura T, Hillary FG, Biswal BB. Resting network plasticity following brain injury. PloS one. 2009;4(12):e8220.

[70] Hayasaka S, Laurienti PJ. Comparison of characteristics between region-and voxel-based network analyses in resting-state fMRI data. Neuroimage. 2010;50(2):499–508.

[71] Rössler OE. Continuous chaos—four prototype equations. Annals of the New York Academy of Sciences. 1979;316(1):376–392.

[72] Aydore S, Pantazis D, Leahy RM. Phase synchrony in multivariate Gaussian data with applications to cortical networks. In: Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on. IEEE; 2012. p. 1547–1550.

[73] Buzsáki G, Mizuseki K. The log-dynamic brain: how skewed distributions affect network operations. Nature Reviews Neuroscience. 2014;15(4):264–278.

[74] Rand WM. Objective criteria for the evaluation of clustering methods. Journal of the American Statistical association. 1971;66(336):846–850.

[75] Nagasaka Y, Shimoda K, Fujii N. Multidimensional recording (MDR) and data sharing: an ecological open research and educational platform for neuroscience. PloS one. 2011;6(7):e22561.

[76] Theiler J, Eubank S, Longtin A, Galdrikian B, Farmer JD. Testing for nonlinearity in time series: the method of surrogate data. Physica D: Nonlinear Phenomena. 1992;58(1-4):77–94.

[77] Bullmore E, Sporns O. Complex brain networks: graph theoretical analysis of structural and functional systems. Nature Reviews Neuroscience. 2009;10(3):186–198.

[78] Latora V, Marchiori M. Efficient behavior of small-world networks. Physical review letters. 2001;87(19):198701.

[79] Avena-Koenigsberger A, Misic B, Sporns O. Communication dynamics in complex brain networks. Nature Reviews Neuroscience. 2018;19(1):17.

[80] Rubinov M, Sporns O. Complex network measures of brain connectivity: uses and interpretations. Neuroimage. 2010;52(3):1059–1069.

[81] Newman ME. Modularity and community structure in networks. Proceedings of the national academy of sciences. 2006;103(23):8577–8582.

[82] Watts DJ, Strogatz SH. Collective dynamics of 'small-world'networks. nature. 1998;393(6684):440–442.

[83] Latora V, Marchiori M. Economic small-world behavior in weighted networks. The European Physical Journal B-Condensed Matter and Complex Systems. 2003;32(2):249–263.

[84] Tononi G. Consciousness as integrated information: a provisional manifesto. The Biological Bulletin. 2008;215(3):216–242.

[85] Tononi G. Consciousness, information integration, and the brain. Progress in brain research. 2005;150:109–126.

[86] Tononi G, Sporns O. Measuring information integration. BMC neuroscience. 2003;4(1):31.

[87] Holcombe AO. Seeing slow and seeing fast: two limits on perception. Trends in cognitive sciences. 2009;13(5):216–221.

[88] Bachmann T. Microgenetic approach to the conscious mind. vol. 25. John Benjamins Publishing; 2000.

[89] Finlay B, Brodsky P. Cortical Evolution as the Expression of a Program for Disproportionate Growth and the Proliferation of Areas-3.05. 2006;.

[90] Schelter B, Winterhalder M, Dahlhaus R, Kurths J, Timmer J. Partial phase synchronization for multivariate synchronizing systems. Physical review letters. 2006;96(20):208103.

[91] Huang L, Chen Q, Lai YC, Pecora LM. Generic behavior of master-stability functions in coupled nonlinear dynamical systems. Physical Review E. 2009;80(3):036204.

[92] Buscarino A, Fortuna L, Frasca M. Essentials of Nonlinear Circuit Dynamics with MATLAB® and Laboratory Experiments. CRC Press; 2017.

[93] Lancaster, G., Iatsenko, D., Pidde, A., Ticcinelli, V. & Stefanovska, A. Surrogate data for hypothesis testing of physical systems. *Physics Reports* (2018).

[94] Timmer, J. Power of surrogate data testing with respect to nonstationarity. *Physical Review E* **58**, 5153 (1998).

[95] Patterson, K. *Unit Root Tests in Time Series Volume 1: Key Concepts and Problems* (Springer, 2011).

[96] Patterson, K. *Unit Root Tests in Time Series Volume 2: Extensions and Developments*, vol. 2 (Palgrave Macmillan, 2012).

[97] Dickey, D. A. & Fuller, W. A. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association* **74**, 427–431 (1979).

[98] Kwiatkowski, D., Phillips, P. C., Schmidt, P. & Shin, Y. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of econometrics* **54**, 159–178 (1992).

[99] Leybourne, S. J. & McCabe, B. P. A consistent test for a unit root. *Journal of Business & Economic Statistics* **12**, 157–166 (1994).

[100] Lo, A. W. & MacKinlay, A. C. Stock market prices do not follow random walks: Evidence from a simple specification test. *The review of financial studies* **1**, 41–66 (1988).

[101] Lo, A. W. & MacKinlay, A. C. The size and power of the variance ratio test in finite samples: A monte carlo investigation. *Journal of econometrics* **40**, 203–238 (1989).

[102] Breitung, J. Nonparametric tests for unit roots and cointegration. *Journal of econometrics* **108**, 343–363 (2002).

[103] Chan, K.-S. & Tong, H. *Chaos: a statistical perspective* (Springer Science & Business Media, 2013).

[104] Theiler, J., Eubank, S., Longtin, A., Galdrikian, B. & Farmer, J. D. Testing for nonlinearity in time series: the method of surrogate data. *Physica D: Nonlinear Phenomena* **58**, 77–94 (1992).

[105] Theiler, J. On the evidence for low-dimensional chaos in an epileptic electroencephalogram. *Physics Letters A* **196**, 335–341 (1994).

[106] Jamšek, J., Paluš, M. & Stefanovska, A. Detecting couplings between interacting oscillators with time-varying basic frequencies: Instantaneous wavelet bispectrum and information theoretic approach. *Physical Review E* **81**, 036207 (2010).

[107] Timmer, J., Häußler, S., Lauk, M. & Lücking, C.-H. Pathological tremors: Deterministic chaos or nonlinear stochastic oscillators? *Chaos: An Interdisciplinary Journal of Nonlinear Science* **10**, 278–288 (2000).

[108] Schreiber, T. Extremely simple nonlinear noise-reduction method. *Physical Review E* **47**, 2401 (1993).

[109] Matthews, P. 0 - 1 test for chaos. https://tinyurl.com/qruo8y2. (2009).

[110] Eyébé Fouda, J. S. A., Bodo, B., Sabat, S. L. & Effa, J. Y. A modified 0-1 test for chaos detection in oversampled time series observations. *International Journal of Bifurcation and Chaos* **24**, 1450063 (2014).

[111] Huebner, U., Abraham, N. & Weiss, C. Dimensions and entropies of chaotic intensity pulsations in a single-mode far-infrared nh 3 laser. *Physical Review A* **40**, 6354 (1989).

[112] Lindner, J. F. *et al.* Strange nonchaotic stars. *Physical review letters* **114**, 054101 (2015).

[113] BenSaïda, A. A practical test for noisy chaotic dynamics. *SoftwareX* **3**, 1–5 (2015).

[114] Rosenstein, M. T., Collins, J. J. & De Luca, C. J. A practical method for calculating largest Lyapunov exponents from small data sets. *Physica D: Nonlinear Phenomena* **65**, 117–134 (1993).

[115] Eckmann, J.-P., Kamphorst, S. O., Ruelle, D. & Ciliberto, S. Liapunov exponents from time series. *Physical Review A* **34**, 4971 (1986).

[116] Amigó, J. M. The equality of Kolmogorov–Sinai entropy and metric permutation entropy generalized. *Physica D: Nonlinear Phenomena* **241**, 789–793 (2012).

[117] White, J. A., Rubinstein, J. T. & Kay, A. R. Channel noise in neurons. *Trends in neurosciences* **23**, 131–137 (2000).

[118] Dawes, J. & Freeland, M. The '0–1 test for chaos' and strange nonchaotic attractors. *preprint* (2008).

[119] Kulp, C. & Zunino, L. Discriminating chaotic and stochastic dynamics through the permutation spectrum test. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **24**, 033116 (2014).

[120] Hu, J., Tung, W.-w., Gao, J. & Cao, Y. Reliability of the 0-1 test for chaos. *Physical Review E* **72**, 056207 (2005).

[121] Sauer, T., Yorke, J. A. & Casdagli, M. Embedology. *Journal of statistical Physics* **65**, 579–616 (1991).

[122] Zunino, L. & Kulp, C. W. Detecting nonlinearity in short and noisy time series using the permutation entropy. *Physics Letters A* **381**, 3627–3635 (2017).

[123] Gottwald, G. A. & Melbourne, I. A new test for chaos in deterministic systems. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 460, 603–611 (The Royal Society, 2004).

[124] Gottwald, G. A. & Melbourne, I. Testing for chaos in deterministic systems with noise. *Physica D: Nonlinear Phenomena* **212**, 100–110 (2005).

[125] Gottwald, G. A. & Melbourne, I. On the implementation of the 0–1 test for chaos. *SIAM Journal on Applied Dynamical Systems* **8**, 129–145 (2009).

[126] Gottwald, G. A. & Melbourne, I. Comment on "reliability of the 0-1 test for chaos". *Physical Review E* **77**, 028201 (2008).

[127] Bernardini, D. & Litak, G. An overview of 0–1 test for chaos. *Journal of the Brazilian Society of Mechanical Sciences and Engineering* **38**, 1433–1450 (2016).

[128] Bandt, C. & Pompe, B. Permutation entropy: a natural complexity measure for time series. *Physical review letters* **88**, 174102 (2002).

[129] Trostel, M. L., Misplon, M. Z., Aragoneses, A. & Pattanayak, A. K. Characterizing complex dynamics in the classical and semi-classical duffing oscillator using ordinal patterns analysis. *Entropy* **20**, 40 (2018).

[130] Politi, A. Quantifying the dynamical complexity of chaotic time series. *Physical review letters* **118**, 144101 (2017).

[131] Bandt, C., Keller, G. & Pompe, B. Entropy of interval maps via permutations. *Nonlinearity* **15**, 1595 (2002).

[132] Keller, K., Unakafov, A. M. & Unakafova, V. A. On the relation of KS entropy and permutation entropy. *Physica D: Nonlinear Phenomena* **241**, 1477–1481 (2012).

[133] Pesin, Y. B. Characteristic Lyapunov exponents and smooth ergodic theory. *Russian Mathematical Surveys* **32**, 55–114 (1977).

[134] Linde, A. D. Eternally existing self-reproducing chaotic inflanationary universe. *Physics Letters B* **175**, 395–400 (1986).

[135] Wisdom, J., Peale, S. J. & Mignard, F. The chaotic rotation of hyperion. *Icarus* **58**, 137–152 (1984).

[136] Hudson, J. & Mankin, J. Chaos in the belousov–zhabotinskii reaction. *The Journal of Chemical Physics* **74**, 6171–6177 (1981).

[137] Kaplan, D. T. *et al.* Subthreshold dynamics in periodically stimulated squid giant axons. *Physical Review Letters* **76**, 4074 (1996).

[138] Glass, L. Introduction to controversial topics in nonlinear science: Is the normal heart rate chaotic? (2009).

[139] Goldberger, A. L. & West, B. J. Applications of nonlinear dynamics to clinical cardiology. *Annals of the New York Academy of Sciences* **504**, 195–213 (1987).

[140] Skarda, C. A. & Freeman, W. J. How brains make chaos in order to make sense of the world. *Behavioral and brain sciences* **10**, 161–173 (1987).

[141] Pool, R. Is it healthy to be chaotic? *Science* **243**, 604–607 (1989).

[142] Korn, H. & Faure, P. Is there chaos in the brain? ii. experimental evidence and related models. *Comptes rendus biologies* **326**, 787–840 (2003).

[143] Destexhe, A. Oscillations, complex spatiotemporal behavior, and information transport in networks of excitatory and inhibitory neurons. *Physical Review E* **50**, 1594 (1994).

[144] Lizier, J. T., Prokopenko, M. & Zomaya, A. Y. The information dynamics of phase transitions in random boolean networks. In *ALIFE*, 374–381 (2008).

[145] Nagao, N., Nishimura, H. & Matsui, N. A neural chaos model of multistable perception. *Neural Processing Letters* **12**, 267–276 (2000).

[146] Rabinovich, M. & Abarbanel, H. The role of chaos in neural systems. *Neuroscience* **87**, 5–14 (1998).

[147] Heltberg, M. L., Krishna, S. & Jensen, M. H. On chaotic dynamics in transcription factors and the associated effects in differential gene regulation. *Nature communications* **10**, 71 (2019).

[148] Glass, L. & Mackey, M. C. *From clocks to chaos: The rhythms of life* (Princeton University Press, 1988).

[149] Guevara, M. R., Glass, L. & Shrier, A. Phase locking, period-doubling bifurcations, and irregular dynamics in periodically stimulated cardiac cells. *Science* **214**, 1350–1353 (1981).

[150] Hayashi, H. & Ishizuka, S. Chaotic nature of bursting discharges in the onchidium pacemaker neuron. *Journal of Theoretical Biology* **156**, 269–291 (1992).

[151] Hayashi, H., Nakao, M. & Hirakawa, K. Chaos in the self-sustained oscillation of an excitable biological membrane under sinusoidal stimulation. *Physics Letters A* **88**, 265–266 (1982).

[152] Mackey, M. C. & Glass, L. Oscillation and chaos in physiological control systems. *Science* **197**, 287–289 (1977).

[153] Steyn-Ross, M. L., Steyn-Ross, D. A. & Sleigh, J. W. Interacting Turing-Hopf instabilities drive symmetry-breaking transitions in a mean-field model of the cortex: a mechanism for the slow oscillation. *Physical Review X* **3**, 021005 (2013).

[154] Chan, K.-S. & Tong, H. *Chaos: a statistical perspective* (Springer Science & Business Media, 2013).

[155] Dawes, J. & Freeland, M. The '0–1 test for chaos' and strange nonchaotic attractors. *preprint* (2008).

[156] Hu, J., Tung, W.-w., Gao, J. & Cao, Y. Reliability of the 0-1 test for chaos. *Physical Review E* **72**, 056207 (2005).

[157] Kulp, C. & Zunino, L. Discriminating chaotic and stochastic dynamics through the permutation spectrum test. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **24**, 033116 (2014).

[158] Timmer, J., Häußler, S., Lauk, M. & Lücking, C.-H. Pathological tremors: Deterministic chaos or nonlinear stochastic oscillators? *Chaos: An Interdisciplinary Journal of Nonlinear Science* **10**, 278–288 (2000).

[159] Uenohara, S. *et al.* Experimental distinction between chaotic and strange nonchaotic attractors on the basis of consistency. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **23**, 023110 (2013).

[160] Huebner, U., Abraham, N. & Weiss, C. Dimensions and entropies of chaotic intensity pulsations in a single-mode far-infrared nh 3 laser. *Physical Review A* **40**, 6354 (1989).

[161] Lindner, J. F. *et al.* Strange nonchaotic stars. *Physical review letters* **114**, 054101 (2015).

[162] Hurrell, J. W., Kushnir, Y., Ottersen, G. & Visbeck, M. An Overview of the North Atlantic Oscillation. *The North Atlantic Oscillation: climatic significance and environmental impact* **134**, 1–35 (2003).

[163] Goldberger, A. L. *et al.* Physiobank, Physiotoolkit, and Physionet: components of a new research resource for complex physiologic signals. *Circulation* **101**, e215–e220 (2000).

[164] Bandt, C. & Pompe, B. Permutation entropy: a natural complexity measure for time series. *Physical review letters* **88**, 174102 (2002).

[165] Theiler, J., Eubank, S., Longtin, A., Galdrikian, B. & Farmer, J. D. Testing for nonlinearity in time series: the method of surrogate data. *Physica D: Nonlinear Phenomena* **58**, 77–94 (1992).

[166] Lancaster, G., Iatsenko, D., Pidde, A., Ticcinelli, V. & Stefanovska, A. Surrogate data for hypothesis testing of physical systems. *Physics Reports* (2018).

[167] Jamšek, J., Paluš, M. & Stefanovska, A. Detecting couplings between interacting oscillators with time-varying basic frequencies: Instantaneous wavelet bispectrum and information theoretic approach. *Physical Review E* **81**, 036207 (2010).

[168] Schreiber, T. Extremely simple nonlinear noise-reduction method. *Physical Review E* **47**, 2401 (1993).

[169] Gottwald, G. A. & Melbourne, I. A new test for chaos in deterministic systems. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 460, 603–611 (The Royal Society, 2004).

[170] Gottwald, G. A. & Melbourne, I. Testing for chaos in deterministic systems with noise. *Physica D: Nonlinear Phenomena* **212**, 100–110 (2005).

[171] Gottwald, G. A. & Melbourne, I. On the implementation of the 0–1 test for chaos. *SIAM Journal on Applied Dynamical Systems* **8**, 129–145 (2009).

[172] Gottwald, G. A. & Melbourne, I. Comment on "reliability of the 0-1 test for chaos". *Physical Review E* **77**, 028201 (2008).

[173] Guevara, M. R. & Glass, L. Phase locking, period doubling bifurcations and chaos in a mathematical model of a periodically driven oscillator: A theory for the entrainment of biological oscillators and the generation of cardiac dysrhythmias. *Journal of mathematical biology* **14**, 1–23 (1982).

[174] Horio, Y., Aihara, K. & Yamamoto, O. Neuron-synapse IC chip-set for large-scale chaotic neural networks. *IEEE Transactions on Neural Networks* **14**, 1393–1404 (2003).

[175] Trostel, M. L., Misplon, M. Z., Aragoneses, A. & Pattanayak, A. K. Characterizing complex dynamics in the classical and semi-classical duffing oscillator using ordinal patterns analysis. *Entropy* **20**, 40 (2018).

[176] Amigó, J. M. The equality of Kolmogorov–Sinai entropy and metric permutation entropy generalized. *Physica D: Nonlinear Phenomena* **241**, 789–793 (2012).

[177] Baillie, R. T., Cecen, A. A. & Erkal, C. Normal heartbeat series are nonchaotic, nonlinear, and multifractal: New evidence from semiparametric and parametric tests. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **19**, 028503 (2009).

[178] Zhang, J., Holden, A., Monfredi, O., Boyett, M. R. & Zhang, H. Stochastic vagal modulation of cardiac pacemaking may lead to erroneous identification of cardiac "chaos". *Chaos: An Interdisciplinary Journal of Nonlinear Science* **19**, 028509 (2009).

[179] Ovchinnikov, I. Introduction to supersymmetric theory of stochastics. *Entropy* **18**, 108 (2016).

[180] Letellier, C. & Aguirre, L. A. Investigating nonlinear dynamics from time series: The influence of symmetries and the choice of observables. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **12**, 549–558 (2002).

[181] Izhikevich, E. M. Simple model of spiking neurons. *IEEE Transactions on neural networks* **14**, 1569–1572 (2003).

[182] Venkatesan, A. & Lakshmanan, M. Interruption of torus doubling bifurcation and genesis of strange nonchaotic attractors in a quasiperiodically forced map: Mechanisms and their characterizations. *Physical Review E* **63**, 026219 (2001).

[183] Grebogi, C., Ott, E., Pelikan, S. & Yorke, J. A. Strange attractors that are not chaotic. *Physica D: Nonlinear Phenomena* **13**, 261–268 (1984).

[184] May, R. M. Simple mathematical models with very complicated dynamics. *Nature* **261**, 459 (1976).

[185] Lorenz, E. N. Deterministic nonperiodic flow. *Journal of the atmospheric sciences* **20**, 130–141 (1963).

[186] Richter, H. The generalized henon maps: examples for higher-dimensional chaos. *International Journal of Bifurcation and Chaos* **12**, 1371–1384 (2002).

[187] Freitas, U. S., Letellier, C. & Aguirre, L. A. Failure in distinguishing colored noise from chaos using the "noise titration" technique. *Physical Review E* **79**, 035201 (2009).

[188] Nicolau, J. Stationary processes that look like random walks—the bounded random walk process in discrete and continuous time. *Econometric Theory* **18**, 99–118 (2002).

[189] Timmer, J. Power of surrogate data testing with respect to nonstationarity. *Physical Review E* **58**, 5153 (1998).

[190] Zhivomirov, H. A method for colored noise generation. *Romanian Journal of Acoustics and Vibration* **15**, 14–19 (2018).

[191] Rössler, O. E. An equation for continuous chaos. *Physics Letters A* **57**, 397–398 (1976).

[192] Hammel, S., Jones, C. & Moloney, J. Global dynamical behavior of the optical field in a ring cavity. *JOSA B* **2**, 552–564 (1985).

[193] Hénon, M. A two-dimensional mapping with a strange attractor. In *The Theory of Chaotic Attractors*, 94–102 (Springer, 1976).

[194] Kulp, C. W. & Smith, S. Characterization of noisy symbolic time series. *Physical Review E* **83**, 026201 (2011).

[195] Bernardini, D. & Litak, G. An overview of 0–1 test for chaos. *Journal of the Brazilian Society of Mechanical Sciences and Engineering* **38**, 1433–1450 (2016).

[196] Zunino, L. & Kulp, C. W. Detecting nonlinearity in short and noisy time series using the permutation entropy. *Physics Letters A* **381**, 3627–3635 (2017).

[197] Riedl, M., Müller, A. & Wessel, N. Practical considerations of permutation entropy. *The European Physical Journal Special Topics* **222**, 249–262 (2013).

[198] Leontitsis, A. Chaotic systems toolbox, https://tinyurl.com/uqd9bp4. (2004).

[199] Matthews, P. 0 - 1 test for chaos. https://tinyurl.com/txzk2uj (2009).

[200] Eyébé Fouda, J. S. A., Bodo, B., Sabat, S. L. & Effa, J. Y. A modified 0-1 test for chaos detection in oversampled time series observations. *International Journal of Bifurcation and Chaos* **24**, 1450063 (2014).

[201] Hawkins, R. Permutation entropies and chaos (2015).

[202] Politi, A. Quantifying the dynamical complexity of chaotic time series. *Physical review letters* **118**, 144101 (2017).

[203] Bandt, C., Keller, G. & Pompe, B. Entropy of interval maps via permutations. *Nonlinearity* **15**, 1595 (2002).

[204] Keller, K., Unakafov, A. M. & Unakafova, V. A. On the relation of KS entropy and permutation entropy. *Physica D: Nonlinear Phenomena* **241**, 1477–1481 (2012).

[205] Pesin, Y. B. Characteristic Lyapunov exponents and smooth ergodic theory. *Russian Mathematical Surveys* **32**, 55–114 (1977).

[206] Liley, D. T., Cadusch, P. J. & Wright, J. J. A continuum theory of electro-cortical activity. *Neurocomputing* **26**, 795–800 (1999).

[207] Liley, D. T., Cadusch, P. J. & Dafilis, M. P. A spatially continuous mean field theory of electrocortical activity. *Network: Computation in Neural Systems* **13**, 67–113 (2002).

[208] Izhikevich, E. M. Which model to use for cortical spiking neurons? *IEEE transactions on neural networks* **15**, 1063–1070 (2004).

[209] Gatti, A. *et al.* Cyclically changing phenomenon of cancer cells in an untreated hemophilia patient. *Blood* **41**, 771–773 (1973).

[210] Farmer, J. D. Chaotic attractors of an infinite-dimensional dynamical system. *Physica D: Nonlinear Phenomena* **4**, 366–393 (1982).

[211] Haken, H. Analogy between higher instabilities in fluids and lasers. *Physics Letters A* **53**, 77–78 (1975).

[212] Poland, D. Cooperative catalysis and chemical chaos: a chemical model for the Lorenz equations. *Physica D: Nonlinear Phenomena* **65**, 86–99 (1993).

[213] Ramasubramanian, K. & Sriram, M. A comparative study of computation of Lyapunov spectra with different algorithms. *Physica D: Nonlinear Phenomena* **139**, 72–86 (2000).

[214] Patterson, K. *Unit Root Tests in Time Series Volume 2: Extensions and Developments*, vol. 2 (Palgrave Macmillan, 2012).

[215] Patterson, K. *Unit Root Tests in Time Series Volume 1: Key Concepts and Problems* (Springer, 2011).

[216] Ikeda, K., Daido, H. & Akimoto, O. Optical turbulence: chaotic behavior of transmitted light from a ring cavity. *Physical Review Letters* **45**, 709 (1980).

[217] Galias, Z. Rigorous investigation of the Ikeda map by means of interval arithmetic. *Nonlinearity* **15**, 1759 (2002).

[218] Ivancevic, V. G. & Ivancevic, T. T. *Computational mind: a complex dynamics perspective*, vol. 60 (Springer, 2007).

[219] Glass, L. Resetting and entraining biological rhythms. In *Nonlinear dynamics in physiology and medicine*, 123–148 (Springer, 2003).

[220] Barnett, L. & Seth, A. K. The MVGC multivariate Granger causality toolbox: a new approach to Granger-causal inference. *Journal of neuroscience methods* **223**, 50–68 (2014).

[221] Aihara, K., Takabe, T. & Toyoda, M. Chaotic neural networks. *Physics letters A* **144**, 333–340 (1990).

[222] Fernández, I., Hernández, C. N. & Pacheco, J. M. Is the North Atlantic Oscillation just a pink noise? *Physica A: Statistical Mechanics and its Applications* **323**, 705–714 (2003).

[223] Stephenson, D. B., Pavan, V. & Bojariu, R. Is the North Atlantic Oscillation a random walk? *International Journal of Climatology* **20**, 1–18 (2000).

[224] Collette, C. & Ausloos, M. Scaling analysis and evolution equation of the North Atlantic Oscillation index fluctuations. *International journal of modern physics C* **15**, 1353–1366 (2004).

[225] Lind, P. G., Mora, A., Haase, M. & Gallas, J. A. Minimizing stochasticity in the NAO index. *International Journal of Bifurcation and Chaos* **17**, 3461–3466 (2007).

[226] Martínez Santafé, M. D., Lana Pons, F. J., Burgueño, A. & Serra de Larrocha, C. Predictability of the monthly North Atlantic Oscillation index based on fractal analyses and dynamic system theory. *Nonlinear processes in geophysics* **17**, 93–101 (2010).

[227] Fernández, I., Pacheco, J. M. & Quintana, M. P. Pinkness of the North Atlantic Oscillation signal revisited. *Physica A: Statistical Mechanics and its Applications* **389**, 5801–5807 (2010).

[228] Lynch, S. *Dynamical systems with applications using Matlab*® (Springer, 2014).

[229] Carhart-Harris, R. L. *et al.* The entropic brain: a theory of conscious states informed by neuroimaging research with psychedelic drugs. *Frontiers in human neuroscience* **8**, 20 (2014).

[230] Carhart-Harris, R. L. The entropic brain-revisited. *Neuropharmacology* **142**, 167–178 (2018).

[231] Lempel, A. & Ziv, J. On the complexity of finite sequences. *IEEE Transactions on information theory* **22**, 75–81 (1976).

[232] Cover, T. M. & Thomas, J. A. *Elements of information theory* (John Wiley & Sons, 2012).

[233] Bai, Y., Liang, Z., Li, X., Voss, L. J. & Sleigh, J. W. Permutation lempel–ziv complexity measure of electroencephalogram in gabaergic anaesthetics. *Physiological measurement* **36**, 2483 (2015).

[234] Bai, Y., Liang, Z. & Li, X. A permutation lempel-ziv complexity measure for eeg analysis. *Biomedical Signal Processing and Control* **19**, 102–114 (2015).

[235] Casali, A. G. *et al.* A theoretically based index of consciousness independent of sensory processing and behavior. *Science translational medicine* **5**, 198ra105–198ra105 (2013).

[236] Mateos, D., Erra, R. G., Wennberg, R. & Velazquez, J. P. Measures of entropy and complexity in altered states of consciousness. *Cognitive neurodynamics* **12**, 73–84 (2018).

[237] Schartner, M. *et al.* Complexity of multi-dimensional spontaneous eeg decreases during propofol induced general anaesthesia. *PloS one* **10**, e0133532 (2015).

[238] Schartner, M. M., Carhart-Harris, R. L., Barrett, A. B., Seth, A. K. & Muthukumaraswamy, S. D. Increased spontaneous meg signal diversity for psychoactive doses of ketamine, lsd and psilocybin. *Scientific Reports* **7**, 46421 (2017).

[239] Schartner, M. M. *et al.* Global and local complexity of intracranial eeg decreases during nrem sleep. *Neuroscience of Consciousness* **2017**, niw022 (2017).

[240] Zhang, X.-S., Roy, R. J. & Jensen, E. W. Eeg complexity as a measure of depth of anesthesia for patients. *IEEE transactions on biomedical engineering* **48**, 1424–1433 (2001).

[241] Timmermann, C. *et al.* Neural correlates of the dmt experience as assessed via multivariate eeg. *bioRxiv* 706283 (2019).

[242] Hudetz, A. G., Liu, X., Pillay, S., Boly, M. & Tononi, G. Propofol anesthesia reduces lempel-ziv complexity of spontaneous brain activity in rats. *Neuroscience letters* **628**, 132–135 (2016).

[243] Gosseries, O. *et al.* Automated eeg entropy measurements in coma, vegetative state/unresponsive wakefulness syndrome and minimally conscious state. *Functional neurology* **26**, 25 (2011).

[244] Rowlinson, J. Thomas andrews and the critical point. *Nature* **224**, 541 (1969).

[245] Curie, P. *Propriétés magnétiques des corps a diverses températures.* 4 (Gauthier-Villars et fils, 1895).

[246] Shew, W. L. & Plenz, D. The functional benefits of criticality in the cortex. *The neuroscientist* **19**, 88–100 (2013).

[247] Beggs, J. M. The criticality hypothesis: how local cortical networks might optimize information processing. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **366**, 329–343 (2007).

[248] Sornette, D. *Critical phenomena in natural sciences: chaos, fractals, selforganization and disorder: concepts and tools* (Springer Science & Business Media, 2006).

[249] Beggs, J. M. & Plenz, D. Neuronal avalanches in neocortical circuits. *Journal of neuroscience* **23**, 11167–11177 (2003).

[250] Friedman, N. *et al.* Universal critical dynamics in high resolution neuronal avalanche data. *Physical review letters* **108**, 208102 (2012).

[251] Hahn, G. *et al.* Spontaneous cortical activity is transiently poised close to criticality. *PLoS computational biology* **13**, e1005543 (2017).

[252] Mazzoni, A. *et al.* On the dynamics of the spontaneous activity in neuronal networks. *PloS one* **2**, e439 (2007).

[253] Pasquale, V., Massobrio, P., Bologna, L., Chiappalone, M. & Martinoia, S. Self-organization and neuronal avalanches in networks of dissociated cortical neurons. *Neuroscience* **153**, 1354–1369 (2008).

[254] Priesemann, V. *et al.* Spike avalanches in vivo suggest a driven, slightly subcritical brain state. *Frontiers in systems neuroscience* **8**, 108 (2014).

[255] Shriki, O. *et al.* Neuronal avalanches in the resting meg of the human brain. *Journal of Neuroscience* **33**, 7079–7090 (2013).

[256] Tagliazucchi, E., Balenzuela, P., Fraiman, D. & Chialvo, D. R. Criticality in large-scale brain fmri dynamics unveiled by a novel point process analysis. *Frontiers in physiology* **3**, 15 (2012).

[257] Ponce-Alvarez, A., Jouary, A., Privat, M., Deco, G. & Sumbre, G. Whole-brain neuronal activity displays crackling noise dynamics. *Neuron* **100**, 1446–1459 (2018).

[258] Shew, W. L., Yang, H., Yu, S., Roy, R. & Plenz, D. Information capacity and transmission are maximized in balanced cortical networks with neuronal avalanches. *Journal of neuroscience* **31**, 55–63 (2011).

[259] Fagerholm, E. D. *et al.* Cortical entropy, mutual information and scale-free dynamics in waking mice. *Cerebral cortex* **26**, 3945–3952 (2016).

[260] Tkačik, G. *et al.* The simplest maximum entropy model for collective behavior in a neural network. *Journal of Statistical Mechanics: Theory and Experiment* **2013**, P03011 (2013).

[261] Ribeiro, T. L. *et al.* Spike avalanches exhibit universal dynamics across the sleep-wake cycle. *PloS one* **5**, e14129 (2010).

[262] Meisel, C., Storch, A., Hallmeyer-Elgner, S., Bullmore, E. & Gross, T. Failure of adaptive self-organized criticality during epileptic seizure attacks. *PLoS computational biology* **8**, e1002312 (2012).

[263] Suffczynski, P., Kalitzin, S. & Da Silva, F. L. Dynamics of non-convulsive epileptic phenomena modeled by a bistable neuronal network. *Neuroscience* **126**, 467–484 (2004).

[264] Fröhlich, F., Sejnowski, T. J. & Bazhenov, M. Network bistability mediates spontaneous transitions between normal and pathological brain states. *Journal of Neuroscience* **30**, 10734–10743 (2010).

[265] Breakspear, M. *et al.* A unifying explanation of primary generalized seizures through nonlinear brain modeling and bifurcation analysis. *Cerebral Cortex* **16**, 1296–1313 (2005).

[266] Robinson, P., Rennie, C. & Rowe, D. Dynamics of large-scale brain activity in normal arousal states and epileptic seizures. *Physical Review E* **65**, 041924 (2002).

[267] Osorio, I., Frei, M. G., Sornette, D., Milton, J. & Lai, Y.-C. Epileptic seizures: quakes of the brain? *Physical Review E* **82**, 021919 (2010).

[268] Worrell, G. A., Cranstoun, S. D., Echauz, J. & Litt, B. Evidence for self-organized criticality in human epileptic hippocampus. *Neuroreport* **13**, 2017–2021 (2002).

[269] Atasoy, S. *et al.* Connectome-harmonic decomposition of human brain activity reveals dynamical repertoire re-organization under lsd. *Scientific reports* **7**, 17661 (2017).

[270] Tagliazucchi, E., Carhart-Harris, R., Leech, R., Nutt, D. & Chialvo, D. R. Enhanced repertoire of brain dynamical states during the psychedelic experience. *Human brain mapping* **35**, 5442–5456 (2014).

[271] Alexander, D. M. & Globus, G. G. Edge-of-chaos dynamics in recursively organized neural systems. *Fractals of brain, fractals of mind* 31–73 (1996).

[272] Bertschinger, N. & Natschläger, T. Real-time computation at the edge of chaos in recurrent neural networks. *Neural computation* **16**, 1413–1436 (2004).

[273] Legenstein, R. & Maass, W. Edge of chaos and prediction of computational performance for neural circuit models. *Neural Networks* **20**, 323–334 (2007).

[274] Kadmon, J. & Sompolinsky, H. Transition to chaos in random neuronal networks. *Physical Review X* **5**, 041030 (2015).

[275] Aljadeff, J., Stern, M. & Sharpee, T. Transition to chaos in random networks with cell-type-specific connectivity. *Physical review letters* **114**, 088101 (2015).

[276] Steyn-Ross, M. L., Steyn-Ross, D. A. & Sleigh, J. W. Interacting turing-hopf instabilities drive symmetry-breaking transitions in a mean-field model of the cortex: a mechanism for the slow oscillation. *Physical Review X* **3**, 021005 (2013).

[277] Kanders, K., Lorimer, T. & Stoop, R. Avalanche and edge-of-chaos criticality do not necessarily co-occur in neural networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **27**, 047408 (2017).

[278] Crutchfield, J. P. & Young, K. Computation at the onset of chaos. In *The Santa Fe Institute, Westview* (Citeseer, 1988).

[279] Langton, C. G. Computation at the edge of chaos: phase transitions and emergent computation. *Physica D: Nonlinear Phenomena* **42**, 12–37 (1990).

[280] Boedecker, J., Obst, O., Lizier, J. T., Mayer, N. M. & Asada, M. Information processing in echo state networks at the edge of chaos. *Theory in Biosciences* **131**, 205–213 (2012).

[281] Roli, A., Villani, M., Filisetti, A. & Serra, R. Dynamical criticality: overview and open questions. *Journal of Systems Science and Complexity* **31**, 647–663 (2018).

[282] Korn, H. & Faure, P. Is there chaos in the brain? ii. experimental evidence and related models. *Comptes rendus biologies* **326**, 787–840 (2003).

[283] Glass, L. & Mackey, M. C. *From clocks to chaos: The rhythms of life* (Princeton University Press, 1988).

[284] Coombes, S., beim Graben, P., Potthast, R. & Wright, J. *Neural fields: theory and applications* (Springer, 2014).

[285] Deco, G., Jirsa, V. K., Robinson, P. A., Breakspear, M. & Friston, K. The dynamic brain: from spiking neurons to neural masses and cortical fields. *PLoS computational biology* **4**, e1000092 (2008).

[286] Breakspear, M. Dynamic models of large-scale brain activity. *Nature neuroscience* **20**, 340 (2017).

[287] Steyn-Ross, M. L., Steyn-Ross, D. A., Sleigh, J. W. & Liley, D. Theoretical electroencephalogram stationary spectrum for a white-noise-driven cortex: evidence for a general anesthetic-induced phase transition. *Physical Review E* **60**, 7299 (1999).

[288] Martinet, L.-E. *et al.* Human seizures couple across spatial scales through travelling wave dynamics. *Nature communications* **8**, 14896 (2017).

[289] Voss, L. J., Jacobson, G., Sleigh, J. W., Steyn-Ross, A. & Steyn-Ross, M. Excitatory effects of gap junction blockers on cerebral cortex seizure-like activity in rats and mice. *Epilepsia* **50**, 1971–1978 (2009).

[290] Wilson, M. T., Sleigh, J. W., Steyn-Ross, D. A. & Steyn-Ross, M. L. General anesthetic-induced seizures can be explained by a mean-field model of cortical dynamics. *Anesthesiology: The Journal of the American Society of Anesthesiologists* **104**, 588–593 (2006).

[291] Steyn-Ross, M. L., Steyn-Ross, D. A. & Sleigh, J. W. Gap junctions modulate seizures in a mean-field model of general anesthesia for the cortex. *Cognitive neurodynamics* **6**, 215–225 (2012).

[292] Steyn-Ross, M. L., Steyn-Ross, D. A., Sleigh, J. W. & Wilcocks, L. C. Toward a theory of the general-anesthetic-induced phase transition of the cerebral cortex. i. a thermodynamics analogy. *Physical Review E* **64**, 011917 (2001).

[293] Steyn-Ross, D. A., Steyn-Ross, M. L., Wilcocks, L. C. & Sleigh, J. W. Toward a theory of the general-anesthetic-induced phase transition of the cerebral cortex. ii. numerical simulations, spectral entropy, and correlation times. *Physical Review E* **64**, 011918 (2001).

[294] Wilson, M. T. *et al.* An analysis of the transitions between down and up states of the cortical slow oscillation under urethane anaesthesia. *Journal of biological physics* **36**, 245–259 (2010).

[295] Gottwald, G. A. & Melbourne, I. A new test for chaos in deterministic systems. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 460, 603–611 (The Royal Society, 2004).

[296] Gottwald, G. A. & Melbourne, I. Testing for chaos in deterministic systems with noise. *Physica D: Nonlinear Phenomena* **212**, 100–110 (2005).

[297] Gottwald, G. A. & Melbourne, I. On the implementation of the 0–1 test for chaos. *SIAM Journal on Applied Dynamical Systems* **8**, 129–145 (2009).

[298] Gottwald, G. A. & Melbourne, I. Comment on "reliability of the 0-1 test for chaos". *Physical Review E* **77**, 028201 (2008).

[299] Dawes, J. & Freeland, M. The '0–1 test for chaos' and strange nonchaotic attractors. *preprint* (2008).

[300] Eyébé Fouda, J. S. A., Bodo, B., Sabat, S. L. & Effa, J. Y. A modified 0-1 test for chaos detection in oversampled time series observations. *International Journal of Bifurcation and Chaos* **24**, 1450063 (2014).

[301] Carhart-Harris, R. L. *et al.* Neural correlates of the lsd experience revealed by multimodal neuroimaging. *Proceedings of the National Academy of Sciences* **113**, 4853–4858 (2016).

[302] Aghajanian, G. & Marek, G. Serotonin induces excitatory postsynaptic potentials in apical dendrites of neocortical pyramidal cells. *Neuropharmacology* **36**, 589–599 (1997).

[303] Puig, M. V., Celada, P., Díaz-Mataix, L. & Artigas, F. In vivo modulation of the activity of pyramidal neurons in the rat medial prefrontal cortex by 5-ht2a receptors: relationship to thalamocortical afferents. *Cerebral Cortex* **13**, 870–882 (2003).

[304] Nichols, D. E. Psychedelics. *Pharmacological reviews* **68**, 264–355 (2016).

[305] Béïque, J.-C., Imad, M., Mladenovic, L., Gingrich, J. A. & Andrade, R. Mechanism of the 5-hydroxytryptamine 2a receptor-mediated facilitation of synaptic activity in prefrontal cortex. *Proceedings of the National Academy of Sciences* **104**, 9870–9875 (2007).

[306] Shew, W. L., Yang, H., Petermann, T., Roy, R. & Plenz, D. Neuronal avalanches imply maximum dynamic range in cortical networks at criticality. *Journal of neuroscience* **29**, 15595–15600 (2009).

[307] Yang, H., Shew, W. L., Roy, R. & Plenz, D. Maximal variability of phase synchrony in cortical networks with neuronal avalanches. *Journal of neuroscience* **32**, 1061–1072 (2012).

[308] Chen, W., Hobbs, J. P., Tang, A. & Beggs, J. M. A few strong connections: optimizing information retention in neuronal avalanches. *BMC neuroscience* **11**, 3 (2010).

[309] Ovchinnikov, I. V. *et al.* Criticality or supersymmetry breaking? *arXiv preprint arXiv:1609.00001* (2016).

[310] Ovchinnikov, I. Introduction to supersymmetric theory of stochastics. *Entropy* **18**, 108 (2016).

[311] Ovchinnikov, I. V., Schwartz, R. N. & Wang, K. L. Topological supersymmetry breaking: The definition and stochastic generalization of chaos and the limit of applicability of statistics. *Modern Physics Letters B* **30**, 1650086 (2016).

[312] Parisi, G. & Sourlas, N. Random magnetic fields, supersymmetry, and negative dimensions. *Physical Review Letters* **43**, 744 (1979).

[313] Parisi, G. & Sourlas, N. Supersymmetric field theories and stochastic differential equations. *Nuclear Physics B* **206**, 321–332 (1982).

[314] Gozzi, E. & Reuter, M. Classical mechanics as a topological field theory. *Physics Letters B* **240**, 137–144 (1990).

[315] Ovchinnikov, I. V. Topological field theory of dynamical systems. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **22**, 033134 (2012).

[316] Ovchinnikov, I. V. & Ensslin, T. A. Kinematic dynamo, supersymmetry breaking, and chaos. *Physical Review D* **93**, 085023 (2016).

[317] Stam, C., Pijn, J., Suffczynski, P. & Da Silva, F. L. Dynamics of the human alpha rhythm: evidence for non-linearity? *Clinical Neurophysiology* **110**, 1801–1813 (1999).

[318] Breakspear, M. & Terry, J. Detection and description of non-linear interdependence in normal multichannel human eeg data. *Clinical neurophysiology* **113**, 735–753 (2002).

[319] Theiler, J. & Rapp, P. E. Re-examination of the evidence for low-dimensional, nonlinear structure in the human electroencephalogram. *Electroencephalography and clinical Neurophysiology* **98**, 213–222 (1996).

[320] Andrzejak, R. G. *et al.* Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Physical Review E* **64**, 061907 (2001).

[321] Friston, K. J., Jezzard, P. & Turner, R. Analysis of functional mri time-series. *Human brain mapping* **1**, 153–171 (1994).

[322] Uenohara, S. *et al.* Experimental distinction between chaotic and strange nonchaotic attractors on the basis of consistency. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **23**, 023110 (2013).

[323] Wood, J., Kim, Y. & Moghaddam, B. Disruption of prefrontal cortex large scale neuronal activity by different classes of psychotomimetic drugs. *Journal of Neuroscience* **32**, 3022–3031 (2012).

[324] Muthukumaraswamy, S. D. *et al.* Broadband cortical desynchronization underlies the human psychedelic state. *Journal of Neuroscience* **33**, 15171–15183 (2013).

[325] Deco, G. *et al.* Whole-brain multimodal neuroimaging model using serotonin receptor maps explains non-linear functional effects of lsd. *Current Biology* **28**, 3065–3074 (2018).

[326] Janowsky, A. *et al.* Mefloquine and psychotomimetics share neurotransmitter receptor and transporter interactions in vitro. *Psychopharmacology* **231**, 2771–2783 (2014).

[327] Rörig, B. & Sutor, B. Serotonin regulates gap junction coupling in the developing rat somatosensory cortex. *European Journal of Neuroscience* **8**, 1685–1695 (1996).

[328] Marona-Lewicka, D. & Nichols, D. E. Potential serotonin 5-ht1a and dopamine d4 receptor modulation of the discriminative stimulus effects of amphetamine in rats. *Behavioural pharmacology* **22**, 508 (2011).

[329] Vollenweider, F. X., Vollenweider-Scherpenhuyzen, M. F., Bäbler, A., Vogel, H. & Hell, D. Psilocybin induces schizophrenia-like psychosis in humans via a serotonin-2 agonist action. *Neuroreport* **9**, 3897–3902 (1998).

[330] Moreno, J. L. *et al.* Persistent effects of chronic clozapine on the cellular and behavioral responses to lsd in mice. *Psychopharmacology* **225**, 217–226 (2013).

[331] Onn, S.-P. & Grace, A. A. Repeated treatment with haloperidol and clozapine exerts differential effects on dye coupling between neurons in subregions of striatum and nucleus accumbens. *Journal of Neuroscience* **15**, 7024–7036 (1995).

[332] Franco-Pérez, J., Ballesteros-Zebadúa, P. & Manjarrez-Marmolejo, J. Anticonvulsant effects of mefloquine on generalized tonic-clonic seizures induced by two acute models in rats. *BMC neuroscience* **16**, 7 (2015).

[333] Sturkenboom, M., Pepplinkhuizen, L., Stricker, B. *et al.* Mefloquine increases the risk of serious psychiatric events during travel abroad: a nationwide case-control study in the netherlands. *The Journal of clinical psychiatry* **66**, 199–204 (2005).

[334] Cruikshank, S. J. *et al.* Potent block of cx36 and cx50 gap junction channels by mefloquine. *Proceedings of the National Academy of Sciences* **101**, 12364–12369 (2004).

[335] Iglesias, R. *et al.* P2x7 receptor-pannexin1 complex: pharmacology and signaling. *American Journal of Physiology-Cell Physiology* **295**, C752–C760 (2008).

[336] Wang, Y. *et al.* Neuronal gap junctions are required for nmda receptor–mediated excitotoxicity: Implications in ischemic stroke. *Journal of neurophysiology* **104**, 3551–3556 (2010).

[337] Sarihi, A. *et al.* Cell type-specific, presynaptic ltp of inhibitory synapses on fast-spiking gabaergic neurons in the mouse visual cortex. *Journal of Neuroscience* **32**, 13189–13199 (2012).

[338] Sporns, O. *Networks of the Brain* (MIT press, 2010).

[339] Steriade, M., McCormick, D. A. & Sejnowski, T. J. Thalamocortical oscillations in the sleeping and aroused brain. *Science* **262**, 679–685 (1993).

[340] Liley, D. T., Cadusch, P. J. & Wright, J. J. A continuum theory of electro-cortical activity. *Neurocomputing* **26**, 795–800 (1999).

[341] Liley, D. T., Cadusch, P. J. & Dafilis, M. P. A spatially continuous mean field theory of electrocortical activity. *Network: Computation in Neural Systems* **13**, 67–113 (2002).

[342] Van Vreeswijk, C. & Sompolinsky, H. Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science* **274**, 1724–1726 (1996).

[343] Vreeswijk, C. v. & Sompolinsky, H. Chaotic balanced state in a model of cortical circuits. *Neural computation* **10**, 1321–1371 (1998).

[344] Li, J. & Shew, W. Tuning network dynamics from criticality to the chaotic balanced state. *BioRxiv* 551457 (2019).

[345] Monteforte, M. & Wolf, F. Dynamical entropy production in spiking neuron networks in the balanced state. *Physical review letters* **105**, 268104 (2010).

[346] Brunel, N. Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *Journal of computational neuroscience* **8**, 183–208 (2000).

[347] Trapp, P., Echeveste, R. & Gros, C. Ei balance emerges naturally from continuous hebbian learning in autonomous neural networks. *Scientific reports* **8**, 8939 (2018).

[348] Hesse, J. & Gross, T. Self-organized criticality as a fundamental property of neural systems. *Frontiers in systems neuroscience* **8**, 166 (2014).

[349] Gireesh, E. D. & Plenz, D. Neuronal avalanches organize as nested theta-and beta/gamma-oscillations during development of cortical layer 2/3. *Proceedings of the National Academy of Sciences* **105**, 7576–7581 (2008).

[350] Agrawal, V. *et al.* Robust entropy requires strong and balanced excitatory and inhibitory synapses. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **28**, 103115 (2018).

[351] Shinohara, K., Hiruma, H., Funabashi, T. & Kimura, F. Gabaergic modulation of gap junction communication in slice cultures of the rat suprachiasmatic nucleus. *Neuroscience* **96**, 591–596 (2000).

[352] Wentlandt, K., Samoilova, M., Carlen, P. L. & El Beheiry, H. General anesthetics inhibit gap junction communication in cultured organotypic hippocampal slices. *Anesthesia & Analgesia* **102**, 1692–1698 (2006).

[353] Kitamura, A., Marszalec, W., Yeh, J. Z. & Narahashi, T. Effects of halothane and propofol on excitatory and inhibitory synaptic transmission in rat cortical neurons. *Journal of Pharmacology and Experimental Therapeutics* **304**, 162–171 (2003).

[354] Pais, I. *et al.* Sharp wave-like activity in the hippocampus in vitro in mice lacking the gap junction protein connexin 36. *Journal of neurophysiology* **89**, 2046–2054 (2003).

[355] Jacobson, G. M. *et al.* Connexin36 knockout mice display increased sensitivity to pentylenetetrazol-induced seizure-like behaviors. *Brain research* **1360**, 198–204 (2010).

[356] Yang, L. & Ling, D. S. Carbenoxolone modifies spontaneous inhibitory and excitatory synaptic transmission in rat somatosensory cortex. *Neuroscience letters* **416**, 221–226 (2007).

[357] Gajda, Z., Gyengési, E., Hermesz, E., Ali, K. S. & Szente, M. Involvement of gap junctions in the manifestation and control of the duration of seizures in rats in vivo. *Epilepsia* **44**, 1596–1600 (2003).

[358] Gajda, Z., Szupera, Z., Blazsó, G. & Szente, M. Quinine, a blocker of neuronal cx36 channels, suppresses seizure activity in rat neocortex in vivo. *Epilepsia* **46**, 1581–1591 (2005).

[359] Nilsen, K. E., Kelso, A. R. & Cock, H. R. Antiepileptic effect of gap-junction blockers in a rat model of refractory focal cortical epilepsy. *Epilepsia* **47**, 1169–1175 (2006).

[360] Jahromi, S. S., Wentlandt, K., Piran, S. & Carlen, P. L. Anticonvulsant actions of gap junctional blockers in an in vitro seizure model. *Journal of neurophysiology* **88**, 1893–1902 (2002).

[361] Bennett, M. V. & Zukin, R. S. Electrical coupling and neuronal synchronization in the mammalian brain. *Neuron* **41**, 495–511 (2004).

[362] Destexhe, A. Oscillations, complex spatiotemporal behavior, and information transport in networks of excitatory and inhibitory neurons. *Physical Review E* **50**, 1594 (1994).

[363] Obrist, W. D. The electroencephalogram of normal aged adults. *Electroencephalography & Clinical Neurophysiology* (1954).

[364] Klimesch, W. Eeg alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain research reviews* **29**, 169–195 (1999).

[365] Haegens, S., Cousijn, H., Wallis, G., Harrison, P. J. & Nobre, A. C. Inter-and intra-individual variability in alpha peak frequency. *Neuroimage* **92**, 46–55 (2014).

[366] Samaha, J. & Postle, B. R. The speed of alpha-band oscillations predicts the temporal resolution of visual perception. *Current Biology* **25**, 2985–2990 (2015).

[367] Mierau, A., Klimesch, W. & Lefebvre, J. State-dependent alpha peak frequency shifts: Experimental evidence, potential mechanisms and functional implications. *Neuroscience* **360**, 146–154 (2017).

[368] Bullock, T. Neuroethology deserves more study of evoked responses. *Neuroscience* **6**, 1203–1215 (1981).

[369] Haller, M. *et al.* Parameterizing neural power spectra. *bioRxiv* 299859 (2018).

[370] Miller, K. J., Sorensen, L. B., Ojemann, J. G. & Den Nijs, M. Power-law scaling in the brain surface electric potential. *PLoS computational biology* **5**, e1000609 (2009).

[371] Kulp, C. W. & Smith, S. Characterization of noisy symbolic time series. *Physical Review E* **83**, 026201 (2011).

[372] Ihle, M. *et al.* Epilepsiae–a european epilepsy database. *Computer methods and programs in biomedicine* **106**, 127–138 (2012).

[373] Nagasaka, Y., Shimoda, K. & Fujii, N. Multidimensional recording (mdr) and data sharing: an ecological open research and educational platform for neuroscience. *PloS one* **6**, e22561 (2011).