OPEN ACCESS Freely available online

PLoS COMPUTATIONAL BIOLOGY

Practical Measures of Integrated Information for Time-Series Data

Adam B. Barrett*, Anil K. Seth

Sackler Centre for Consciousness Science and School of Informatics, University of Sussex, Brighton, United Kingdom

Abstract

A recent measure of 'integrated information', Φ_{DM}, quantifies the extent to which a system generates more information than the sum of its parts as it transitions between states, possibly reflecting levels of consciousness generated by neural systems. However, Φ_{DM} is defined only for discrete Markov systems, which are unusual in biology; as a result, Φ_{DM} can rarely be measured in practice. Here, we describe two new measures, Φ_E and Φ_{AR}, that overcome these limitations and are easy to apply to time-series data. We use simulations to demonstrate the in-practice applicability of our measures, and to explore their properties. Our results provide new opportunities for examining information integration in real and model systems and carry implications for relations between integrated information, consciousness, and other neurocognitive processes. However, our findings pose challenges for theories that ascribe physical meaning to the measured quantities.

Citation: Barrett AB, Seth AK (2011) Practical Measures of Integrated Information for Time-Series Data. PLoS Comput Biol 7(1): e1001052. doi:10.1371/journal.pcbi.1001052

Editor: Olaf Sporns, Indiana University, United States of America

Received June 20, 2010; **Accepted** December 6, 2010; **Published** January 20, 2011

Funding: AKS is supported by EPSRC Leadership Fellowship EP/G007543/1, which also supports the work of ABB. Support is also gratefully acknowledged from the Dr. Mortimer and Theresa Sackler Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: adam.barrett@sussex.ac.uk

Introduction

How can the complex dynamics exhibited by networks of interconnected elements best be measured? Answering this question promises to shed substantial new light on many complex systems, biological and non-biological. Neural systems in particular are characterized by richly interconnected elements exhibiting complex dynamics at multiple spatiotemporal scales [1], which have been associated with a variety of behavioral, cognitive, and phenomenal properties [2,3,4]. Characterizing dynamical complexity for such systems therefore presents a key challenge for developing new theoretical accounts [5] and for designing and evaluating new experiments. A common and attractive intuition is that dynamical complexity consists in the coexistence of *differentiation* (subsets of a system are dynamically distinct) and *integration* (the system as a whole exhibits coherence) in a system's dynamics. Applied to neural systems, this intuition may underpin notions of cognitive and behavioral flexibility. A system that is able to respond specifically and selectively to a broad range of stimuli, in an integrated way, may require conjoined functional integration and differentiation [6,7]. More ambitiously, the intuition may also characterize basic aspects of conscious experience [8]. At the phenomenal level, each conscious scene is composed of many different parts and is different from every other conscious scene ever experienced (differentiation), yet each conscious scene is experienced as a coherent whole (integration). Therefore, dynamical complexity in neural systems may actually *account for* (and not merely correlate with) fundamental aspects of consciousness [9].

Several measures now exist which operationalize the above intuition under different assumptions and with varying practical applicability [5]. In this paper, we critically evaluate 'integrated information' (Φ) [10,11], a candidate measure that has received significant recent attention, especially in the domain of consciousness science [12,13,14,15]. We present new versions of this measure that are both theoretically well-grounded and, in contrast to previous versions, practically applicable given time-series data. Φ has been proposed as a measure of the amount of information that is integrated by a system, where 'information' reflects the differentiated states of a system and 'integration' their global cohesion. According to the 'integrated information theory of consciousness' (IITC), this quantity is identical to the quantity of consciousness generated by the system; in other words, on the IITC, consciousness *is* integrated information [12,14]. This dramatic claim invites a close examination of the in-principle and in-practice properties of Φ.

A first version of Φ (which we call Φ_C, 'Φ-capacity',) was conceived as a measure of the *capacity* of a system to integrate information, and did not take into account time or changing dynamics [10,12]. Also, measuring Φ_C requires flexible, repeated, and reversible perturbation of arbitrary system subsets, which is infeasible for non-trivial systems (except in simulation). We do not discuss this measure any further. Recently, a new version of Φ has been introduced in the context of the IITC, which we call Φ_{DM}, 'Φ-discrete/Markov' [11]. In contrast to Φ_C, Φ_{DM} is defined for systems of discrete elements that evolve through time with Markovian transitions. Specifically, Φ_{DM} measures the information generated when a system transitions to one particular state out of a repertoire of possible states, but only to the extent that this information is generated by the whole system, over and above the information generated independently by the parts [11]. Importantly, Φ_{DM} measures information as reduction in entropy from a prior *maximum entropy* distribution, which is taken to represent the repertoire of possible states.

Author Summary

A key feature of the human brain is its ability to represent a vast amount of information, and to integrate this information in order to produce specific and selective behaviour, as well as a stream of unified conscious scenes. Attempts have been made to quantify so-called 'integrated information' by formalizing in mathematics the extent to which a system as a whole generates more information than the sum of its parts. However, so far, the resulting measures have turned out to be inapplicable to real neural systems. In this paper we introduce two new measures that can be applied to both realistic neural models and to time-series data garnered from a broad range of neuroimaging and electrophysiological methods. Our work provides new opportunities for examining the role of integrated information in cognition and consciousness, and indeed in the function of any complex biological system. However, our results also pose challenges for theories that ascribe a direct physical meaning to any version of integrated information so far described.

It has been shown, using simulations, that Φ_{DM} behaves consistently with several intuitions about dynamical complexity [11]. In particular, high values of Φ_{DM} are generated by networks that exhibit both differentiation and integration in their dynamics. However, Φ_{DM} is defined only for idealized discrete Markovian systems (a Markovian system is one for which the future depends only on the present, and not on the past). This in-principle restriction severely limits its in-practice applicability because complex biological systems are typically continuous (or are measured as continuous) and are non-Markovian). This limitation in turn imposes a serious obstacle for developing and evaluating theories, such as the IITC, which depend on quantifying integrated information.

In this paper we introduce an alternative measure of integrated information, Φ_E ('Φ-empirical'), which is applicable to time-series data, and to continuous or discrete stochastic systems, Markovian or otherwise (and without perturbation of the studied system). These key features arise because Φ_E is based on the reduction in Shannon entropy from the empirical, as opposed to the maximum entropy, distribution. Our basic formulation of Φ_E therefore addresses the in-principle restrictions of Φ_{DM} mentioned above. Φ_E is best suited for application to stationary systems, for which it provides a single value for a given stationary epoch. However, its in-practice applicability still faces the difficulty of accurately estimating entropies from limited data. This is a problem that scales poorly as the number of elements (variables) increases, especially for continuous systems [16]. Confronting this problem, we show that when states are Gaussian distributed, Φ_E can be computed directly from empirical covariance matrices, rendering it extremely easy to apply in practice for these systems. Meanwhile, for non-Gaussian systems, we introduce a second measure, Φ_{AR} ('auto-regressive Φ'), which is based on auto-regressive prediction error. Φ_{AR} can be understood as measuring how well the present state of a system predicts some previous state, but only to the extent that predictions based on the whole outstrip predictions based on the parts considered independently. Φ_{AR} and Φ_E are constructed analogously, and indeed for Gaussian systems we are able to show, using a connection between linear regression and information theory [17,18], that they are precisely equivalent. Recognizing this equivalence allows us to interpret Φ_E in the same way as Φ_{AR}, i.e., in terms of predictive ability. Importantly, although for non-Gaussian systems Φ_{AR} and Φ_E may differ, the former remains easy to measure in practice from empirical covariance matrices.

The difference between Φ_E/Φ_{AR} and Φ_{DM} is not only a matter of practical applicability. Using the empirical distribution as opposed to the maximum entropy distribution substantially changes possible interpretations of the measure. According to Φ_E, integrated information is a measure of a *process*, since the empirical distribution is a characterization of the actual behavior of the system. According to Φ_{DM} integrated information is to some extent a measure of *capacity* [14], since the maximum entropy distribution is maximally agnostic about the behavior of the system, representing instead its potential or capacity.

The above distinction carries implications for theories, such as the IITC, that ascribe physical meaning to measures of integrated information. Under the IITC, consciousness is explicitly characterized in terms of the capacity of a system [14], and not, following William James [19], as a process. Our new measures imply a Jamesian modification of the IITC by considering consciousness as a process; they also challenge the identity relation between consciousness and integrated information assumed in the IITC. More generally, many other brain-based phenomena are best considered in terms of process rather than capacity, and may admit useful interpretations in terms of integrated information. For example, multi-modal binding and perceptual categorization [20] could involve integrated information in the perceptual domain, and action selection (decision making) [21] may require the integration of sensory, cognitive and motor processes, while retaining differentiation among competing alternatives. In these and other cases, having a measure of integrated information framed in terms of process, that is practically applicable to time-series data, will permit the formulation of testable hypotheses and synthetic models relating information integration to cognitive and neural operations.

Results

The 'Results' section is organized as follows. In the 'Notation, conventions and preliminaries' section we lay out our notation and introduce some necessary mathematical concepts. In the section 'The previous measure, Φ_{DM}' we review Φ_{DM} using our current notation, noting its limitations especially with respect to discrete Markovian systems. The section 'The new measure, Φ_E' describes the new measure Φ_E and provides practical recipes for its computation either numerically from time-series or analytically, given a generative model of the system, both under Gaussian assumptions. We note that for non-Gaussian systems Φ_E remains well-defined even if it is more challenging to calculate. The section 'Φ_E for Markovian Gaussian systems' presents the results of various simulations, designed to illustrate the in-practice applicability of Φ_E and to explore its properties. We compute Φ_E for some canonical networks, optimize connectivity under simple dynamics, and examine the numerical stability of the measure. We also compare Φ_E with a version of Φ_{DM} modified to apply to continuous systems, showing quantitative congruence in most cases. The section 'Extension to multiple lags and to $\mathrm{MVAR}(p)$ processes' describes some additional simulation results, showing how Φ_E can measure integrated information over arbitrary time-steps (lags). In the section 'Auto-regressive Φ (Φ_{AR})' we describe Φ_{AR} and explain its derivation in terms of relations among conditional entropy, covariance, and linear regression prediction error. We demonstrate the utility of Φ_{AR} by calculating integrated information for representative systems animated by exponentially distributed (i.e., non-Gaussian) dynamics.

Notation, conventions and preliminaries

We use bold upper-case letters to denote multivariate random variables, and corresponding bold lower-case letters to denote actualizations of random variables. Matrices are denoted by upper-case letters. The n-dimensional identity matrix is denoted by I_n and the n-dimensional square matrix of zeros by O_n. The transpose operator is denoted by '$^{\mathrm{T}}$', and the determinant by 'det'. Our convention for logarithms is to take them to the natural base e, and to denote them by 'log'.

Let $\boldsymbol{X}=\left(X^1,\ldots,X^n\right)^{\mathrm{T}}$ be a random variable that takes values in the space $\boldsymbol{\Omega_X}$. Then we denote the probability density function by $P_{\boldsymbol{X}}$, the mean by $\bar{\boldsymbol{x}}$ and the $n\times n$ matrix of covariances, $\mathrm{cov}(X^i,X^j)$, by $\Sigma(\boldsymbol{X})$. Let $\boldsymbol{Y}=\left(Y^1,Y^2,\ldots,Y^m\right)^{\mathrm{T}}$ be a second random variable. Then we denote the $n\times m$ matrix of cross-covariances, $\mathrm{cov}(X^i,Y^j)$, by $\Sigma(\boldsymbol{X},\boldsymbol{Y})$. The following quantity will be useful:

$$\Sigma(\boldsymbol{X}|\boldsymbol{Y}) =: \Sigma(\boldsymbol{X})-\Sigma(\boldsymbol{X},\boldsymbol{Y})\Sigma(\boldsymbol{Y})^{-1}\Sigma(\boldsymbol{X},\boldsymbol{Y})^{\mathrm{T}}. \tag{0.1}$$

We call this the partial covariance of $\boldsymbol{X}$ given $\boldsymbol{Y}$, and it is well-defined when $\Sigma(\boldsymbol{Y})$ is invertible. If $\boldsymbol{X}$ and $\boldsymbol{Y}$ are both multivariate Gaussian variables then the partial covariance $\Sigma(\boldsymbol{X}|\boldsymbol{Y})$ is precisely the covariance matrix of the conditional variable $\boldsymbol{X}|\boldsymbol{Y}=\boldsymbol{y}$, for any $\boldsymbol{y}$:

$$\boldsymbol{X}|(\boldsymbol{Y}=\boldsymbol{y})\sim\mathcal{N}\left[\boldsymbol{\mu_y},\Sigma(\boldsymbol{X}|\boldsymbol{Y})\right], \tag{0.2}$$

where $\boldsymbol{\mu_y}=\bar{\boldsymbol{x}}+\Sigma(\boldsymbol{X},\boldsymbol{Y})\Sigma(\boldsymbol{Y})^{-1}(\boldsymbol{y}-\bar{\boldsymbol{y}})$.

Entropy H characterizes uncertainty, and is given by

$$H(\boldsymbol{X}) =: -\sum_{\boldsymbol{x}\in\Omega_{\boldsymbol{X}}} P_{\boldsymbol{X}}(\boldsymbol{x})\log P_{\boldsymbol{X}}(\boldsymbol{x}), \tag{0.3}$$

if $\boldsymbol{X}$ is a discrete random variable, or

$$H(\boldsymbol{X}) =: -\int_{\mathbb{R}^n} P_{\boldsymbol{X}}(\boldsymbol{x})\log P_{\boldsymbol{X}}(\boldsymbol{x})\mathrm{d}^n\boldsymbol{x} \tag{0.4}$$

if $\boldsymbol{X}$ is a continuous random variable. (Note, strictly, Eq. (0.4) is the differential entropy, since entropy itself is infinite for continuous variables. However, considering continuous variables as continuous limits of discrete variable approximations, entropy differences and hence information remain well-defined in the continuous limit and may be consistently measured using Eq. (0.4) [16]. Moreover, this equation assumes that $\boldsymbol{X}$ has a density with respect to the Lebesgue measure $\mathrm{d}^n\boldsymbol{x}$; this assumption is upheld whenever we discuss continuous random variables.)

We write $H(\boldsymbol{X}|\boldsymbol{Y}=\boldsymbol{y})$ for the conditional entropy of $\boldsymbol{X}$ given that $\boldsymbol{Y}=\boldsymbol{y}$, and $H(\boldsymbol{X}|\boldsymbol{Y})$ for the expected conditional entropy of $\boldsymbol{X}$ given $\boldsymbol{Y}$, i.e.,

$$H(\boldsymbol{X}|\boldsymbol{Y}) =: \sum_{\boldsymbol{y}\in\Omega_{\boldsymbol{Y}}} H(\boldsymbol{X}|\boldsymbol{Y}=\boldsymbol{y})P_{\boldsymbol{Y}}(\boldsymbol{y}), \tag{0.5}$$

if $\boldsymbol{Y}$ is discrete; for continuous $\boldsymbol{Y}$ replace the summation by integration. The mutual information $I(\boldsymbol{X};\boldsymbol{Y})$ between $\boldsymbol{X}$ and $\boldsymbol{Y}$ is the average information, or reduction in uncertainty (entropy), about $\boldsymbol{X}$, knowing the outcome of $\boldsymbol{Y}$:

$$I(\boldsymbol{X};\boldsymbol{Y})=H(\boldsymbol{X})-H(\boldsymbol{X}|\boldsymbol{Y}). \tag{0.6}$$

Mutual information can also be written in the useful form

$$I(\boldsymbol{X};\boldsymbol{Y})=H(\boldsymbol{X})+H(\boldsymbol{Y})-H(\boldsymbol{X},\boldsymbol{Y}), \tag{0.7}$$

from which it follows that mutual information is symmetric in $\boldsymbol{X}$ and $\boldsymbol{Y}$ [16]. If $\boldsymbol{X}$ and $\boldsymbol{Y}$ are both Gaussian,

$$H(\boldsymbol{X})=\frac{1}{2}\log[\det\Sigma(\boldsymbol{X})]+\frac{1}{2}n\log(2\pi e), \tag{0.8}$$

$$H(\boldsymbol{X}|\boldsymbol{Y}=\boldsymbol{y})=\frac{1}{2}\log[\det\Sigma(\boldsymbol{X}|\boldsymbol{Y})]+\frac{1}{2}n\log(2\pi e),\ \forall\boldsymbol{y}\in\mathbb{R}^m, \tag{0.9}$$

$$I(\boldsymbol{X};\boldsymbol{Y})=\frac{1}{2}\log\left[\frac{\det\Sigma(\boldsymbol{X})}{\det\Sigma(\boldsymbol{X}|\boldsymbol{Y})}\right]. \tag{0.10}$$

All these quantities are straightforward to compute empirically from the empirical covariance matrices $\Sigma(\boldsymbol{X})$ and $\Sigma(\boldsymbol{X},\boldsymbol{Y})$, and the expression (0.1).

The Kullback-Leibler (KL) divergence $D_{\mathrm{KL}}(P_{\boldsymbol{X}}||P_{\boldsymbol{Y}})$ is a (non-symmetric) measure of the difference between two probability distributions $P_{\boldsymbol{X}}$ and $P_{\boldsymbol{Y}}$ (well-defined when the variables take values in the same space, $\boldsymbol{\Omega_X}=\boldsymbol{\Omega_Y}$). It is given by

$$D_{\mathrm{KL}}(P_{\boldsymbol{X}}||P_{\boldsymbol{Y}}) =: \sum_{\boldsymbol{x}\in\Omega_{\boldsymbol{X}}} P_{\boldsymbol{X}}(\boldsymbol{x})\log\left[\frac{P_{\boldsymbol{X}}(\boldsymbol{x})}{P_{\boldsymbol{Y}}(\boldsymbol{x})}\right], \tag{0.11}$$

if the variables are discrete, or

$$D_{\mathrm{KL}}(P_{\boldsymbol{X}}||P_{\boldsymbol{Y}}) =: \int_{\mathbb{R}^n} P_{\boldsymbol{X}}(\boldsymbol{x})\log\left[\frac{P_{\boldsymbol{X}}(\boldsymbol{x})}{P_{\boldsymbol{Y}}(\boldsymbol{x})}\right]\mathrm{d}^n\boldsymbol{x}, \tag{0.12}$$

if the variables are continuous.

We examine integrated information generated by systems of interconnected dynamical elements. We use the letter X to denote such a system, and the number of elements in the system is denoted by $|X|$. A partition $\mathcal{P}=\{M^1,\ldots,M^r\}$ divides the elements of X into non-overlapping, non-trivial sub-systems, $X=M^1\cup M^2\cup\cdots\cup M^r$. The state of X at time t is a $|X|$-dimensional random vector denoted by $\boldsymbol{X}_t$, with entries corresponding to states of individual elements of X. Time is discretized, so t takes integer values. We denote the set of possible states of X by S_X, and the size of this set by $|S_X|$. Analogous notation is used for the states of sub-systems of X.

A stationary system is one for which the probability density function for $\boldsymbol{X}_t$ does not change with time t. For such systems $\Sigma(\boldsymbol{X})$ denotes the stationary covariance matrix, and $\Gamma_\tau(\boldsymbol{X})$ the auto-covariance matrix with time-lag τ:

$$\Gamma_\tau(\boldsymbol{X}) =: \Sigma(\boldsymbol{X}_{t-\tau},\boldsymbol{X}_t). \tag{0.13}$$

The previous measure, Φ_{DM}

In this section we review, following Ref. [11], the most recent version of Φ within integrated information theory, using our current notation. This measure, which we call Φ_{DM} ('Φ-discrete/Markovian'), was defined for discrete, Markovian systems, i.e. systems with (i) a discrete set of possible states, and (ii) dynamics for which the current state depends only on the state at the previous time-step. After laying out the formal description of Φ_{DM}, we

briefly discuss these limitations, which motivate our new measures Φ_{E} and Φ_{AR}.

Let X be a discrete, Markovian system. Φ_{DM} compares the information generated by the whole system to information generated by its parts, when the system transitions to a particular state $\boldsymbol{X}_1=\boldsymbol{x}$ from a preceding state $\boldsymbol{X}_0$ characterized by the maximum entropy distribution for the system. This is performed by use of KL divergence to compare (i) the conditional probability distribution for the preceding state of the whole given the current state; (ii) the joint distribution for the preceding states of parts given their respective current states.

The *effective information*, $\varphi_{\mathrm{DM}}[X;\boldsymbol{x},\mathcal{P}]$, generated by X being in state $\boldsymbol{x}$, with respect to the partition $\mathcal{P}=\{M^1,\ldots,M^r\}$, is given by

$$\varphi_{\mathrm{DM}}[X;\boldsymbol{x},\mathcal{P}] =: D_{\mathrm{KL}}\left(P_{\boldsymbol{X}_0|\boldsymbol{X}_1=\boldsymbol{x}} \,||\, \prod_{k=1}^{r} P_{\boldsymbol{M}_0^k|\boldsymbol{M}_1^k=\boldsymbol{m}^k}\right). \qquad (0.14)$$

Here $\boldsymbol{m}^k$ is the state of the k^{th} sub-system of the partition when X has state $\boldsymbol{x}$.

To specify the probability distributions in (0.14), one must use Bayes' rule. For the distribution of the whole system the formula is

$$P_{\boldsymbol{X}_0|\boldsymbol{X}_1=\boldsymbol{x}}(\boldsymbol{x}') = \frac{P_{\boldsymbol{X}_1|\boldsymbol{X}_0=\boldsymbol{x}'}(\boldsymbol{x})P_{\boldsymbol{X}_0}(\boldsymbol{x}')}{P_{\boldsymbol{X}_1}(\boldsymbol{x})}. \qquad (0.15)$$

Here $P_{\boldsymbol{X}_0}(\boldsymbol{x}')$ is the maximum entropy distribution, so

$$P_{\boldsymbol{X}_0}(\boldsymbol{x}') = \frac{1}{|S_X|}, \qquad (0.16)$$

for all possible initial states $\boldsymbol{x}'\in S_X$. $P_{\boldsymbol{X}_1|\boldsymbol{X}_0=\boldsymbol{x}'}$ is the conditional probability density for the state at time $t=1$ given that the state at time $t=0$ is $\boldsymbol{x}'$. Given a generative model of the system, this distribution can be derived analytically by examining the transitions allowed by the model. In the absence of a generative model the distribution can be obtained by empirical measurement of the equivalent distribution $P_{\boldsymbol{X}_t|\boldsymbol{X}_{t-1}=\boldsymbol{x}'}$. Note that in neither case is perturbation of the system required, although in the latter case the system must visit all possible states multiple times to allow reasonable estimation of $P_{\boldsymbol{X}_t|\boldsymbol{X}_{t-1}=\boldsymbol{x}'}$. Finally, the denominator $P_{\boldsymbol{X}_1}(\boldsymbol{x})$ is computed from

$$P_{\boldsymbol{X}_1}(\boldsymbol{x}) = \sum_{\boldsymbol{\xi}\in S_X} P_{\boldsymbol{X}_1|\boldsymbol{X}_0=\boldsymbol{\xi}}(\boldsymbol{x})P_{\boldsymbol{X}_0}(\boldsymbol{\xi}). \qquad (0.17)$$

For a part M the analogous Bayes' rule formula is

$$P_{\boldsymbol{M}_0|\boldsymbol{M}_1=\boldsymbol{m}}(\boldsymbol{m}') = \frac{P_{\boldsymbol{M}_1|\boldsymbol{M}_0=\boldsymbol{m}'}(\boldsymbol{m})P_{\boldsymbol{M}_0}(\boldsymbol{m}')}{P_{\boldsymbol{M}_1}(\boldsymbol{m})}. \qquad (0.18)$$

Here $P_{\boldsymbol{M}_0}$ is the maximum entropy distribution on S_M. To compute the conditional probability distribution $P_{\boldsymbol{M}_1|\boldsymbol{M}_0=\boldsymbol{m}'}$ for the state at time 1 given the state at time 0 it is necessary to average over states external M. Let N denote the complement of M within X, so $\boldsymbol{X}_t=(\boldsymbol{M}_t,\boldsymbol{N}_t)^{\mathrm{T}}$. Then we have

$$P_{\boldsymbol{M}_1|\boldsymbol{M}_0=\boldsymbol{m}',\boldsymbol{N}_0=\boldsymbol{n}'}(\boldsymbol{m}) = \sum_{\boldsymbol{n}\in S_N} P_{\boldsymbol{X}_1|\boldsymbol{X}_0=(\boldsymbol{m}',\boldsymbol{n}')}(\boldsymbol{m},\boldsymbol{n}), \qquad (0.19)$$

$$P_{\boldsymbol{M}_1|\boldsymbol{M}_0=\boldsymbol{m}'}(\boldsymbol{m}) = \sum_{\boldsymbol{n}'\in S_N} P_{\boldsymbol{M}_1|\boldsymbol{M}_0=\boldsymbol{m}',\boldsymbol{N}_0=\boldsymbol{n}'}(\boldsymbol{m})P_{\boldsymbol{N}_0}(\boldsymbol{n}'). \qquad (0.20)$$

(Note that in Ref. [11] φ_{DM} is instead computed using a perturbed version of the sub-system M, for which the joint distribution of the noise in all the afferent connections ('wires') to M is taken to be maximum entropy. Here we instead assign the maximum entropy distribution to *states* external to the sub-system. By doing so, we eliminate the step of perturbing sub-systems, and need only perturb the whole system once, namely to impose the maximum entropy distribution as the initial state of the whole system. This choice enables simpler notation and description and does not affect the qualitative behavior of the measure [11].) Finally, $P_{\boldsymbol{M}_1}(\boldsymbol{m})$ is given by

$$P_{\boldsymbol{M}_1}(\boldsymbol{m}) = \sum_{\boldsymbol{\mu}\in S_M} P_{\boldsymbol{M}_1|\boldsymbol{M}_0=\boldsymbol{\mu}}(\boldsymbol{m})P_{\boldsymbol{M}_0}(\boldsymbol{\mu}). \qquad (0.21)$$

Given the probability distributions $P_{\boldsymbol{X}_0|\boldsymbol{X}_1=\boldsymbol{x}}$ and $P_{\boldsymbol{M}_0^k|\boldsymbol{M}_1^k=\boldsymbol{m}^k}$, $k=1,\ldots,r$, the effective information is computed using the formula (0.11) for the KL divergence.

The *integrated information* is defined as the effective information with respect to the minimum information partition (MIP). The MIP, $\mathcal{P}^{\mathrm{MIP}}(\boldsymbol{x})$, is defined as the partition that minimizes the effective information when it is normalized by

$$K_{\mathrm{M}}(\{M^1,\ldots,M^r\}) =: (r-1)\cdot\min_k\left[H(\boldsymbol{M}_0^k)\right]. \qquad (0.22)$$

Normalization is necessary because sub-systems that are almost as large as the whole system typically generate almost as much information as the whole system. Therefore, without normalization, most systems would have a highly imbalanced MIP, (e.g., one element versus the remainder of the system) and a trivially small value for integrated information. The normalization K_{M} ensures that integrated information is specified using a partition defined using a weighted minimization of the effective information, with a bias towards partitions into sub-systems of roughly equal size. We will discuss the importance of normalization further in the section 'Φ_{E} for Markovian Gaussian systems'. Thus, $\mathcal{P}^{\mathrm{MIP}}(\boldsymbol{x})$ is given by

$$\mathcal{P}^{\mathrm{MIP}}(\boldsymbol{x}) =: \arg_{\mathcal{P}}\min\left\{\frac{\varphi_{\mathrm{DM}}[X;\boldsymbol{x},\mathcal{P}]}{K_{\mathrm{M}}(\mathcal{P})}\right\}. \qquad (0.23)$$

Given the MIP, the integrated information $\Phi_{\mathrm{DM}}(X;\boldsymbol{x})$ generated by the system X entering state $\boldsymbol{x}$ is simply the *non-normalized* effective information with respect to the MIP,

$$\Phi_{\mathrm{DM}}[X;\boldsymbol{x}] =: \varphi_{\mathrm{DM}}\left[X;\boldsymbol{x},\mathcal{P}^{\mathrm{MIP}}(\boldsymbol{x})\right]. \qquad (0.24)$$

Importantly, the value of $\Phi_{\mathrm{DM}}[X;\boldsymbol{x}]$ is furnished by the non-normalized effective information because it is supposed to represent a physically meaningful property of the system in the corresponding 'integrated information theory' [14].

For a state-independent alternative to Φ_{DM}, one can replace the effective information with its expectation with respect to the current state $\boldsymbol{x}$, and define the *expected integrated information*, $\bar{\Phi}_{\mathrm{DM}}$, as the expected effective information across the partition that minimizes the normalized expected effective information [11]. The expected effective information, $\bar{\varphi}_{\mathrm{DM}}$, is given by [11]

$$\bar{\varphi}_{\mathrm{DM}}\left[X;\left\{M^1,\ldots,M^r\right\}\right]\equiv\sum_{k=1}^{r}H\left(\boldsymbol{M}_0^k|\boldsymbol{M}_1^k\right)-H(\boldsymbol{X}_0|\boldsymbol{X}_1)\,, \quad (0.25)$$

or equivalently

$$\bar{\varphi}_{\mathrm{DM}}\left[X;\left\{M^1,\ldots,M^r\right\}\right]\equiv I(\boldsymbol{X}_0;\boldsymbol{X}_1)-\sum_{k=1}^{r}I\left(\boldsymbol{M}_0^k;\boldsymbol{M}_1^k\right). \quad (0.26)$$

Note that the second expression (0.26), but not the first (0.25), requires that $\boldsymbol{X}_0$ have the maximum entropy distribution [11]. To derive (0.26) from (0.25), one uses that the maximum entropy distribution is uniform, so that

$$H(\boldsymbol{X}_0)=\sum_k H\left(\boldsymbol{M}_0^k\right). \quad (0.27)$$

This ensures that one can add $H(\boldsymbol{X}_0)$ to the second term on the RHS of (0.25) and subtract $\sum_k H\left(\boldsymbol{M}_0^k\right)$ from the first term, and then use Eq. (0.6) to obtain the expression (0.26).

We emphasize that Φ_{DM} was defined only for systems that are both discrete and Markovian. The measure can not be applied to continuous systems (except those with a compact i.e. closed and bounded set of states) because there is no uniquely defined maximum entropy distribution for a continuous random variable defined on the real number line [16]. (In fact, the measure is also not applicable to discrete systems with an infinite set of states.) Φ_{DM} can only be applied to Markovian systems because for a non-Markovian system it is not clear how to impose the maximum entropy distribution as an initial condition, implying that the conditional probability distribution $P_{\boldsymbol{X}_1|\boldsymbol{X}_0=\boldsymbol{x}'}$ cannot be uniquely specified by any generative model. For instance three alternatives are (i) to make all past states independent and maximum entropy; (ii) to set all past states to zero except the most recent; (iii) to just set one past state to maximum entropy and obtain the distribution for other past states from the generative model. There is no immediately apparent way to choose among these alternatives. Taken together, these limitations are important because complex (e.g. neural) systems are typically non-Markovian, and neural signals are often recorded as continuous variables. In 'Methods' we describe an extension to Φ_{DM} that renders it well-defined for stationary continuous, but still Markovian, systems by choosing a maximum entropy distribution based on the stationary variances of the states of individual elements. This enables us to compare Φ_{DM} with our new measure Φ_{E} for some example cases.

The new measure, Φ_{E}

The general case. In this section we introduce a new measure of integrated information, Φ_{E}, constructed analogously to Φ_{DM}, but with modifications to broaden its applicability, both in theory and in practice. Φ_{E} is designed for stochastic stationary systems, for which it provides a single time- and state-independent value (given a timescale of measurement, discussed below). The measure is particularly easy to apply to stationary Gaussian systems, either from time-series data or from a generative model.

The key modification is that rather than measuring information generated by transitions from a hypothetical maximum entropy past state, Φ_{E} instead utilizes the actual distribution of the past state; hence the name Φ_{E}, 'Φ-empirical'. This ensures that the measure does not suffer from the in-principle restrictions that pertain to Φ_{DM}, and can be applied to both discrete and continuous systems with either Markovian or non-Markovian dynamics. (More specifically, Φ_{E} will be well-defined as long as the states $\boldsymbol{X}_t$ of the system are either discrete or have continuous probability densities with respect to a Lebesgue measure $\mathrm{d}^n\boldsymbol{x}$.) A second difference is that, in order to be state-independent, Φ_{E} is based on the *average* information generated by the current state about the past state, as opposed to information generated by a particular current state. Finally, Φ_{E} is defined so as to enable a choice of timescale (indicated by τ) over which integrated information is measured. Thus $\Phi_{\mathrm{E}}[X;\tau]$ is the integrated information generated by the current state of the system about the state τ time-steps in the past.

We now define Φ_{E} for a stochastic system with stationary dynamics. As for Φ_{DM}, Φ_{E} is defined via 'effective information'. For the new measure we define the effective information generated by the current state $\boldsymbol{X}_t$ about the state τ time-steps ago, with respect to bipartition $\mathcal{B}=\left\{M^1,M^2\right\}$, to be the mutual information generated by the whole system minus the sum of the mutual information generated by the parts within the bipartition. Thus

$$\varphi[X;\tau,\mathcal{B}]=:I(\boldsymbol{X}_{t-\tau};\boldsymbol{X}_t)-\sum_{k=1}^{2}I\left(\boldsymbol{M}_{t-\tau}^k;\boldsymbol{M}_t^k\right). \quad (0.28)$$

The integrated information $\Phi_{\mathrm{E}}[X;\tau]$ is then the non-normalized effective information with respect to the minimum information bipartition (MIB),

$$\Phi_{\mathrm{E}}[X;\tau]=:\varphi\left[X;\tau,\mathcal{B}^{\mathrm{MIB}}(X;\tau)\right], \quad (0.29)$$

where

$$\mathcal{B}^{\mathrm{MIB}}(X;\tau)=:\arg_{\mathcal{B}}\min\left\{\frac{\varphi[X;\tau,\mathcal{B}]}{K(\mathcal{B})}\right\}, \quad (0.30)$$

and

$$K\left(\left\{M^1,M^2\right\}\right)=:\min\left[H\left(\boldsymbol{M}_t^1\right),H\left(\boldsymbol{M}_t^2\right)\right]. \quad (0.31)$$

Φ_{E} can either be computed analytically from a generative model, or estimated numerically from time-series data. In either case, one must first obtain estimates of the probability distributions for the states $\boldsymbol{X}_{t-\tau}$ and $\boldsymbol{X}_t$, and their joint distribution $P_{(\boldsymbol{X}_{t-\tau},\boldsymbol{X}_t)^{\mathrm{T}}}$, as well as the corresponding distributions for all sub-systems. Then, given these distributions, the corresponding entropies can be computed using Eq. (0.3), for a system with discrete states, or Eq. (0.4) for a system with continuous states. Having obtained these entropies, Eq. (0.7) can be used to obtain the mutual information $I(\boldsymbol{X}_{t-\tau};\boldsymbol{X}_t)$ between the past and current state of the system, and likewise for all sub-systems. Given these quantities, Φ_{E} can then be obtained directly from Eqs. (0.28)–(0.31).

For numerical computation, the required probability distributions can in principle be obtained directly from data, although in practice it may be difficult to obtain sufficient data to enable accurate estimation of all the relevant entropies. As we explain in the section 'Computing Φ_{E} empirically under Gaussian assumptions', this difficulty can be readily overcome if states are Gaussian distributed.

For analytic computation of Φ_{E} given a generative model, we note that the probability distributions for $\boldsymbol{X}_{t-\tau}$ and $\boldsymbol{X}_t$ *individually* are both simply equal to the stationary distribution for the state of the system. Obtaining the joint distribution for $\boldsymbol{X}_{t-\tau}$ and $\boldsymbol{X}_t$

together will depend on the details of the generative model. Once again the situation is much easier in practice for Gaussian systems, in which case only the covariance matrix of each probability distribution is needed (see equation (0.8)). As we show in the section 'Computing Φ_E analytically for a Gaussian system', these matrices can be derived easily from a generative model expressed as a generalized connectivity matrix, assuming Gaussian dynamics.

A few further remarks about Φ_E are worth making. First, that Φ_E remains well-defined as a time-dependent quantity for non-stationary stochastic systems; we focus on the stationary case for simplicity, and because of our interest in empirical measurement of Φ_E via sampling from time-series data. Second, unlike Φ_{DM}, Φ_E is not defined for deterministic systems. This is because it does not incorporate a perturbation through which to introduce probabilities into a deterministic system. Third, we restrict attention to bipartitions for computational efficiency. This is standard practice for computing Φ_{DM} [11,14]. Extension to general partitions is trivial, albeit computationally expensive. Finally, since mutual information is symmetric in its two arguments (0.7), effective information as given by (0.28) can alternatively be read in terms of information generated by the past state $\boldsymbol{X}_{t-\tau}$ about the current state $\boldsymbol{X}_t$.

Our definition (0.28) for the effective information, φ, is based on the expression (0.26) for the *expected* effective information, $\bar{\varphi}_{DM}$ in the construction of Φ_{DM}. A viable alternative would be to instead use

$$\tilde{\varphi}\left[X;\tau,\left\{M^1,M^2\right\}\right] = :\sum_{k=1}^{2} H\left(\boldsymbol{M}_{t-\tau}^k|\boldsymbol{M}_t^k\right) - H(\boldsymbol{X}_{t-\tau}|\boldsymbol{X}_t), \quad (0.32)$$

the analogue of (0.25). This quantity has previously been defined in Ref. [22] as 'stochastic interaction. It is the average KL divergence between (i) the past of the whole given the present of the whole, and (ii) the product of this for parts [11]. Replacing φ with $\tilde{\varphi}$ in the definition of Φ_E leads to a second measure $\tilde{\Phi}_E$. In general, $\tilde{\varphi}$ will not be exactly equal to φ. (Equality of their analogues for Φ_{DM} relies on the past state being maximum entropy, see section 'The previous measure, Φ_{DM}'.) However, we show in Table 1 that $\tilde{\Phi}_E$ behaves very similarly to Φ_E for the examples we consider in this paper. We choose to focus on Φ_E because it explicitly operationalizes the concept of 'information generated by the whole minus the sum of information generated by the parts' (0.28).

In summary, we have defined a new measure of integrated information Φ_E that is broadly well-defined, and which is easy to measure under Gaussian dynamics, either from time-series data or given a generative model (see below). In contrast, the previous measure Φ_{DM} is only defined for discrete, Markovian systems. As a consequence, Φ_E but not Φ_{DM} is applicable to realistic continuous non-Markovian stochastic models of neural systems.

Computing Φ_E empirically under Gaussian assumptions. Under Gaussian assumptions, equation (0.10) furnishes an expression for Φ_E simply in terms of covariance matrices, enabling straightforward empirical computation. The effective information is given by

$$\varphi\left[X;\tau,\left\{M^1,M^2\right\}\right] = \frac{1}{2}\log\left\{\frac{\det\Sigma(\mathrm{X})}{\det\Sigma(\mathrm{X}_{t-\tau}|\mathrm{X}_t)}\right\} - \sum_{k=1}^{2}\frac{1}{2}\log\left\{\frac{\det\Sigma\left(\mathrm{M}^k\right)}{\det\Sigma\left(\mathrm{M}_{t-\tau}^k|\mathrm{M}_t^k\right)}\right\}, \quad (0.33)$$

and the normalization factor K by

$$K\left(\left\{M^1,M^2\right\}\right) = \frac{1}{2}\log\min_k\left\{(2\pi e)^{\left|M^k\right|}\det\Sigma\left(\boldsymbol{M}^k\right)\right\}. \quad (0.34)$$

In practice, the procedure for computing Φ_E is as follows. First one obtains empirically the covariance matrices $\Sigma(\boldsymbol{X})$, $\Sigma(\boldsymbol{X}_{t-\tau},\boldsymbol{X}_t)$ and analogues for all sub-systems. Then one uses Eq. (0.1) to obtain the partial covariance $\Sigma(\boldsymbol{X}_{t-\tau}|\boldsymbol{X}_t)$ and its sub-system analogues. Given these quantities, equations (0.33) and (0.34) furnish estimates for the effective information and normalized effective information with respect to any given bipartition. These estimates allow identification of the MIB and Φ_E, via equations (0.29) and (0.30).

Computing Φ_E analytically for a Gaussian system. In this section we describe analytical computation of Φ_E for Gaussian systems, assuming that the generative model is known. We first recognize that a generative model for a Gaussian stationary system is always equivalent to an $\mathrm{MVAR}(p)$ (multivariate auto-regressive) process [18]

$$\boldsymbol{X}_t = A_1\cdot\boldsymbol{X}_{t-1} + A_2\cdot\boldsymbol{X}_{t-2} + \cdots + A_p\cdot\boldsymbol{X}_{t-p} + \boldsymbol{E}_t, \quad (0.35)$$

where the A_i, $i=1,\ldots,p$, can be understood as generalized connectivity matrices acting at different time-lags, and $\boldsymbol{E}_t$ is a stationary multivariate Gaussian 'white noise' source with zero mean and vanishing auto-covariance function, $\Gamma_\tau(\boldsymbol{E})=0$, $\tau\neq 0$. (Technically, there also exists the case $p=\infty$, but we do not consider this here, because in practical application there will always be an optimal range of finite p for model fitting.) Below, we show how to calculate Φ_E for an MVAR(1) system at timescale $\tau=1$. Extension to the general p, general τ case is given in the 'Methods' section. Consider the generative model

$$\boldsymbol{X}_t = A\cdot\boldsymbol{X}_{t-1} + \boldsymbol{E}_t. \quad (0.36)$$

Taking the covariance of both sides of (0.36) gives

$$\Sigma(\boldsymbol{X}) = A\Sigma(\boldsymbol{X})A^{\mathrm{T}} + \Sigma(\boldsymbol{E}). \quad (0.37)$$

Noticing that this equation is the discrete-time Lyapunov equation, $\Sigma(\boldsymbol{X})$ can be computed numerically, given A, for example, in Matlab via use of the 'dlyap' command. To compute the partial covariance $\Sigma(\boldsymbol{X}_{t-1}|\boldsymbol{X}_t)$ we need the single time-step auto-covariance matrix

$$\Gamma_1(\boldsymbol{X}) \equiv \Sigma(\boldsymbol{X}_{t-1},\boldsymbol{X}_t) = \langle\boldsymbol{X}_{t-1}(A\boldsymbol{X}_{t-1}+\boldsymbol{E}_t)^{\mathrm{T}}\rangle = \Sigma(\boldsymbol{X})A^{\mathrm{T}}. \quad (0.38)$$

We can then use equation (0.1) to obtain the partial covariance as

$$\Sigma(\boldsymbol{X}_{t-1}|\boldsymbol{X}_t) = \Sigma(\boldsymbol{X}) - \Gamma_1(\boldsymbol{X})\Sigma(\boldsymbol{X})^{-1}\Gamma_1(\boldsymbol{X})^{\mathrm{T}}. \quad (0.39)$$

Having values for $\Sigma(\boldsymbol{X})$ and $\Sigma(\boldsymbol{X}_{t-1}|\boldsymbol{X}_t)$ allows calculation of the first term in the RHS of (0.33). Calculation of the second term, and of the normalization factor, requires consideration of sub-systems. For a sub-system M, we consider the bipartition $\{M,N\}$, and the block decomposition of vectors and matrices according to $\boldsymbol{X}_t=(\boldsymbol{M}_t,\boldsymbol{N}_t)^{\mathrm{T}}$. The matrices $\Sigma(\boldsymbol{X})$ and $\Gamma_1(\boldsymbol{X})$ can then be written in the form

Table 1. Integrated information computed in various ways for the networks shown in Figs. 1 and 2.

Network	(i) Φ_E	(ii) Φ_E (3000 data)	(iii) Φ_E (10,000 data)	(iv) $\tilde{\Phi}_E$	(v) $\bar{\Phi}_{DM}$	(vi) Φ_{AR} (3000 data)
1(a)	0.0323	0.037 ± 0.004 (10)	0.034 ± 0.003 (10)	0.0323	0.0323	0.038 ± 0.002 (9) 0.031 (1)
1(b)	0.0645	0.063 ± 0.004 (10)	0.063 ± 0.003 (10)	0.0645	0.0645	0.061 ± 0.005 (10)
1(c)	0.1283	0.122 ± 0.008 (10)	0.124 ± 0.003 (10)	0.1387	0.1313	0.125 ± 0.004 (10)
1(d)	0.0795	0.072 ± 0.006 (10)	0.075 ± 0.004 (10)	0.0894	0.0755	0.073 ± 0.006 (10)
1(e)	0.1285	0.136 ± 0.012 (10)	0.129 ± 0.002 (10)	0.1376	0.1303	0.135 ± 0.013 (10)
1(f)	0.1294	0.132 ± 0.008 (10)	0.133 ± 0.004 (10)	0.1383	0.1307	0.131 ± 0.012 (10)
1(g)	0.1266	0.128 ± 0.010 (6) 0.093 ± 0.005 (4)	0.098 ± 0.004 (6) 0.129 ± 0.006 (4)	0.1362	0.1288	0.129 ± 0.013 (7) 0.101 ± 0.009 (3)
2(a)	0.2502	0.244 ± 0.010 (10)	0.246 ± 0.004 (9) 0.125 (1)	0.2652	0.1254	0.245 ± 0.009 (9) 0.128 (1)
2(b)	0.2965	0.291 ± 0.013 (10)	0.2902 ± 0.004 (10)	0.3012	0.2647	0.287 ± 0.010 (7) 0.142 ± 0.005 (3)

Methods of computation are (i) Φ_E computed analytically; (ii) Φ_E computed numerically from 10 trials of 3000 data points each; (iii) Φ_E computed numerically from 10 trials of 10,000 data points each; (iv) $\tilde{\Phi}_E$ computed analytically; (v) (extended) $\bar{\Phi}_{DM}$ computed analytically, and (vi) Φ_{AR} computed numerically from 10 trials of 3000 data points each, with the noise exponentially distributed. For numerical computation, means and standard deviations are given; the number of trials resulting in each value is given in parentheses. In all cases $\tau = 1$.

doi:10.1371/journal.pcbi.1001052.t001

$$\Sigma(X) = \begin{pmatrix} \Sigma(X)_{MM} & \Sigma(X)_{MN} \\ \Sigma(X)_{NM} & \Sigma(X)_{NN} \end{pmatrix}, \qquad \Gamma_1(\mathrm{X}) = \begin{pmatrix} \Gamma_1(\mathrm{X})_{MM} & \Gamma_1(\mathrm{X})_{MN} \\ \Gamma_1(\mathrm{X})_{NM} & \Gamma_1(\mathrm{X})_{NN} \end{pmatrix}, \tag{0.40}$$

and we can use that

$$\Sigma(\boldsymbol{M}) = \Sigma(\boldsymbol{X}), \quad \Gamma_1(\boldsymbol{M}) = \Gamma_1(\boldsymbol{X})_{MM}. \tag{0.41}$$

Then, again from (0.1), the partial covariance is given by

$$\Sigma(\mathrm{M}_{t-1}|\mathrm{M}_t) = \Sigma(X)_{MM} - \Gamma_1(\mathrm{X})_{MM}\left[\Sigma(X)_{MM}\right]^{-1}\Gamma_1(X)^{\mathrm{T}}_{MM}. \tag{0.42}$$

Equations (0.37)–(0.42) together furnish the covariance matrices needed to compute the effective information and normalized effective information from the formulae (0.33) and (0.34) valid for Gaussian systems. Finally, the MIB and Φ_E are obtained from Eqs. (0.29) and (0.30).

Φ_E for Markovian Gaussian systems

Canonical examples. We present results from computing Φ_E, for timescale $\tau = 1$, for some example Markovian Gaussian systems. Results are given for analytical computation given the generative model, and for numerical computation given simulated time-series data. The example systems are characterized by the MVAR(1) dynamics

$$\boldsymbol{X}_t = A \cdot \boldsymbol{X}_{t-1} + \boldsymbol{E}_t, \tag{0.43}$$

where $\boldsymbol{X}_t$ contains 8 variables, A is the connectivity matrix, and each component of $\boldsymbol{E}_t$ is an independent Gaussian random variable of mean 0 and variance 1. We considered seven systems, with connectivity as shown in Fig. 1(a)–(g); we refer to these systems '1(a)', '1(b)', and so on. The corresponding values of Φ_E are given in Fig. 1(h) and Table 1. For analytic computation, we performed the procedure described in the section 'Computing Φ_E analytically for a Gaussian system'. For simulated measurements, we first obtained time-series data from equation (0.43), and then computed Φ_E using the recipe described in the section 'Computing Φ_E empirically under Gaussian assumptions'. To examine numerical stability of simulation measurements, we performed 10 trials for each network with 3000 post-equilibrium data points and a separate set of 10 trials with 10,000 post-equilibrium data points.

For all systems, except 1(g) (which we discuss below), the analytically derived (true) value of Φ_E lay within ≈ 1 standard deviation of the mean value obtained via the simulations, both for 3000 and 10,000 data points (see Fig. 1(h) and Table 1). This correspondence confirms the consistency of the numerical and analytical approaches described above.

The values of integrated information mostly correspond with expectations. For example, a ring of reciprocal connections (1(c)) integrates approximately twice as much information as a ring of unidirectional connections (1(b)), which itself integrates approximately twice as much information as a (non-closed) chain of unidirectional connections (1(a)). Also as expected, the homogenous system 1(d) has a low Φ_E value. Perhaps in contrast to expectations, adding sparse long-range 'short-cut' connections to a reciprocal ring (1(e)–1(g)), in the style of a so-called 'small world' network [23,24,25], does not increase Φ_E (compare with network 1(c)).

For values of Φ_E to be meaningful it is essential that they are stable with respect to numerical computation. To assess numerical stability, we calculated the coefficient of variation (the standard deviation divided by mean) across each set of 10 trials. For all networks other than 1(g), and for trial sets of both 3,000 and 10,000 data points, the coefficient of variation was less than 0.11, confirming that empirical calculation of Φ_E from time-series data is stable for these networks.

Network 1(g) exhibited instability when measuring Φ_E from simulation. As shown in Fig. 1(h), the corresponding values of Φ_E fell close to one of two values, one of which was the true

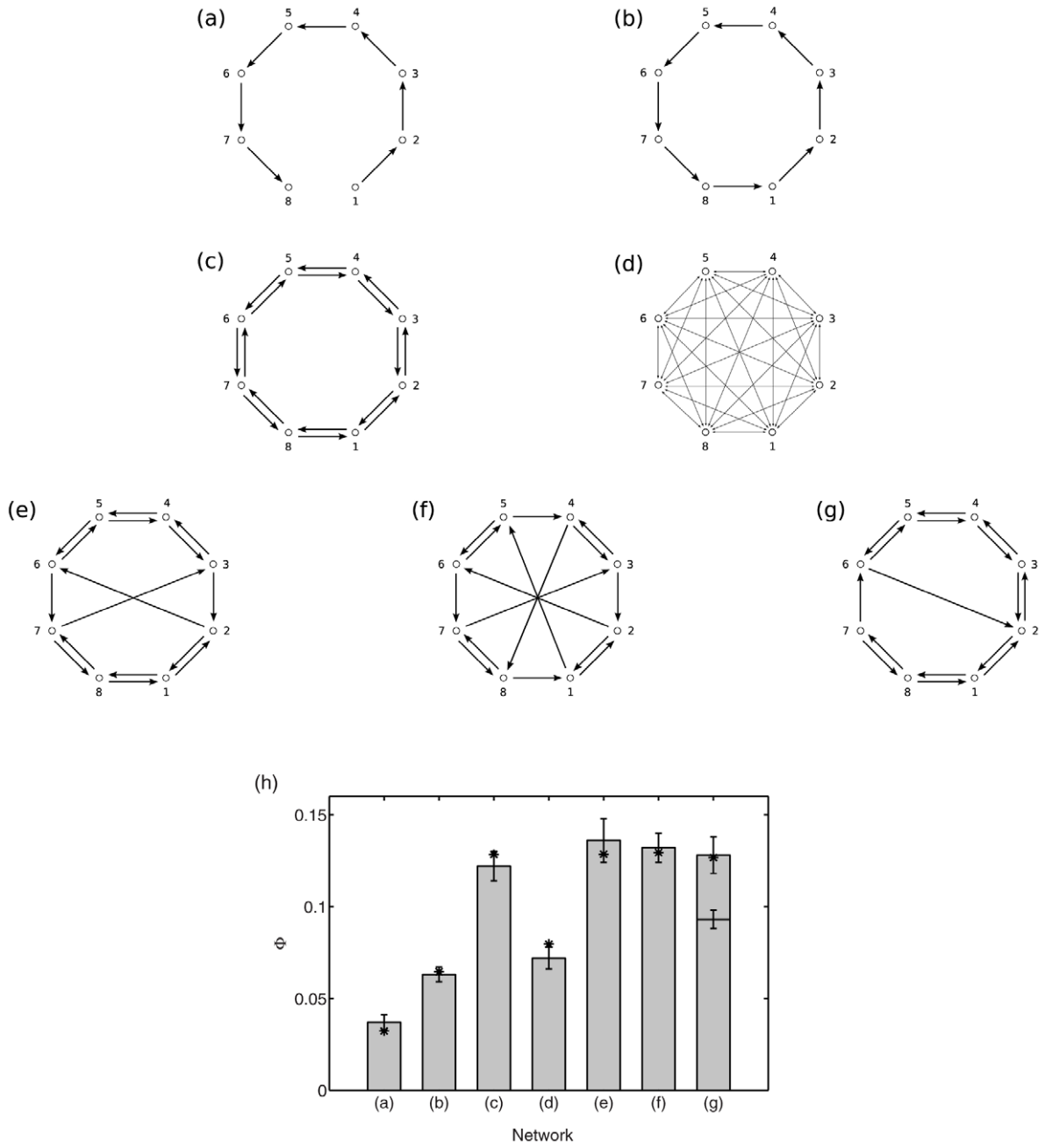

Figure 1. Integrated information in Markovian Gaussian systems. (a)–(g) Connectivity diagrams for seven systems as specified by the corresponding connectivity matrices A. Arrow widths reflect connection strengths: for (a)–(c) and (e)–(g), all connection strengths are 0.25; for system (d) each connection strength is 1/14, thus the total afferent connection to each element is 0.5. (h) Integrated information, as measured by Φ_E ($\tau = 1$) for each of the systems (a)–(g), via simulated data (bars) and analytically via the generative model (asterisks). For simulated data, 10 trials were performed, with each trial generating 3000 data points. Bars show mean values; error bars show plus/minus one standard deviation. For system 1(g), sizes of sub-systems in the MIB varied across trials, falling into two distinct groups which are shown separately (the top bar reflects a group of 6 trials; the bottom bar, 4 trials).
doi:10.1371/journal.pcbi.1001052.g001

(analytically derived) value. For simulations of 3,000 data points 6/10 trials produced Φ_E estimates close to the true value; for 10,000 data points 4/10 trials provided such estimates. This instability arises from the use of *normalized* effective information (φ) in identifying the MIB, but *non-normalized* φ in specifying the corresponding value of Φ_E. Given finite data, estimates of φ cannot be guaranteed to be accurate. As a result, inter-trial variation in measuring Φ_E from data can arise when (i) there are two (or more) partitions with similar values of normalized φ close to the true minimum (used to identify the MIB), and (ii) these partitions have substantially different values for non-normalized φ. The latter condition will typically hold when partitions with similar normalized φ have significantly different sub-system sizes (see the section 'The previous measure, Φ_{DM}'). Network 1(g) illustrates this difficulty. For this network, the true MIB is the bipartition $\{\{1,6,7,8\},\{2,3,4,5\}\}$, for which the normalized φ is 0.0213.

However, there is an uneven bipartition, {{1,2,3,4,5},{6,7,8}} for which the normalized φ is 0.0218, i.e., very similar to the value of φ for the true MIB. However, the non-normalized φ for the MIB (i.e. Φ_E) is 0.1266, whereas the value for the uneven bipartition is 0.0966. Fig. 1(h) and Table 1 show that empirical measurements of Φ_E cluster around these two values.

One may consider that this problem of instability could be avoided by using non-normalized φ to identify the MIB. However, as discussed in the section 'The previous measure, Φ_{DM}', in this case Φ_E would always be trivially small because, for any non-trivial system X, a bipartition of the form $\{\{1\},\{2,3,\ldots,|X|\}\}$ would generate almost as much information as the whole system. A second solution would be to specify Φ_E in terms of normalized φ. However, in this case the meaning of Φ_E would be substantially altered inasmuch as it could no longer be considered a measure of the quantity of information generated (or integrated) by a system.

Optimization of networks for generating high Φ_E. To examine whether network structures other than reciprocally connected rings could generate high levels of Φ_E, we performed numerical optimizations using a genetic algorithm (GA). Specifically, we used Φ_E ($\tau = 1$) as an objective function for evolving populations of networks with dynamics governed by MVAR(1) processes (see Eq. (0.43)). We performed two sets of optimizations under different constraints on the connectivity matrix A. In the first set, all connection strengths were fixed ('fixed' condition; two afferents per element each with strength 0.25). In the second set, connection strengths were allowed to vary ('vary' condition; total afferent to each element equal to 0.5, all afferents to a given element equal and positive). Each condition consisted of 20 separate GAs, each with 30 randomly initialized networks in the population; (in the 'vary' condition networks were initialized with elements having on average 2 afferent connections). Each GA ran for 200 generations, allowing fitness to asymptote. Within each generation, the fitness of each network was determined by analytical computation of Φ_E; networks were then ranked by fitness and a new population was formed by rank-based selection and mutation. In the 'fixed' condition, each network was mutated by rearranging 2 connections; in the 'vary' condition each network was mutated by (with equal probability) adding, removing, or swapping 2 connections, followed by renormalization of total afference to each element to 0.5.

The results of the optimizations are shown in Fig. 2 and Table 1. Network 2(a) is the fittest (highest Φ_E) across all 20 GAs in the 'fixed' condition; this network topology was discovered by 6 out of the 20 GAs in this condition. The network has $\Phi_E = 0.2502$, approximately twice the value of the reciprocal ring networks shown in Fig. 1. Network 2(b) is the fittest across all 20 GAs in the 'vary' condition, exhibiting $\Phi_E = 0.2965$, i.e., substantially higher again. This particular topology was discovered by only 2/20 GAs, perhaps due to the larger search-space in this condition. It is noteworthy that both of these 'fittest' networks show highly heterogeneous connectivity patterns, consistent with the intuition that integrated information is characterized by the coexistence of differentiated and integrated dynamics.

The observation that the fittest network found in each condition was only reached by a minority of GAs suggests that the Φ_E landscape across MVAR(1) systems has local maxima and may exhibit ruggedness and discontinuities. To characterize this landscape, we first plotted the distribution of fitness values across all networks in the final populations from GAs that yielded the (fittest) networks 2(a) and 2(b). Figs. 3(a,b) show that in both cases the modal value of Φ_E was substantially less than the maximum value, indicating a lack of convergence suggestive of local maxima and/or ruggedness [26]. We next examined the sensitivity of Φ_E to single mutations. Figs. 3(c) and 3(d) show the percentage decrease in Φ_E following 200 separate mutations of networks 2(a) and 2(b) respectively (the corresponding mutation type was used in each case, i.e., 'fixed' for 2(a) and 'vary' for 2(b)). For network 2(a), post-mutation fitness decreases cluster in the range 10–20%, with a few instances of $\approx 60\%$. For network 2(b), more than 20% of mutations resulted in a fitness decrease of 50% or more. Together, these observations show that the value of Φ_E generated by a network is highly sensitive to small changes in topology and connection strength, further pointing to the ruggedness of the Φ_E landscape.

The instability arising from using normalized effective information to find the MIB, (see 'Canonical examples'), suggests that there may be discontinuities, as well as ruggedness, in the Φ_E

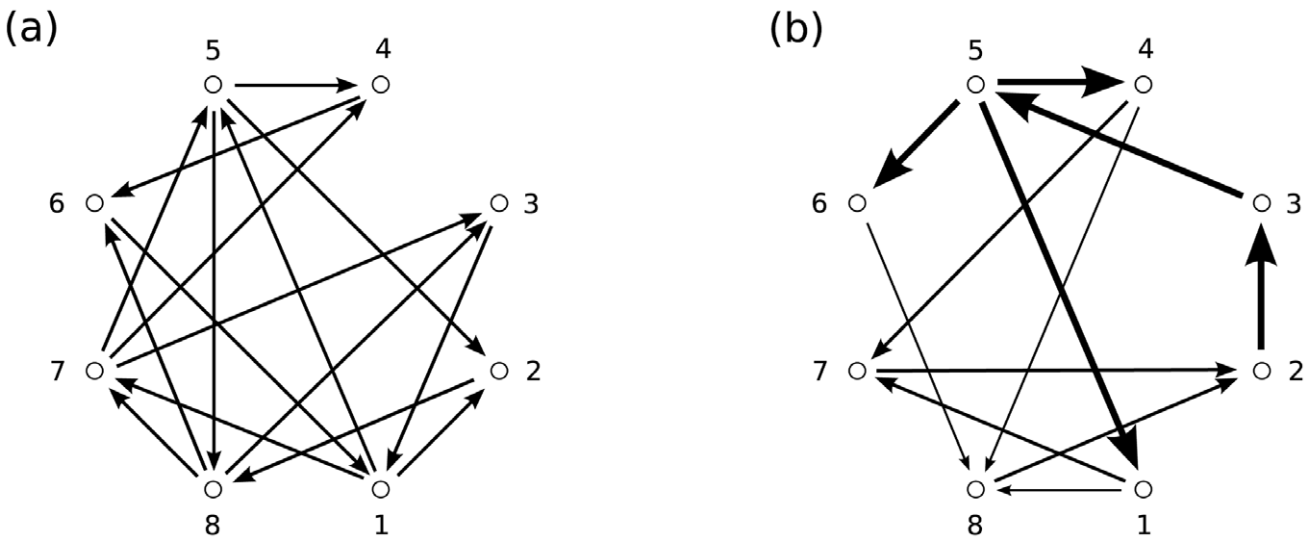

Figure 2. Networks optimized for high integrated information. (a) Optimal network for 2 afferents of 0.25 to each node. This has $\Phi_E = 0.2502$. (b) Optimal network for total afferent of 0.5 to each node, and all connections to a given node equal. This has $\Phi_E = 0.2965$.
doi:10.1371/journal.pcbi.1001052.g002

landscape. We were able to confirm the existence of such discontinuities by incrementally perturbing a specific connection in the example network 2(a). The MIB for this network is the bipartition {{1,4,5,6},{2,3,7,8}}, for which the normalized effective information is 0.0421. However, there is an uneven bipartition, {{1,2,3,5,7,8},{4,6}} with the very similar normalized effective information of 0.0424. We incrementally weakened the connection between the two sub-systems in this uneven bipartition, finding that there is a discontinuous change in Φ_E at the point at which the uneven bipartition becomes the MIB (see Fig. 3(e)).

Comparison with Φ_{DM}, $\tilde{\Phi}_E$, and full table of MVAR(1) results. It is instructive to compare results obtained using Φ_E with those obtained from the version of $\bar{\Phi}_{DM}$ extended to apply to stationary continuous (but still Markovian) systems (see sections 'The previous measure, Φ_{DM}' and 'Methods'). Table 1 shows (extended) $\bar{\Phi}_{DM}$ values for the various networks discussed above, as well as the corresponding Φ_E values. For networks 1(a) and 1(b) the two measures are exactly equivalent, which is explained by the stationary and maximum entropy distributions coinciding. For the remaining networks, (except network 2(a), discussed below), the two measures remain very similar, confirming Φ_E as a valid and useful measure of integrated information.

The network 2(a) has a value for Φ_E that is approximately double that of the corresponding $\bar{\Phi}_{DM}$. This discrepancy can also be attributed to the instability arising from normalization. Specifically, the difference between the stationary and maximum entropy distributions in this case is sufficient to lead to two different MIBs, with constituent sub-systems of different sizes. In fact, use of $\bar{\Phi}_{DM}$ leads to the MIB {{1,2,3,5,7,8},{4,6}} of the *perturbed* version of this network discussed in 'Optimization of networks for generating high Φ_E'.

We also compared results obtained using Φ_E with those obtained using $\tilde{\Phi}_E$, the measure constructed using the alternative expression (0.32) for the effective information (Table 1). We found that the two measures behave in qualitatively the same way across all examples.

Extension to multiple lags and to MVAR(p) processes

The analyses in the previous section were concerned with integrated information measured across a single time-step for MVAR(1) processes. However, Φ_E is well-defined for general MVAR(p) processes and can measure integrated information over any number of time-steps (lags). Here we illustrate this property using three simple examples in which Φ_E was computed analytically, via the method outlined in 'Computing Φ_E analytically for a Gaussian system' and 'Methods'. Fig. 4(a) shows Φ_E measured for various values of τ, (where τ specifies the lag), for the network 1(c). Fig. 4(b) shows the same analysis conducted for network 2(b). Note that both of these networks are animated by MVAR(1) processes, which explains why Φ_E peaks at $\tau = 1$ in both cases, (in other words, for these networks, most of the integrated information generated about past states by the current state is generated about the most recent past state (i.e. $\tau = 1$)).

Fig. 4(c) shows Φ_E as a function of τ for the MVAR(3) process

$$\boldsymbol{X}_t = A_1 \cdot \boldsymbol{X}_{t-1} + A_2 \cdot \boldsymbol{X}_{t-2} + A_3 \cdot \boldsymbol{X}_{t-3} + \boldsymbol{E}_t, \qquad (0.44)$$

where A_1, A_2 and A_3 are respectively the connectivity matrices of networks 1(c), 2(b) and 2(a), each divided by 2. Note that this generalized connectivity matrix was chosen purely to provide an example of an MVAR(3) process. For this system, Φ_E peaks at $\tau = 2$, indicating that most information is integrated about the state two time-steps previous to the current state. These examples verify

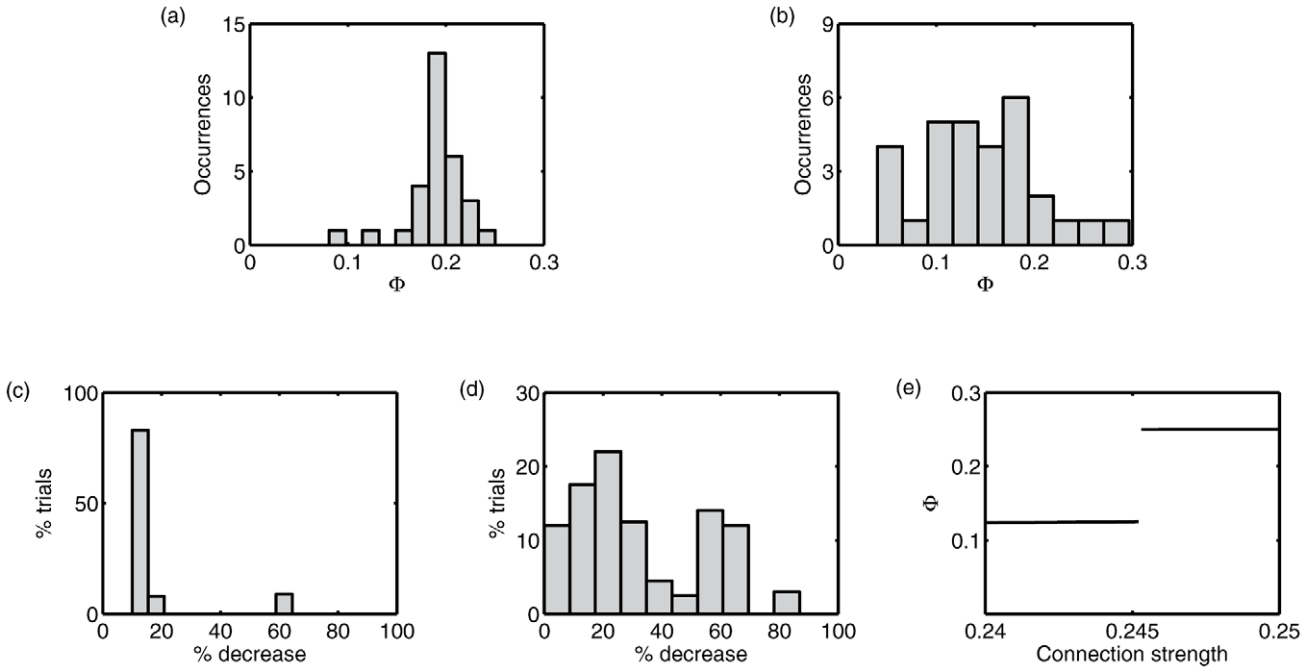

Figure 3. Examination of the Φ_E landscape with respect to network connectivity. (a) Histogram of Φ_E for the 30 networks in the final population of a GA that yielded optimal network 2(a). (b) Histogram of Φ_E for the 30 networks in the final population of a GA that yielded optimal network 2(b). (c) Histogram of percentage decrease in Φ_E following single mutations of network 2(a) (200 evaluations). (d) Histogram of percentage decrease in Φ_E following single mutations of network 2(b) (200 evaluations). (e) Discontinuity in Φ_E as connection strength from element 6 to element 1 continuously changes (network 2(a); (all other connections fixed at 0.25)).
doi:10.1371/journal.pcbi.1001052.g003

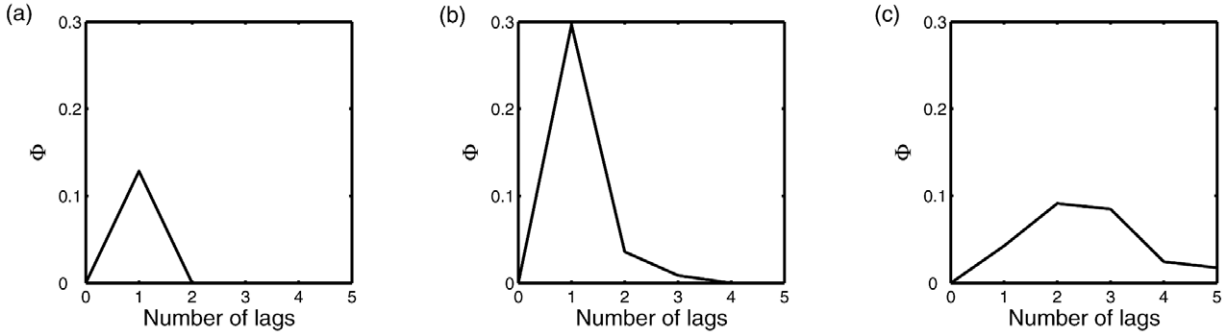

Figure 4. Integrated information, Φ_E measured for states multiple time-steps in the past, i.e. for varying τ. (a) Network 1(c). (b) Network 2(b). (c) Example MVAR(3) process, see Eq. (0.44).
doi:10.1371/journal.pcbi.1001052.g004

that Φ_E can be applied at arbitrary lags to MVAR(p) processes, and that it does detect integrated information at time-scales corresponding to a system's underlying generative mechanism.

Auto-regressive Φ (Φ_{AR})

We have presented a measure of integrated information, Φ_E, that is practical to measure from time-series data under Gaussian assumptions. However, in the case of stationary, non-Gaussian distributed time-series, Φ_E can no longer be obtained directly from empirical covariance matrices, and the required entropies must be obtained via estimation of the corresponding probability distributions. For non-trivial systems accurate entropy estimation may typically require the collection of more data than is practical.

We now describe how, even for the non-Gaussian case, the recipe used to calculate Φ_E under Gaussian assumptions can nonetheless lead to a meaningful quantity reflecting integrated information. We call this quantity Φ_{AR} ('auto-regressive Φ'). By construction, Φ_{AR} is equivalent to Φ_E for Gaussian systems, however, for non-Gaussian systems it may differ. In all cases, because it is based on empirical covariance matrices, it remains easy to measure in practice. The motivation for considering Φ_{AR} as a useful measure of integrated information rests on relations between conditional entropy, partial covariance and linear regression prediction error, explained below [17].

First we rehearse the concept of linear regression. Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be two multivariate random variables. Then the linear regression of $\boldsymbol{X}$ on $\boldsymbol{Y}$ is the expression

$$\boldsymbol{X} = \boldsymbol{\alpha} + A \cdot \boldsymbol{Y} + \boldsymbol{E}\,, \tag{0.45}$$

where A is termed the regression matrix, $\boldsymbol{\alpha}$ is a vector of constants, and $\boldsymbol{E}$ is the prediction error (or 'residual') [27,28,29,17]. The residual is a random vector uncorrelated with $\boldsymbol{Y}$. This representation is unique given the distributions of $\boldsymbol{X}$ and $\boldsymbol{Y}$, with A and $\boldsymbol{\alpha}$ given by

$$A = \Sigma(\boldsymbol{X},\boldsymbol{Y})\Sigma(\boldsymbol{Y})^{-1}\,, \tag{0.46}$$

$$\boldsymbol{\alpha} = \bar{\boldsymbol{x}} - A \cdot \bar{\boldsymbol{y}}\,. \tag{0.47}$$

The residual has zero mean and, importantly, its covariance matrix is precisely the partial covariance of $\boldsymbol{X}$ given $\boldsymbol{Y}$ [17], thus

$$\Sigma(\boldsymbol{E}) = \Sigma(\boldsymbol{X}|\boldsymbol{Y})\,. \tag{0.48}$$

Note that this identity holds for any $\boldsymbol{X}$ and $\boldsymbol{Y}$, Gaussian or otherwise. For the case that $\boldsymbol{X}$ and $\boldsymbol{Y}$ are Gaussian, we can use Eq. (0.9) to obtain, for all $\boldsymbol{y}$,

$$H(\boldsymbol{X}|\boldsymbol{Y}=\boldsymbol{y}) = \frac{1}{2}\log[\det\Sigma(\boldsymbol{E})] + \frac{1}{2}n\log(2\pi e)\,, \tag{0.49}$$

where n is the dimension of $\boldsymbol{X}$. This relation between conditional entropy and linear regression prediction error implies that, for Gaussian systems, Φ_E can be re-expressed in terms of linear regression prediction errors. Thus, the formula (0.33) for effective information can be re-written as

$$\varphi\left[X;\tau,\left\{M^1,M^2\right\}\right] = \frac{1}{2}\log\left\{\frac{\det\Sigma(\mathrm{X})}{\det\Sigma\left(\mathrm{E}^X\right)}\right\} - \sum_{k=1}^{2}\frac{1}{2}\log\left\{\frac{\det\Sigma\left(\mathrm{M}^k\right)}{\det\Sigma\left(\mathrm{E}^{M^k}\right)}\right\}, \tag{0.50}$$

where $\boldsymbol{E}^{M^k}$, $k=1,2$, and $\boldsymbol{E}^X$ are the residuals in the regressions

$$\boldsymbol{M}^k_{t-\tau} = A^{M^k}\cdot\boldsymbol{M}^k_t + \boldsymbol{E}^{M^k}_t\,, \tag{0.51}$$

$$\boldsymbol{X}_{t-\tau} = A^X\cdot\boldsymbol{X}_t + \boldsymbol{E}^X_t\,. \tag{0.52}$$

For a non-Gaussian system, although Eq. (0.50) does not hold, its RHS nonetheless constitutes a quantity that is easy to measure empirically. This quantity forms the basis of the alternative measure Φ_{AR}, which we now define. Let X be a stationary, not necessarily Gaussian, system, and let $\varphi_{AR}\left[X;\tau,\left\{M^1,M^2\right\}\right]$ be the RHS of Eq. (0.50), i.e.

$$\varphi_{AR}\left[X;\tau,\left\{M^1,M^2\right\}\right] =: \frac{1}{2}\log\left\{\frac{\det\Sigma(\mathrm{X})}{\det\Sigma\left(\mathrm{E}^X\right)}\right\} - \sum_{k=1}^{2}\frac{1}{2}\log\left\{\frac{\det\Sigma\left(\mathrm{M}^k\right)}{\det\Sigma\left(\mathrm{E}^{M^k}\right)}\right\}, \tag{0.53}$$

where $\boldsymbol{E}^{M^k}$, $k=1,2$, and $\boldsymbol{E}^X$ are the residuals in the regressions (0.51) and (0.52). Then Φ_{AR} is simply φ_{AR} for the bipartition that minimizes φ_{AR} divided by the normalization factor

$$L(\{M^1,M^2\}) =: \frac{1}{2}\log \min_k \left\{ (2\pi e)^{|M^k|} \det \Sigma(\boldsymbol{M}^k) \right\}. \quad (0.54)$$

Thus,

$$\Phi_{\mathrm{AR}}[X;\tau] =: \varphi_{\mathrm{AR}}\left[X;\tau,\mathcal{B}^{\min}(\tau)\right], \quad (0.55)$$

$$\mathcal{B}^{\min}(\tau) =: \arg_{\mathcal{B}} \min \left\{ \frac{\varphi_{\mathrm{AR}}[X;\tau,\mathcal{B}]}{L(\mathcal{B})} \right\}. \quad (0.56)$$

For Gaussian systems, Φ_{E} and Φ_{AR} are exactly equal. For non-Gaussian systems the two measures differ, because the relation (0.49) between conditional entropy and linear regression prediction error no longer holds. However the equivalence (0.48) between partial covariance and prediction error does still hold. Hence, for *any* stationary system, the recipe for computing Φ_{E} under Gaussian assumptions (as laid out in 'Computing Φ_{E} empirically under Gaussian assumptions') yields precisely Φ_{AR}. Notably, this recipe implies that it is not necessary to explicitly carry out the linear regressions; rather, the equivalence (0.48) shows that Φ_{AR} can be calculated using empirical covariance matrices.

Φ_{AR} is meaningful as a measure of integrated information because of its formulation in terms of linear regression prediction error. Φ_{AR} compares the whole system to the sum of its parts in terms of the log-ratio of the variance of the past state to the variance of the residual of a linear regression of the past on the present. In other words, Φ_{AR} can be understood as a measure of the extent to which the *present* global state of the system predicts the *past* global state of the system, as compared to predictions based on the most informative decomposition of the system into its component parts. When Gaussian conditions are satisfied, the interpretation of Φ_{AR} in terms of (backwards) prediction becomes exactly equivalent to the interpretation of Φ_{E} in terms of Shannon information. Note that in fact, by the symmetry of mutual information (0.7), (0.28), Φ_{AR} could also be expressed in terms of entirely analogous linear regressions in which the *present* is used to predict the *future*. Understood this way, Φ_{AR} provides an interesting complement to complexity measures based on Granger causality, such as *causal density* [5], which are also based on linear regression models [30,5,18] (see 'Comparison with causal density and neural complexity').

To demonstrate the use of Φ_{AR} as distinct from Φ_{E}, we re-animated the networks 1(a)–1(g), 2(a) and 2(b) with non-Gaussian dynamics. Specifically, we replaced the Gaussian noise sources $\boldsymbol{E}_t$ in Eq. (0.43) with independent random variables drawn from exponential distributions with mean (and variance) 1. This selection was motivated by the observation that aggregate assemblies of Poissonian spiking neurons typically follow an exponential distribution [31]. Fig. 5 shows representative examples of single-element empirical stationary distributions resulting from this modified dynamics; all show a large deviation from the Gaussian. For each network we computed Φ_{AR} empirically from 10 trials of 3000 data points each. The results, shown in Table 1, suggest that in each case Φ_{AR} for the non-Gaussian dynamics is approximately equal to Φ_{E} ($=\Phi_{\mathrm{AR}}$) for the Gaussian dynamics. This finding provides support for Φ_{AR} as a useful alternative to Φ_{E}, applicable to non-Gaussian dynamics.

Discussion

In this paper we have presented two new measures of integrated information, Φ_{E} and Φ_{AR}. As with a previous measure, Φ_{DM}, our measures quantify the information generated by a system over and above that which can be accounted for by its parts acting independently [11]. However, whereas Φ_{DM} is defined only for discrete Markovian systems, and is therefore difficult to measure in practice, our quantities are well defined much more generally, and are easily applicable to stationary time-series data. Our key innovations are (i) to treat information in terms of reduction in uncertainty from the empirical as opposed to the maximum entropy distribution (Φ_{E}), and (ii) to interpret integrated information in terms of predictive ability of the present of a system with respect to its past (Φ_{AR}). Simulations showed that our measures conform to intuitions regarding conjoined dynamical integration and segregation; where comparisons could be made, in most cases our measures quantitatively aligned with Φ_{DM}. By showing how to measure integrated information from time-series data and for non-trivial non-Markovian systems, our results provide new opportunities for examining the role of integrated information in complex biological systems of all kinds, and carry implications for integrated information theories of consciousness. In the following discussion, we use the symbol Φ to refer to integrated information independently of its method of measurement.

Empirical and maximum entropy distributions

As mentioned, many of the restrictions in applicability of Φ_{DM} arise from the use of the maximum entropy distribution to measure information. The maximum entropy distribution is maximally agnostic with respect to the behavior of a system, and represents, in some sense, its potential, or 'capacity' (see 'Integrated information as a measure of consciousness' and

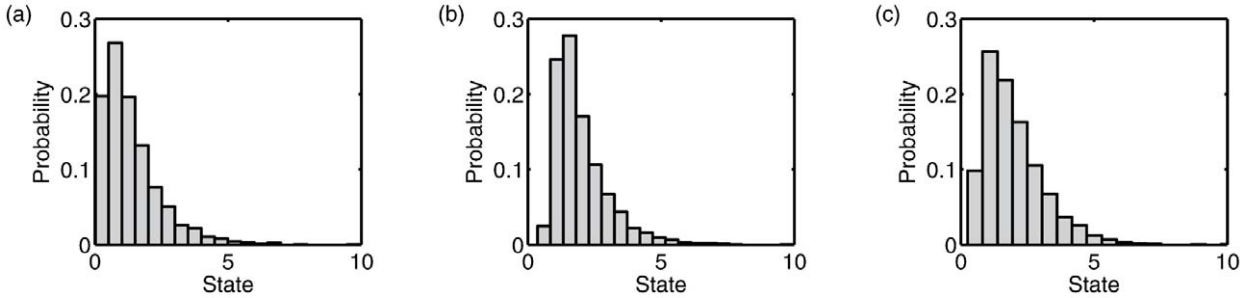

Figure 5. Stationary distributions for elements in networks animated with exponentially distributed noise. Each panel shows an empirical probability distribution as a histogram taken from 3000 data points from element 1 in (a) network 1(b), (b) network 1(d), and (c) network 2(b).
doi:10.1371/journal.pcbi.1001052.g005

'Comparison with causal density and neural complexity'). However, since the maximum entropy distribution typically does not arise spontaneously, it must be introduced as the distribution of a hypothetical initial state [11]. To compute Φ_{DM} one therefore has to characterize evolution from all possible initial states of the system. However, for most practical purposes, especially in biology, it is only possible to experimentally examine systems in the context of their ongoing evolution as a sequence of states. Unless the system is Markovian, evolution from a state with history is not the same as evolution from a hypothetical initial state, implying that Φ_{DM} cannot be applied to non-Markovian systems (with the exception of idealized simulated systems for which a separate generative model can be written down for evolution from the initial state). Equally important, but easier to appreciate, is that it is not possible to apply Φ_{DM} to continuous systems (except those with a compact, i.e. closed and bounded, set of states) because there is no uniquely defined maximum entropy distribution for a continuous random variable defined on the real number line [16].

Our new measure Φ_E eliminates the need to consider the maximum entropy distribution by being based instead on the information generated by the current state of the system about the *actual* state of the system some number of time-steps in the past. This approach lifts the conditions that the system be discrete and Markovian. (Note however that Φ_{DM} but not Φ_E is applicable to deterministic systems, by virtue of introducing probabilities via the maximum entropy initial state.)

In principle, use of the empirical distribution de-emphasizes the notion of 'capacity' because the generation of information is measured with respect to what the system *has done* rather than what it *could do*. However, over large samples and for ergodic systems, this distinction becomes increasingly blurred. In practice, computing Φ_E via sampling from time-series requires the data to be stationary. We recognize that not all complex biological systems generate stationary dynamics (see, e.g., Ref. [32]). However, stationarity is a common pre-requisite for statistical analysis of time-series data [33], and neural data can often be brought into this form, for example by detrending, taking first-differences and/or binning observations into short time windows [34]. Furthermore, neural dynamics are often characterized as a series of 'metastable' states [35,36,37], each of which may be locally stationary. Stationarity can also depend on the spatiotemporal granularity of observation. Dynamics that appear non-stationary at one time scale may exhibit stationarity when sampled over different time scales, underlining the principle that data acquisition should be guided by the constraints of subsequent analysis methods.

Use of the empirical, rather than maximum entropy distribution also changes the means by which Φ is computed. To compute Φ_{DM}, one requires the conditional probability distributions for the past state given the present state, but with an *a priori* maximum entropy distribution on the past state. Because of the maximum entropy condition (which represents 'perturbation' of the system), these distributions cannot be obtained empirically, but they can be obtained by applying Bayes' rule given a forward dynamical model estimated from the data (i.e. conditional probability distributions for the present state, given the past state). By contrast, computation of Φ_E does not require Bayes' rule because, in the absence of (maximum entropy) perturbation, one can obtain the full joint distribution for the past and present directly from the data.

Practical applicability and Gaussian dynamics

Φ_E is particularly easy to apply to data under Gaussian assumptions. This is because the relevant entropies can be estimated directly from empirical covariance matrices. It is also possible to compute Φ_E analytically from a generative model for a Gaussian system, (i.e., to any desired level of accuracy, without explicitly simulating or observing its dynamics); in that case, one obtains the necessary covariance matrices analytically. This means that Φ_E can be evaluated in practice for a broad range of biological systems.

While Gaussian dynamics are common in biology (and the assumption of Gaussianity even more so), many systems depart from this assumption. For example, the spiking activity of populations of neurons typically exhibit exponentially distributed dynamics. For the non-Gaussian case, one can still in principle calculate Φ_E by obtaining the necessary entropies directly from data. However, in practice, accurately obtaining all of the underlying probability distributions may typically require the collection of more data than is practical. To overcome this, we introduced the second measure Φ_{AR}. This is constructed analogously to Φ_E, but with information replaced by the reduction in the generalized covariance of the past state under prediction via linear regression on the current state. Φ_{AR} is interpreted as measuring how well the present state of a system predicts some previous state, but only to the extent that predictions based on the whole outstrip predictions based on parts independently. Φ_{AR} and Φ_E are equivalent for Gaussian systems, but otherwise differ; (recall however that Φ_{AR} can be obtained for any system by using the recipe for computing Φ_E for a Gaussian system). In our examples, Φ_{AR} was in fact insensitive to a change from Gaussian noise to exponentially distributed noise, supporting its use as an alternative to Φ_E.

Normalization and instability

All versions of Φ require a normalization step. Specifically, Φ is determined by the *non-normalized* effective information (φ) across a minimum information bipartition (MIB) which is specified as the bipartition which minimizes the *normalized* φ (the informational 'weakest link'). Normalization enforces a bias towards bipartitions consisting of sub-systems of roughly equal size. Without normalization, MIBs would typically divide systems into single elements versus the remainder of the system, leading to trivially small values of Φ. On the other hand, it remains important to determine the value of Φ using the non-normalized φ in order to allow Φ to be interpreted as a quantity of information.

The use of normalization, as just described, leads to instabilities. Our simulations have shown that Φ_E can be (i) discontinuous under a continuous perturbation of dynamics, and (ii) highly sensitive to the accuracy of entropy estimation from finite data. In our examples, these instabilities arose precisely when there were multiple partitions with similar values of normalized φ close to the true minimum *and* these partitions had substantially different values of non-normalized φ. This instability does not arise for all systems, and indeed for most of our examples Φ_E is numerically stable. Nonetheless, the embedding of normalization within the definition of Φ challenges ascription of physical meaning to any measured value of Φ. This is because the value of Φ is in all cases dependent to some arbitrary degree on the normalization process involved in determining the MIB.

Integrated information as a measure of consciousness

Previous measures of integrated information (Φ_C and Φ_{DM}) were formulated in the context of a theory of consciousness, the 'integrated information theory of consciousness' (IITC). According to the IITC, consciousness *is* integrated information, and has the status of a fundamental property of the universe, equivalent to mass, charge, and the like [14]. On this theory a low value of integrated information would correspond to a low conscious 'level' (e.g., coma, general anesthesia, deep dreamless sleep) and a high

value to normal conscious wakefulness. If one subscribes to the theory using Φ_{DM}, then one must interpret consciousness (integrated information) as a function of state transitions [11]; accordingly, one cannot ask about the conscious level of a system *per se*. By contrast, if one applies Φ_E or Φ_{AR} to a stationary system then they are state-independent and so, subscribing to the IITC with these measures involves viewing integrated information as a property of the system's dynamics. This in turn would imply that (i) conscious level is constant during each stationary epoch in brain activity, and (ii) conscious level changes when functional connectivity changes, modifying the stationary statistics. This view recalls William James' notion of consciousness as a process [19] and is consistent with a large amount of empirical evidence showing correlations between conscious level and plausibly stationary epochs of brain activity. For example, normal conscious wakefulness is characterized by low-amplitude high-frequency oscillations in the cortical EEG [38], whereas epileptic absence seizures are characterized instead by increased synchrony in thalamocortical systems [39]. As mentioned in the section 'Empirical and maximum entropy distributions', neural dynamics may be metastable [35,36,37], with locally stationary periods corresponding to a conscious state with a particular level and content. Our results now make it possible to measure the integrated information corresponding to these various states and to compare these values with other indices of consciousness, both subjective (e.g., verbal reports, confidence ratings, etc.) and objective (e.g., EEG synchrony, widespread brain activity, etc.) [40]. Importantly, it is now possible to quantitatively compare integrated information with other measures of neural dynamics that operationalize in different ways the notion that consciousness conjoins dynamical integration and differentiation, such as 'causal density' [41] and 'neural complexity' [8] (see 'Comparison with causal density and neural complexity').

An important feature of the IITC as previously expressed is that consciousness *qua* Φ is best considered as a capacity (equivalently a potential, or disposition), and not as an 'object' or a process [14]. The original Φ_C operationalized the notion of capacity by subjecting a system to all possible perturbations and examining its responses. The recent Φ_{DM} measures information as a reduction in entropy from the maximum entropy distribution, which can be taken to correspond to the capacity of a system. However, because Φ_{DM} is specified by state transitions it is not a 'pure' measure of capacity; rather, it is a measure of capacity modulated by a system's dynamics. By measuring Φ with reference to the stationary distribution, our measures depart from the notion of consciousness as a capacity. The stationary distribution characterizes the capacity of a system only to the extent that it is realized in the system's behaviour. Φ_E and Φ_{AR} can therefore be construed as measures of a process modulated by capacity, aligning more closely with the Jamesian intuition.

The notion that Φ exists as a 'fundamental property' deserves comment. As described in the section 'Normalization and instability', our results challenge the ascription of physical meaning to Φ, in virtue of its exquisite sensitivity to the normalization process involved in specifying the MIB: this challenge pertains equally to the notion of Φ as a 'fundamental quantity'. A further challenge to the ascription of physical meaning to Φ is the fact that it is not invariant under a change of coordinates, since this leads to a different set of sub-systems over which to minimize the effective information. An interesting question for future work is to examine whether, under certain conditions, the set of coordinates that maximizes Φ could be taken to define 'natural' coordinates, or macroscopic variables, for the system. In any case, it does not seem necessary to consider Φ as a strict physical quantity in order to measure the integrated information corresponding to a system's state transitions or stationary dynamics, nor to relate these measurements to conscious level and content. In other words, one can depart from the IITC by interpreting Φ as accounting for particular aspects of consciousness without the further step of claiming identity [9].

Integrated information in other neurocognitive processes

Although Φ was originally developed in the context of a theory of consciousness, it is plausible that integrated information, and (more generally) conjoined functional integration and differentiation, play key roles in other cognitive and neural processes. Previous formulations (Φ_{DM}, Φ_C) are poorly suited to investigating these roles, not only because of practical inapplicability, but also because they characterize integrated information in terms of capacity rather than process. Whereas consciousness under some theories may be considered as a capacity (see above), neurocognitive properties in general are best considered as processes. Having a measure of Φ that is framed in terms of process, and that is easy to apply in practice, therefore permits the framing of testable hypotheses, and the specification of synthetic models, aimed at examining the role of integrated information in neurocognitive processes broadly construed. For example, multimodal binding and perceptual categorization [20], and action selection (decision making) [21] plausibly involve integrated information and could be profitably analyzed using our methods. Already, related measures of dynamical complexity (neural complexity and causal density, see below) have been correlated with the ability of simulated agents to deploy flexible behavior, suggesting a role for such dynamics in sensorimotor coordination in rich environments [6,41]. Our results now allow integrated information to be applied in similar situations, facilitating comparative analyses.

Comparison with causal density and neural complexity

Φ is one among a family of recent measures that aim to characterize, in different ways, the coexistence of integration and differentiation in a system's dynamics. Two alternative measures are 'causal density' [41] and 'neural complexity' [42]. Here, we briefly summarize the similarities and differences among these measures, in order to set Φ into a broader context.

Causal density, like Φ_{AR} and Φ_E (but in contrast to Φ_{DM} and Φ_C), is a measure of process rather than capacity. In virtue of being based on 'Granger causality', it also shares with Φ a sensitivity to causal interactions within a system. A key difference, however, is that causal density is based on *all* causal interactions, and not just those across a particular partition; thus causal density avoids the normalization problems described above ('Normalization and instability'). Briefly, Granger causality is a statistical measure of causal influence which asserts that a variable X^1 'Granger causes' another variable X^2 if information in the past of X^1 helps predict the future of X^2, above and beyond information already in the past of X^2 (and, optionally, in the past of a set of conditioning variables $X^{3\ldots N}$) [30,43]. Causal density is then the (weighted) fraction of causal interactions among all elements that are statistically significant. High causal density indicates that elements within a system are both globally coordinated in their activity (to be useful for predicting each others' activity) and at the same time dynamically distinct (so that different elements contribute in different ways to these predictions). Granger causality (and causal density) is typically calculated using linear auto-regressive models, which brings about an interesting comparison with Φ_{AR}. In a loose sense, integrated information,

as measured by Φ_{AR} or Φ_E, can be thought of as a variety of 'causal density', that quantifies the strength of the weakest bidirectional causal link between any two halves of the system. Forthcoming work will investigate further the links between Φ_{AR} and causal density.

Neural complexity is calculated as the sum of the average mutual information across all bipartitions of a system [42]. Unlike Φ and causal density, it does not reflect causal interactions within a system, however, like causal density, it is a measure of process rather than capacity. Neural complexity is maximal in a system that is globally integrated at the level of large subsystems, while exhibiting a high degree of segregation between smaller subsystems. (Note: The original papers describing neural complexity contained an error in calculating the covariance matrix from a generative model, which has been subsequently corrected in [44]. However, it appears that this error may still affect extant calculations of Φ_C.) A recent result [17] showing an equivalence between Granger causality and 'transfer entropy' (a time-directed version of mutual information) allows causal density to be related directly to neural complexity. Specifically, one can define a 'bipartition causal density' as a weighted average Granger causality (transfer entropy) across all bipartitions of a system (this definition also requires extension of Granger causality to multivariate variables) [18]. This measure furnishes a 'time-directed' version of neural complexity based on transfer entropy rather than mutual information.

These relations together suggest common foundations for measures of coexisting integration and differentiation. However, further work is needed to fully establish their theoretical interdependencies and their empirical convergences and divergences.

Comparison with other measures

Characterizing complexity is a diverse field, and there are other measures that capture complex properties other than conjoined differentiation and integration. For example, 'thermodynamic depth' [45] can be interpreted as a measure of how hard it is to put a system together, and is based on the joint entropy of all past states, given the current state. Φ by contrast considers only one past state. An interesting further modification to Φ could involve information between the present and the whole past trajectory of the system. Another measure of statistical interdependence, 'informational coherence', considers the optimal predictive state for each time-series, and then measures mutual information between these [46]. In related work by Ay et al., the whole system is compared to the sum of individual elements [22,47,48], and the analysis goes beyond examination of conditional entropies to a more thorough mathematical treatment in terms of information geometry. While it is beyond the present scope to examine the formal correspondences among these measures, other related measures, and the measures described above, the growing interest in quantitative measures of complexity further emphasizes the need to formulate theoretically principled measures that are also simple to apply in practice.

Limitations and extensions

Although our measures represent substantial improvements in practical applicability of measures of integrated information, several limitations remain. Most prominently, the normalization procedure leads to instabilities in the measurement process and undercuts ascription of physical meaning to Φ. Addressing this problem stands as a key theoretical challenge. We have only considered application of our measures to stationary dynamics. Future work may extend consideration to non-stationary (but still continuous and non-Markovian) processes, potentially capturing important non-stationary aspects of neural dynamics. In addition, our measures are applicable only to stochastic systems. While extension to closed deterministic systems may be of some value, most complex biological systems have stochastic components, especially when considered in interaction with a (stochastic) environment [49,50]. Finally, our measures share with previous measures the computational challenge posed by the combinatorial explosion in partitions of a system as the number of elements increases. Possibly, imposing priors on the search for the minimum information partition may mitigate this challenge.

We have only considered a first-order, linear approximation for computing entropies/information from data. While this is useful for drawing comparison with Granger causality and causal density, there now exist more advanced approximation techniques that could be used in future work, for example additive regression [51] or kernel regression [52]. Regarding estimation of entropy and mutual information without employing a regression model, we have only considered this via the intermediate step of density estimation. Again, future work could investigate the applicability of more advanced techniques [53,54] that avoid this step.

As well as addressing the above challenges, future work will (i) empirically examine integrated information for time-series data acquired from neuroimaging and other biological datasets, in order to test intuitions regarding consciousness and other neurocognitive processes; (ii) investigate in models how integrated information is modulated by input and output relations of a system embedded in, and interacting with, a surrounding environment, and (iii) determine theoretically the relations between integrated information and alternative measures of dynamical complexity and metastability.

Methods

Text S1 in 'Supporting Information' contains software enabling calculation of ΦE and ΦAR, as well as functions which allow regeneration of some of the simulations we describe.

Extension and computation of Φ_{DM} for an MVAR(1) process

To extend Φ_{DM} to stationary continuous Markovian systems, we have to address the problem that there is no well-defined maximum entropy distribution for such systems. We do this by replacing the 'maximum entropy distribution' with the distribution for which the state of each element is independent of the states of all other elements, is Gaussian distributed, and has mean and variance equal to those of its corresponding stationary distribution. Thus, we take $\boldsymbol{X}_0 \sim \mathcal{N}\left(\bar{\boldsymbol{x}}, \Sigma^{\mathrm{D}}(\boldsymbol{X})\right)$, where

$$\Sigma^{\mathrm{D}}(\boldsymbol{X})_{ij} = \begin{cases} \Sigma(\boldsymbol{X})_{ij}, & i=j, \\ 0, & i \neq j. \end{cases} \tag{0.57}$$

Having defined a distribution for the initial state $\boldsymbol{X}_0$, we explain how to compute the expected integrated information, $\bar{\Phi}_{DM}$, for MVAR(1) processes (0.36). The computation proceeds analytically, given the generative model, which is specified by the connectivity matrix A and the covariance matrix of the noise, $\Omega =: \Sigma(\boldsymbol{E})$. Alternatively, an estimate of $\bar{\Phi}_{DM}$ from time-series data can be obtained by using estimates of A and Ω. The linear-regression formulae (0.46) and (0.48) yield the estimates

$$\hat{A} = \hat{\Sigma}(\boldsymbol{X}_t, \boldsymbol{X}_{t-1}) \hat{\Sigma}(\boldsymbol{X})^{-1}, \tag{0.58}$$

$$\hat{\boldsymbol{\Omega}}=\hat{\boldsymbol{\Sigma}}(\boldsymbol{X}_t|\boldsymbol{X}_{t-1}), \tag{0.59}$$

where the symbol $\hat{\boldsymbol{\Omega}}$ denotes empirical quantities.

Given A and $\boldsymbol{\Omega}$, (or their estimates $\hat{\boldsymbol{A}}$ and $\hat{\boldsymbol{\Omega}}$), the covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{X})$ can be obtained via the discrete-time Lyapunov equation (0.37),

$$\boldsymbol{\Sigma}(\boldsymbol{X})=A\boldsymbol{\Sigma}(\boldsymbol{X})A^{\mathrm{T}}+\boldsymbol{\Omega}, \tag{0.60}$$

and $\boldsymbol{\Sigma}^{\mathrm{D}}$ from Eq. (0.57).

To compute the conditional probability $P_{\boldsymbol{X}_0|\boldsymbol{X}_1=\boldsymbol{x}}$ we first use the MVAR(1) dynamics (0.36) to obtain the distribution of $\boldsymbol{X}_1|\boldsymbol{X}_0=\boldsymbol{x}'$ as

$$\boldsymbol{X}_1|(\boldsymbol{X}_0=\boldsymbol{x}')\sim\mathcal{N}(A\boldsymbol{x}',\boldsymbol{\Omega}). \tag{0.61}$$

Then we use Bayes' rule (0.15) to obtain

$$P_{\boldsymbol{X}_0|\boldsymbol{X}_1=\boldsymbol{x}}(\boldsymbol{x}')\propto P_{\boldsymbol{X}_1|\boldsymbol{X}_0=\boldsymbol{x}'}(\boldsymbol{x})\cdot P_{\boldsymbol{X}_0}(\boldsymbol{x}') \tag{0.62}$$

$$\propto\exp\left\{-\frac{1}{2}\left[(\boldsymbol{x}-A\boldsymbol{x}')^{\mathrm{T}}\boldsymbol{\Omega}^{-1}(\boldsymbol{x}-A\boldsymbol{x}')+\boldsymbol{x}'^{\mathrm{T}}\left(\boldsymbol{\Sigma}^{\mathrm{D}}\right)^{-1}\boldsymbol{x}'\right]\right\}. \tag{0.63}$$

From the term quadratic in $\boldsymbol{x}'$ we can obtain the inverse of the covariance matrix of (the Gaussian distributed) conditional variable $\boldsymbol{X}_0|\boldsymbol{X}_1=\boldsymbol{x}$ as

$$\boldsymbol{\Sigma}(\boldsymbol{X}_0|\boldsymbol{X}_1=\boldsymbol{x})^{-1}=A^{\mathrm{T}}\boldsymbol{\Omega}^{-1}A+\left(\boldsymbol{\Sigma}^{\mathrm{D}}\right)^{-1}, \tag{0.64}$$

and hence express the conditional entropy $H(\boldsymbol{X}_0|\boldsymbol{X}_1=\boldsymbol{x})$ in terms of the connectivity and stationary covariance matrices:

$$H(\boldsymbol{X}_0|\boldsymbol{X}_1=\boldsymbol{x})=\frac{1}{2}|X|\log(2\pi e)-\frac{1}{2}\log\left\{\det\left[A^{\mathrm{T}}\boldsymbol{\Omega}^{-1}A+\left(\boldsymbol{\Sigma}^{\mathrm{D}}\right)^{-1}\right]\right\}. \tag{0.65}$$

For a given a sub-system M, we have to consider the bipartition $X=\{M,N\}$, and the block decomposition of vectors and matrices according to $\boldsymbol{X}_t=(\boldsymbol{M}_t,\boldsymbol{N}_t)^{\mathrm{T}}$ so that

$$\boldsymbol{\Sigma}^{\mathrm{D}}=\begin{pmatrix}\boldsymbol{\Sigma}^{\mathrm{D}}_M & 0\\ 0 & \boldsymbol{\Sigma}^{\mathrm{D}}_N\end{pmatrix},\quad A=\begin{pmatrix}A_{MM} & A_{MN}\\ A_{NM} & A_{NN}\end{pmatrix}, \tag{0.66}$$

and similarly for $\boldsymbol{\Omega}$ and $\boldsymbol{E}_t$. To obtain the distribution for the conditional random variable $\boldsymbol{M}_1|\boldsymbol{M}_0=\boldsymbol{m}'$, we express $\boldsymbol{M}_1$ in terms of $\boldsymbol{M}_0$ as

$$\boldsymbol{M}_1=A_{MM}\boldsymbol{M}_0+A_{MN}\boldsymbol{N}_0+\boldsymbol{E}_{M1}, \tag{0.67}$$

and note that $\boldsymbol{N}_0\sim\mathcal{N}\left(0,\boldsymbol{\Sigma}^{\mathrm{D}}_N\right)$. Hence

$$\boldsymbol{M}_1|(\boldsymbol{M}_0=\boldsymbol{m}')\sim\mathcal{N}\left(A_{MM}\boldsymbol{m}',\boldsymbol{\Omega}_{MM}+A_{MN}\boldsymbol{\Sigma}^{\mathrm{D}}_NA^{\mathrm{T}}_{MN}\right). \tag{0.68}$$

From Bayes' rule, we can then calculate the inverse of the covariance matrix of (the Gaussian distributed) conditional variable $\boldsymbol{M}_0|\boldsymbol{M}_1=\boldsymbol{m}$ as

$$\boldsymbol{\Sigma}(\mathrm{M}_0|\mathrm{M}_1=\mathrm{m})^{-1}=A^{\mathrm{T}}_{MM}\left(\boldsymbol{\Omega}_{MM}+A_{MN}\boldsymbol{\Sigma}^{\mathrm{D}}_NA^{\mathrm{T}}_{MN}\right)^{-1}A_{MM}+\left(\boldsymbol{\Sigma}^{\mathrm{D}}_{MM}\right)^{-1}, \tag{0.69}$$

and hence

$$H(\mathrm{M}_0|\mathrm{M}_1=\mathrm{m})=\frac{1}{2}|M|\log(2\pi e)-\frac{1}{2}\log\left\{\det\left[A^{\mathrm{T}}_{MM}\left(\boldsymbol{\Omega}_{MM}+A_{MN}\boldsymbol{\Sigma}^{\mathrm{D}}_NA^{\mathrm{T}}_{MN}\right)^{-1}A_{MM}+\left(\boldsymbol{\Sigma}^{\mathrm{D}}_{MM}\right)^{-1}\right]\right\}. \tag{0.70}$$

The entropy formulae (0.65) and (0.70) furnish the sufficient quantities for computing $\bar{\boldsymbol{\Phi}}_{\mathrm{DM}}$ as described in the section 'The previous measure, Φ_{DM}', using the expression (0.25) for the expected effective information. For present purposes, as with Φ_{E}, we restrict attention to bipartitions only.

Analytical computation of Φ_{E} for a general Gaussian case

Here we show how to compute Φ_{E} analytically, for a general stationary Gaussian system, for any timescale τ. Importantly, the generative model for such a system X is always equivalent to an $\mathrm{MVAR}(p)$ process [18]:

$$\boldsymbol{X}_t=A_1\cdot\boldsymbol{X}_{t-1}+A_2\cdot\boldsymbol{X}_{t-2}+\cdots+A_p\cdot\boldsymbol{X}_{t-p}+\boldsymbol{E}_t, \tag{0.71}$$

where the A_i, $i=1,\ldots,p$, can be thought of as generalized connectivity matrices acting at different time-lags, and $\boldsymbol{E}_t$ is a stationary multivariate Gaussian 'white noise' source with zero mean and vanishing auto-covariance function, $\Gamma_\tau(\boldsymbol{E})=0$, $\tau\neq0$. (We ignore the case $p=\infty$ corresponding to an MA(1), i.e. moving average, process.) This system is stationary if and only if the roots of the equation

$$\det\left(I_{|X|}-\sum_{i=1}^{p}z^iA_i\right)=0 \tag{0.72}$$

lie outside the unit circle [33].

The method outlined in 'Computing Φ_{E} analytically for a Gaussian system' for computing Φ_{E} with $\tau=1$ for an MVAR(1) process is easy to extend to the more general $\mathrm{MVAR}(p)$, any τ, case given by equation (0.71). Suppose we wish to compute Φ_{E} for any value of τ up to $\tau=q$, where $q>p$. We first use the fact [33] that the $\mathrm{MVAR}(p)$ process is equivalent to the MVAR(1) process

$$\boldsymbol{\Xi}_t=F\cdot\boldsymbol{\Xi}_{t-1}+\boldsymbol{V}_t, \tag{0.73}$$

involving the block quantities $\boldsymbol{\Xi}_t=:\left(\boldsymbol{X}_t,\boldsymbol{X}_{t-1},\ldots,\boldsymbol{X}_{t-q}\right)^{\mathrm{T}}$, $\boldsymbol{V}_t=:(\boldsymbol{E}_t,0,0,\ldots,0)^{\mathrm{T}}$ and

$$F=:\begin{pmatrix}A_1 & A_2 & A_3 & \cdots & A_p & O_{|X|} & \cdots & O_{|X|} & O_{|X|}\\ I_{|X|} & O_{|X|} & O_{|X|} & \cdots & O_{|X|} & O_{|X|} & \cdots & O_{|X|} & O_{|X|}\\ O_{|X|} & I_{|X|} & O_{|X|} & \cdots & O_{|X|} & O_{|X|} & \cdots & O_{|X|} & O_{|X|}\\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots & \vdots\\ O_{|X|} & O_{|X|} & O_{|X|} & \cdots & O_{|X|} & O_{|X|} & \cdots & I_{|X|} & O_{|X|}\end{pmatrix}. \tag{0.74}$$

The stationary covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\Xi})$ for this process can be obtained from the Lyapunov equation, by analogy with $\boldsymbol{\Sigma}(\boldsymbol{X})$ for

the MVAR(1) case (0.37):

$$\Sigma(\Xi) = F\Sigma(\Xi)F^{\mathrm{T}} + \Sigma(\boldsymbol{V}), \tag{0.75}$$

where $\Sigma(\boldsymbol{V}) = \mathrm{diag}\left[\Sigma(\boldsymbol{E}), O_{|X|}, O_{|X|}, \ldots, O_{|X|}\right]$. Then the stationary covariance $\Sigma(\boldsymbol{X})$ and auto-covariance $\Gamma_\tau(\boldsymbol{X})$ are obtained respectively as the $(1,1)$ and $(\tau+1,1)$ component blocks of $\Sigma(\Xi)$. We can then proceed as for the MVAR(1), $\tau = 1$ case:

$$\Sigma(\boldsymbol{X}_{t-\tau}|\boldsymbol{X}_t) = \Sigma(\boldsymbol{X}) - \Gamma_\tau(\boldsymbol{X})\Sigma(\boldsymbol{X})^{-1}\Gamma_\tau(\boldsymbol{X})^{\mathrm{T}}, \tag{0.76}$$

$$\begin{aligned}\Sigma(\mathrm{X}) &= \begin{pmatrix} \Sigma(\mathrm{X})_{MM} & \Sigma(\mathrm{X})_{MN} \\ \Sigma(\mathrm{X})_{NM} & \Sigma(\mathrm{X})_{NN} \end{pmatrix}, \\ \Gamma_\tau(\mathrm{X}) &= \begin{pmatrix} \Gamma_\tau(\mathrm{X})_{MM} & \Gamma_\tau(\mathrm{X})_{MN} \\ \Gamma_\tau(\mathrm{X})_{NM} & \Gamma_\tau(\mathrm{X})_{NN} \end{pmatrix},\end{aligned} \tag{0.77}$$

$$\Sigma(\boldsymbol{M}) = \Sigma(\boldsymbol{X})_{MM}, \quad \Gamma_\tau(\boldsymbol{M}) = \Gamma_\tau(\boldsymbol{X})_{MM}, \tag{0.78}$$

$$\begin{aligned}&\Sigma(\mathrm{M}_{t-\tau}|\mathrm{M}_t) = \\ &\Sigma(\mathrm{X})_{MM} - \Gamma_\tau(\mathrm{X})_{MM}\left[\Sigma(\mathrm{X})_{MM}\right]^{-1}\Gamma_\tau(\mathrm{X})_{MM}^{\mathrm{T}}.\end{aligned} \tag{0.79}$$

The above expressions furnish the quantities needed to compute Φ_{E} from equations (0.29), (0.30), (0.33) and (0.34).

Supporting Information

Text S1 Toolbox for computing integrated information as Φ_{E} or Φ_{AR}. 'phiemvarp.m' computes Φ_{E} from an MVAR(p) generative model. 'ARphidata.m' computes Φ_{AR} ($=\Phi_{\mathrm{E}}$ if Gaussian), from stationary time-series data. 'statdata.m' creates time-series data from an MVAR(p) generative model. 'A2b.mat' contains the connectivity matrix for the optimal network, Fig. 2(b). 'time-reverse.m' is an m-file for time-reversing the data (required to run ARphidata.m).
Found at: doi:10.1371/journal.pcbi.1001052.s001 (0.01 MB ZIP)

Acknowledgments

We are grateful to Lionel Barnett and Chang-Sub Kim for useful discussions. Nihat Ay, David Balduzzi and one anonymous reviewer provided extremely detailed and valuable comments on a first draft of this paper.

Author Contributions

Conceived and designed the experiments: ABB AKS. Performed the experiments: ABB. Analyzed the data: ABB AKS. Contributed reagents/materials/analysis tools: ABB. Wrote the paper: ABB AKS.

References

1. Honey CJ, Kötter R, Breakspear M, Sporns O (2007) Network structure of cerebral cortex shapes functional connectivity on multiple time scales. Proc Natl Acad Sci U S A 104: 10240–10245.
2. Sporns O, Chialvo D, Kaiser M, Hilgetag C (2004) Organization, development and function of complex brain networks. Trends Cogn Sci 8: 418–425.
3. Bressler SL, Tognoli E (2006) Operational principles of neurocognitive networks. Int J Psychophysiol 60: 139–148.
4. Edelman GM (2003) Naturalizing consciousness: A theoretical framework. Proc Natl Acad Sci U S A 100: 5520–5524.
5. Seth AK, Izhikevich E, Reeke GN, Edelman GM (2006) Theories and measures of consciousness: An extended framework. Proc Natl Acad Sci U S A 103: 10799–10804.
6. Seth AK, Edelman GM (2004) Environment and behavior influence the complexity of evolved neural networks. Adapt Behav 12: 5–21.
7. Lungarella M, Sporns O (2006) Mapping information flow in sensorimotor networks. PLoS Comput Biol 2: e144.
8. Tononi G, Edelman GM (1998) Consciousness and complexity. Science 282: 1846–1851.
9. Seth AK (2009) Explanatory correlates of consciousness: Theoretical and computational challenges. Cogn Comput 1: 50–63.
10. Tononi G, Sporns O (2003) Measuring information integration. BMC Neurosci 4: 31.
11. Balduzzi D, Tononi G (2008) Integrated information in discrete dynamical systems: Motivation and theoretical framework. PLoS Comput Biol 4(6): e1000091.
12. Tononi G (2004) An information integration theory of consciousness. BMC Neurosci 5: 42.
13. Balduzzi D, Tononi G (2009) Qualia: The geometry of integrated information. PLoS Comput Biol 5(8): e1000462.
14. Tononi G (2008) Consciousness as integrated information: A provisional manifesto. Biol Bull 215(3): 216–242.
15. Tononi G (2005) Consciousness, information integration, and the brain. Prog Brain Res 150: 109–126.
16. Cover TM, Thomas JA (1991) Elements of information theory. New York: Wiley-Interscience. 776 p.
17. Barnett L, Barrett AB, Seth AK (2009) Granger causality and transfer entropy are equivalent for Gaussian variables. Phys Rev Lett 103: 238701.
18. Barrett AB, Barnett L, Seth AK (2010) Multivariate Granger causality and generalized variance. Phys Rev E 81: 041907.
19. James W (1904) Does consciousness exist? J Philos Psychol Sci Meth 1: 477–491.
20. Seth AK, McKinstry JL, Edelman GM, Krichmar JL (2004) Visual binding through reentrant connectivity and dynamic synchronization in a brain-based device. Cereb Cortex 14: 1185–1199.
21. Cisek P, Kalaska JF (2010) Neural mechanisms for interacting with a world full of action choices. Annu Rev Neurosci 33: 269–298.
22. Ay N (2001) Information geometry on complexity and stochastic interaction. MPI MIS Preprint 95. Available: http://www.mis.mpg.de/publications/preprints/2001/prepr2001-95.html.
23. Watts DJ, Strogatz SH (1998) Collective dynamics of 'small world' networks. Nature 393: 440–442.
24. Watts DJ (2004) Small worlds: The dynamics of networks between order and randomness (Princeton studies in complexity). Princeton: Princeton University Press. 264 p.
25. Shanahan M (2008) Dynamical complexity in small-world networks of spiking neurons. Phys Rev E 78: 041924.
26. Mitchell M (1997) An introduction to genetic algorithms. CambridgeMA: MIT Press. 221 p.
27. Wilks SS (1932) Certain generalizations in the analysis of variance. Biometrika 24: 471–494.
28. Davidson J (2000) Econometric theory. Oxford: Wiley-Blackwell. 528 p.
29. Ding M, Chen Y, Bressler S (2006) Granger causality: Basic theory and application to neuroscience. In: Schelter S, Winterhalder M, Timmer J, eds. Handbook of time series analysis. Wienheim: Wiley. pp 438–460.
30. Granger CWJ (1969) Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37: 424–438.
31. Dayan P, Abbott LF (2001) Theoretical neuroscience: Computational and mathematical modeling of neural systems. CambridgeMA: The MIT Press. 576 p.
32. Buzsaki G (2006) Rhythms of the brain. Oxford: Oxford University Press. 464 p.
33. Hamilton JD (1994) Time series analysis. Princeton: Princeton University Press. 820 p.
34. Seth AK (2010) A MATLAB toolbox for Granger causal connectivity analysis. J Neurosci Meth 186: 262–273.
35. Werner G (2007) Metastability, criticality and phase transitions in brain and its models. Biosystems 90: 496–508.
36. Freeman WJ, Skarda CA (1985) Spatial EEG patterns, non-linear dynamics and perception: The neo-Sherringtonian view. Brain Res 357: 147–175.
37. Bressler S, Kelso J (2001) Cortical coordination dynamics and cognition. Trends Cogn Sci 5: 26–36.
38. Seth AK, Baars BJ, Edelman DB (2005) Criteria for consciousness in humans and other mammals. Conscious Cogn 14: 119–139.

39. Arthuis M, Valton L, Régis J, Chauvel P, Wendling F, et al. (2009) Impaired consciousness during temporal lobe seizures is related to increased long-distance cortical-subcortical synchronization. Brain 132: 2091–2101.
40. Seth AK, Dienes Z, Cleeremans A, Overgaard M, Pessoa L (2008) Measuring consciousness: Relating behavioural and neurophysiological approaches. Trends Cogn Sci 12: 314–321.
41. Seth AK (2005) Causal connectivity of evolved neural networks during behavior. Network 16: 35–54.
42. Tononi G, Sporns O, Edelman GM (1994) A measure for brain complexity: Relating functional segregation and integration in the nervous system. Proc Natl Acad Sci U S A 91: 5033–5037.
43. Geweke J (1982) Measurement of linear dependence and feedback between multiple time series. J Am Stat Assoc 77: 304–313.
44. Barnett L, Buckley CL, Bullock S (2009) Neural complexity and structural connectivity. Phys Rev E 79: 051914.
45. Pagels H, Lloyd S (1988) Complexity as thermodynamic depth. Ann Phys 188: 186–213.
46. Camperi MF, Klinkner KL, Shalizi CR (2005) Measuring shared information and coordinated activity in neuronal networks. Adv Neural In 18: 667–674.
47. Wennekers T, Ay N (2001) Dynamical properties of strongly interacting markov chains. Neural Netw 16: 1483–1497.
48. Ay N, Olbrich E, Bertschinger N, Jost J (2001) A unifying framework for complexity measures of finite systems. Proceedings of ECCS'06, Santa Fe Institute Working Paper 06-08-028 Available: http://www.santafe.edu/media/workingpapers/06-08-028.pdf.
49. Rolls ET, Deco G (2010) The noisy brain: Stochastic dynamics as a principle of brain function. Oxford: Oxford University Press. 304 p.
50. McDonnell MD, Abbott D (2009) What is stochastic resonance? Definitions, misconceptions, debates, and its relevance to biology. PLoS Comput Biol 5: e1000348.
51. Yao Q, Fan J (2003) Nonlinear time series: Nonparametric and parametric methods. Berlin: Springer-Verlag. 552 p.
52. Bosq D (1998) Nonparametric statistics for stochastic processes: Estimation and prediction, 2nd edn. Berlin: Springer-Verlag. 232 p.
53. Kraskov A, Stoegbauer H, Grassberger P (2004) Estimating mutual information. Phys Rev E 69: 066138.
54. Paninski L (2003) Estimation of entropy and mutual information. Neural Comput 15: 1191–1254.