Practical measures of integrated information for time series data

Barrett, Adam B and Seth, Anil K (2011) Practical measures of integrated information for time series data. PLoS Computational Biology, 7 (1). e1001052. ISSN 1553-734X

This version is available from Sussex Research Online: http://sro.sussex.ac.uk/21423/

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher’s version. Please see the URL above for details on accessing the published version.

Copyright and reuse:
Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

http://sro.sussex.ac.uk
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’, ΦDM, 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, ΦDM is defined only for discrete Markov systems, which are unusual in biology; as a result, ΦDM can rarely be measured in practice. Here, we describe two new measures, ΦC and ΦAR, 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.

Introduction
How can the complex dynamics exhibited by networks of interconnected elements be best 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 selective to a broad range of stimuli, in an integrated way, systems of discrete elements that evolve through time with dramatic claim invites a close examination of the in-principle and in-practice properties of Φ.

A first version of Φ (which we call ΦC, ‘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 ΦC 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 ΦDM, ‘Φ-discrete/Markov’ [11]. In contrast to ΦC, ΦDM is defined for systems of discrete elements that evolve through time with Markovian transitions. Specifically, ΦDM 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, ΦDM measures information as reduction in entropy from a prior maximum entropy distribution, which is taken to represent the repertoire of possible states.

* E-mail: adam.barrett@sussex.ac.uk
It has been shown, using simulations, that $\Phi_{DM}$ behaves consistently with several intuitions about dynamical complexity [11]. In particular, high values of $\Phi_{DM}$ are generated by networks that exhibit both differentiation and integration in their dynamics. However, $\Phi_{DM}$ 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 ITC, which depend on quantifying integrated information.

In this paper we introduce an alternative measure of integrated information, $\Phi_E$ (‘$\Phi$-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 $\Phi_E$ is based on the reduction in Shannon entropy from the empirical, as opposed to the maximum entropy, distribution. Our basic formulation of $\Phi_E$ therefore addresses the in-principle restrictions of $\Phi_{DM}$ mentioned above. $\Phi_E$ 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, $\Phi_E$ 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, $\Phi_{AR}$ (‘auto-regressive $\Phi$’), which is based on auto-regressive prediction error. $\Phi_{AR}$ 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. $\Phi_{AR}$ and $\Phi_E$ 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 $\Phi_E$ in the same way as $\Phi_{AR}$, i.e., in terms of predictive ability. Importantly, although for non-Gaussian systems $\Phi_{AR}$ and $\Phi_E$ may differ, the former remains easy to measure in practice from empirical covariance matrices.

The difference between $\Phi_E/\Phi_{AR}$ and $\Phi_{DM}$ 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 $\Phi_E$, integrated information is a measure of a $\text{process}$, since the empirical distribution is a characterization of the actual behavior of the system. According to $\Phi_{DM}$ integrated information is to some extent a measure of $\text{capacity}$ [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 ITC, that ascribe physical meaning to measures of integrated information. Under the ITC, 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 ITC by considering consciousness as a process; they also challenge the identity relation between consciousness and integrated information assumed in the ITC.

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, $\Phi_{DM}$’ we review $\Phi_{DM}$ using our current notation, noting its limitations especially with respect to discrete Markovian systems. The section ‘The new measure, $\Phi_E$’ describes the new measure $\Phi_E$ 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 $\Phi_E$ remains well-defined even if it is more challenging to calculate. The section ‘$\Phi_E$ for Markovian Gaussian systems’ presents the results of various simulations, designed to illustrate the in-practice applicability of $\Phi_E$ and to explore its properties. We compute $\Phi_E$ for some canonical networks, optimize connectivity under simple dynamics, and examine the numerical stability of the measure. We also compare $\Phi_E$ with a version of $\Phi_{DM}$ modified to apply to continuous systems, showing quantitative congruence in most cases. The section ‘Extension to multiple lags and to MVAR($p$) processes’ describes some additional simulation results, showing how $\Phi_E$ can measure integrated information over arbitrary time-steps (lags). In the section ‘Auto-regressive $\Phi$ ($\Phi_{AR}$)’ we describe $\Phi_{AR}$ and explain its derivation in terms of relations among conditional entropy, covariance, and linear regression prediction error. We demonstrate the utility of $\Phi_{AR}$ 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 $\mathbf{T}$, 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 $X = (X^1, \ldots, X^n)^T$ be a random variable that takes values in the space $\Omega_X$. Then we denote the probability density function by $f_X$, the mean by $\mathbb{E}[X]$ and the $n \times n$ matrix of covariances, $\text{cov}(X^i, X^j)$, by $\Sigma(X)$. Let $Y = (Y^1, Y^2, \ldots, Y^n)^T$ be a second random variable. Then we denote the $n \times m$ matrix of cross-covariances, $\text{cov}(X^i, Y^j)$, by $\Sigma(X, Y)$. The following quantity will be useful:

$$\Sigma(X | Y) = \Sigma(X) - \Sigma(X, Y) \Sigma(Y)^{-1} \Sigma(X, Y)^T. \quad (0.1)$$

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

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

where $\mu_y = \mathbb{E}[X | Y = y] = \mathbb{E}[X] - \Sigma(X, Y) \Sigma(Y)^{-1} (Y - \mathbb{E}[Y])$.

Entropy $H$ characterizes uncertainty, and is given by

$$H(X) = -\sum_{x \in \Omega_X} P_X(x) \log P_X(x), \quad (0.3)$$

if $X$ is a discrete random variable, or

$$H(X) = -\int_{\mathbb{R}^n} P_X(x) \log P_X(x) d^n x \quad (0.4)$$

if $X$ 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 $X$ has a density with respect to the Lebesgue measure $d^n x$; this assumption is upheld whenever we discuss continuous random variables.)

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

$$H(X | Y) = \sum_{y \in \Omega_Y} H(X | Y = y) P_Y(y), \quad (0.5)$$

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

$$I(X; Y) = H(X) - H(X | Y), \quad (0.6)$$

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

$$H(X) = \frac{1}{2} \log |\det \Sigma(X)| + \frac{1}{2} n \log (2\pi e), \quad (0.8)$$

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

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

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

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

$$D_{KL}(P_X || P_Y) = \sum_{x \in \Omega_X} P_X(x) \log \frac{P_X(x)}{P_Y(x)}, \quad (0.11)$$

if the variables are discrete, or

$$D_{KL}(P_X || P_Y) = \int_{\mathbb{R}^n} P_X(x) \log \frac{P_X(x)}{P_Y(x)} d^n x, \quad (0.12)$$

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 $P = \{M^1, \ldots, M^r\}$ divides the elements of $X$ into non-overlapping, non-trivial sub-systems, $X = M^1 \cup \cdots \cup M^r$. The state of $X$ at time $t$ is a $|X|$-dimensional random vector denoted by $X_t$, 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 $\mathcal{X}$, and of the size of this set by $|\mathcal{X}|$. Analogous notation is used for the states of sub-systems of $X$.

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

$$\Gamma_\tau(X) = : \Sigma(X_{t-\tau}, X_t). \quad (0.13)$$

The previous measure, $\Phi_{DM}$

In this section we review, following Ref. [11], the most recent version of $\Phi$ within integrated information theory, using our current notation. This measure, which we call $\Phi_{DM}$ (‘$\Phi$-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 $\Phi_{DM}$, we
brieﬂy discuss these limitations, which motivate our new measures \( \Phi_E \) and \( \Phi_{AR} \).

Let \( X \) be a discrete, Markovian system. \( \Phi_{DM} \) compares the information generated by the whole system to information generated by its parts, when the system transitions to a particular state \( X_1 = x \) from a preceding state \( X_0 \) 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, \( \Phi_{DM}[X:x:P] \), generated by \( X \) being in state \( x \), with respect to the partition \( P = \{ M',\ldots,M' \} \), is given by

\[
\Phi_{DM}[X:x:P] = D_{KL}
\left( P_{X|X_0=x} \parallel \prod_{k=1}^r P_{M'_k|M'_k=m_k} \right).
\] (0.14)

Here \( m' \) is the state of the \( k\)th sub-system of the partition when \( X \) has state \( x \).

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

\[
P_{X_0|X_1=x}(x') = \frac{P_{X_0|X_1=x}(x)P_{X_1}(x')}{P_X(x)}. \] (0.15)

Here \( P_X(x) \) is the maximum entropy distribution, so

\[
P_X(x) = \frac{1}{|S_X|}, \] (0.16)

for all possible initial states \( x' \in S_X \). \( P_{X_0|X_1=x}(x) \) is the conditional probability density for the state at time \( t = 1 \) given that the state at time \( t = 0 \) is \( x' \). 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 \( P_{X_i|X_i=x|} \). 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 \( P_{X_i|X_i=x|} \). Finally, the denominator \( P_X(x) \) is computed from

\[
P_X(x) = \sum_{x' \in S_X} P_{X_0|X_1=x}(x)P_{X_1}(x'). \] (0.17)

For a part \( M \) the analogous Bayes’ rule formula is

\[
P_{M_0|M_0-M'=m'} = \frac{P_{M_0-M'=m'(m_0)}}{P_{M_0}(m)}. \] (0.18)

Here \( P_{M_0} \) is the maximum entropy distribution on \( S_M \). To compute the conditional probability distribution \( P_{M_0-M'=m} \) 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 \( X_0 = (M,N)_1 \). Then we have

\[
P_{M_1|M_0-M'=m}(m,n) = \sum_{m \in S_M} P_{M_0}(m_0,n)(m,n). \] (0.19)

Finally, the denominator

\[
P_{M_1}(m) = \sum_{m \in S_M} P_{M_0-M'=m}(m)P_{M_0}(m). \] (0.21)

Given the probability distributions \( P_{X_0|X_1=x} \) and \( P_{M'_k|M'_k=m_k} \), \( k=1,\ldots,r \), 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, \( \mathcal{P}_{MIP}(x) \), is defined as the partition that minimizes the effective information when it is normalized by

\[
K_M(\{ M',\ldots,M' \}) = (r-1)\min_{\mathcal{P}} [H(M'_k)]. \] (0.22)

Normalization is necessary because sub-systems that are 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 ‘\( \Phi_E \) for Markovian Gaussian systems’. Thus, \( \mathcal{P}_{MIP}(x) \) is given by

\[
\mathcal{P}_{MIP}(x) = \arg\min_{\mathcal{P}} \left\{ \Phi_{DM}[X:x;\mathcal{P}] \right\}. \] (0.23)

Given the MIP, the integrated information \( \Phi_{DM}[X:x] \) generated by the system \( X \) entering state \( x \) is simply the non-normalized effective information with respect to the MIP,

\[
\Phi_{DM}[X:x] = \Phi_{DM}[X:x;\mathcal{P}_{MIP}(x)]. \] (0.24)

Importantly, the value of \( \Phi_{DM}[X:x] \) 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 \( \Phi_{DM} \), one can replace the effective information with its expectation with respect to the current state \( x \), and define the expected integrated information, \( \Phi_{EDM} \), as the expected effective information across the partition that minimizes the normalized expected effective information [11]. The expected effective information, \( \Phi_{EDM} \), is given by [11]...
or equivalently
\[ \phi_{DM}[X; \{ M^1, \ldots, M^r \}] = \sum_{k=1}^{r} H(M_0^k | M_1^k) - H(X_0 | X_1), \quad (0.25) \]

or equivalently
\[ \phi_{DM}[X; \{ M^1, \ldots, M^r \}] = I(X_0; X_1) = \sum_{k=1}^{r} I(M_0^k; M_1^k). \quad (0.26) \]

Note that the second expression (0.26), but not the first (0.25), requires that \( X_0 \) have the maximum entropy distribution [11]. To derive (0.26) from (0.25), one uses that the maximum entropy distribution is uniform, so that
\[ H(X_0) = \sum_{k} H(M_0^k). \quad (0.27) \]

This ensures that one can add \( H(X_0) \) to the second term on the RHS of (0.25) and subtract \( \sum_{k} H(M_0^k) \) from the first term, and then use Eq. (0.6) to obtain the expression (0.26).

We emphasize that \( \phi_{DM} \) 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.) \( \phi_{DM} \) 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 \( P_{X;X_{-\tau}} \) 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 \( \phi_{DM} \) 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 \( \phi_{DM} \) with our new measure \( \Phi_E \) for some example cases.

The new measure, \( \Phi_E \)

The general case. In this section we introduce a new measure of integrated information, \( \Phi_E \), constructed analogously to \( \phi_{DM} \), but with modifications to broaden its applicability, both in theory and in practice. \( \Phi_E \) 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, \( \Phi_E \) instead utilizes the actual distribution of the past state; hence the name \( \Phi_E \). ‘\( \Phi \)-empirical’. This ensures that the measure does not suffer from the in-principle restrictions that pertain to \( \phi_{DM} \), and can be applied to both discrete and continuous systems with either Markovian or non-Markovian dynamics. (More specifically, \( \Phi_E \) will be well-defined as long as the states \( X_i \) of the system are either discrete or have continuous probability densities with respect to a Lebesgue measure \( d^d x \).) A second difference is that, in order to be state-independent, \( \Phi_E \) 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, \( \Phi_E \) is defined so as to enable a choice of timescale (indicated by \( \tau \)) over which integrated information is measured. Thus \( \Phi_E[X; \tau] \) is the integrated information generated by the current state of the system about the state \( \tau \) time-steps in the past.

We now define \( \Phi_E \) for a stochastic system with stationary dynamics. As for \( \phi_{DM} \), \( \Phi_E \) is defined via ‘effective information’. For the new measure we define the effective information generated by the current state \( X_i \) about the state \( \tau \) time-steps ago, with respect to bipartition \( B = \{ M^1, M^2 \} \), 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
\[ \varphi[X; \tau, B] = : I(X_{\tau \rightarrow \tau - 1}; X_i) - \sum_{k=1}^{2} I(M_{\tau \rightarrow \tau - 1}^k; M_i^k). \quad (0.28) \]

The integrated information \( \Phi_E[X; \tau] \) is then the non-normalized effective information with respect to the minimum information bipartition (MIB),
\[ \Phi_E[X; \tau] = : \varphi[X; \tau, B_{MIB}(X; \tau)], \quad (0.29) \]

where
\[ B_{MIB}(X; \tau) = : \arg \min \left\{ \frac{\varphi[X; \tau, B]}{K(B)} \right\}, \quad (0.30) \]

and
\[ K(\{ M^1, M^2 \}) = : \min \left[ H(M_1^1), H(M_2^2) \right]. \quad (0.31) \]

\( \Phi_E \) 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 \( X_{\tau \rightarrow \tau - \tau} \) and \( X_i \), and their joint distribution \( P(X_{\tau \rightarrow \tau - \tau} X_i) \), 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 \( I(X_{\tau \rightarrow \tau - 1}; X_i) \) between the past and current state of the system, and likewise for all sub-systems. Given these quantities, \( \Phi_E \) 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 \( \Phi_E \) empirically under Gaussian assumptions’, this difficulty can be readily overcome if states are Gaussian distributed.

For analytic computation of \( \Phi_E \) given a generative model, we note that the probability distributions for \( X_{\tau \rightarrow \tau - \tau} \) and \( X_i \) individually are both simply equal to the stationary distribution for the state of the system. Obtaining the joint distribution for \( X_{\tau \rightarrow \tau - \tau} \) and \( X_i \)
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 $\Phi_E$ 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 $\Phi_E$ are worth making. First, that $\Phi_E$ 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 $\Phi_E$ via sampling from time-series data. Second, unlike $\Phi_{DM}$, $\Phi_E$ 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 $\Phi_{DM}$ [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 (0.26) can alternatively be read in terms of information generated by the past state $X_{-\tau}$ about the current state $X_{\tau}$.

Our definition (0.26) for the effective information, $\phi$, is based on the expression (0.26) for the expected effective information, $\Phi_{DM}$ in the construction of $\Phi_{DM}$. A viable alternative would be to instead use

$$\tilde{\phi}[X; \tau; \{\mathbf{M}^1, \mathbf{M}^2\}] = \sum_{k=1}^{3} H(\mathbf{M}^k_{t-\tau} | \mathbf{M}^k_t) - H(X_{t-\tau} | X_t),$$

(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 $\phi$ with $\tilde{\phi}$ in the definition of $\Phi_E$ leads to a second measure $\Phi_{E}$. In general, $\tilde{\phi}$ will not be exactly equal to $\phi$. (Equality of their analogues for $\Phi_{DM}$ relies on the past state being maximum entropy, see section ‘The previous measure, $\Phi_{DM}$’.) However, we show in Table 1 that $\Phi_{E}$ behaves very similarly to $\Phi_E$ for the examples we consider in this paper. We choose to focus on $\Phi_{E}$ 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 $\Phi_E$ 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 $\Phi_{DM}$ is only defined for discrete, Markovian systems. As a consequence, $\Phi_E$ but not $\Phi_{DM}$ is applicable to realistic continuous non-Markovian stochastic models of neural systems.

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

$$\phi[X; \tau; \{\mathbf{M}^1, \mathbf{M}^2\}] = \frac{1}{2} \log \left\{ \frac{\det \Sigma(X)}{\det \Sigma(X_{t-\tau} | X_{t})} \right\} - \sum_{k=1}^{3} \frac{1}{2} \log \left\{ \frac{\det \Sigma(\mathbf{M}^k_t)}{\det \Sigma(\mathbf{M}^k_{t-\tau} | \mathbf{M}^k_t)} \right\},$$

(0.33)

and the normalization factor $K$ by

$$K(\{\mathbf{M}^1, \mathbf{M}^2\}) = \frac{1}{2} \log \min\left\{ (2\pi e)^{3k} | \det \Sigma(\mathbf{M}^k) \right\}. \quad (0.34)$$

In practice, the procedure for computing $\Phi_E$ is as follows. First one obtains empirically the covariance matrices $\Sigma(X)$, $\Sigma(X_{t-1} | X_t)$ and analogues for all sub-systems. Then one uses Eq. (0.1) to obtain the partial covariance $\Sigma(X_{t-1} | X_t)$ 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 $\Phi_{E}$, via equations (0.29) and (0.30).

**Computing $\Phi_E$ analytically for a Gaussian system.** In this section we describe analytical computation of $\Phi_E$ 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 MVAR($p$) (multivariate auto-regressive) process [18]

$$X_{\tau} = A_1 X_{t-1} + A_2 X_{t-2} + \cdots + A_p X_{t-p} + E_{\tau}, \quad (0.35)$$

where the $A_i$, $i = 1, \ldots, p$, can be understood as generalized connectivity matrices acting at different time-lags, and $E_{\tau}$ is a stationary multivariate Gaussian white noise source with zero mean and vanishing auto-covariance function, $\Gamma_\tau(E) = 0$, $\tau \neq 0$. (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 $\Phi_E$ for an MVAR(1) system at timescale $\tau = 1$.

Extension to the general $p$, general $\tau$ case is given in the ‘Methods’ section. Consider the generative model

$$X_t = A X_{t-1} + E_t. \quad (0.36)$$

Taking the covariance of both sides of (0.36) gives

$$\Sigma(X) = A \Sigma(X) A^T + \Sigma(E). \quad (0.37)$$

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

$$\Gamma_1(X) \equiv \Sigma(X_{t-1} | X_t) = \langle X_{t-1} (A X_{t-1} + E_{\tau})^T \rangle = \Sigma(X) A^T. \quad (0.38)$$

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

$$\Sigma(X_{t-1} | X_t) = \Sigma(X) - \Gamma_1(X) \Sigma(A)^{-1} \Gamma_1(X)^T. \quad (0.39)$$

Having values for $\Sigma(X)$ and $\Sigma(X_{t-1} | X_t)$ 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 $X_t = (M_t, N_t)^T$. The matrices $\Sigma(X)$ and $\Gamma_1(X)$ can then be written in the form

$$\Sigma(X) = \begin{pmatrix} \Sigma_{M,M} & \Sigma_{M,N} \\ \Sigma_{N,M} & \Sigma_{N,N} \end{pmatrix}, \quad \Gamma_1(X) = \begin{pmatrix} \Gamma_{M,M} & \Gamma_{M,N} \\ \Gamma_{N,M} & \Gamma_{N,N} \end{pmatrix}. \quad (0.40)$$

$$\Sigma_{M,N} = \Sigma_{N,M} = 0.$$
with connectivity as shown in Fig. 1(a)–(g); we refer to these MVAR(1) dynamics for Markovian Gaussian systems.

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

<table>
<thead>
<tr>
<th>Network</th>
<th>(i) $\Phi_E$ (3000 data)</th>
<th>(ii) $\Phi_E$ (10,000 data)</th>
<th>(iii) $\Phi_{E_{\text{ext}}}$ (10,000 data)</th>
<th>(iv) $\Phi_E$</th>
<th>(v) $\Phi_{E_{DM}}$ (3000 data)</th>
<th>(vi) $\Phi_{AR}$ (3000 data)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1(a)</td>
<td>0.0323</td>
<td>0.037 ± 0.004 (10)</td>
<td>0.034 ± 0.003 (10)</td>
<td>0.0323</td>
<td>0.0323</td>
<td>0.038 ± 0.002 (9)</td>
</tr>
<tr>
<td>1(b)</td>
<td>0.0645</td>
<td>0.063 ± 0.004 (10)</td>
<td>0.063 ± 0.003 (10)</td>
<td>0.0645</td>
<td>0.0645</td>
<td>0.061 ± 0.005 (10)</td>
</tr>
<tr>
<td>1(c)</td>
<td>0.1283</td>
<td>0.122 ± 0.008 (10)</td>
<td>0.124 ± 0.003 (10)</td>
<td>0.1387</td>
<td>0.1313</td>
<td>0.125 ± 0.004 (10)</td>
</tr>
<tr>
<td>1(d)</td>
<td>0.0795</td>
<td>0.072 ± 0.006 (10)</td>
<td>0.075 ± 0.004 (10)</td>
<td>0.0894</td>
<td>0.0755</td>
<td>0.073 ± 0.006 (10)</td>
</tr>
<tr>
<td>1(e)</td>
<td>0.1285</td>
<td>0.136 ± 0.012 (10)</td>
<td>0.129 ± 0.002 (10)</td>
<td>0.1376</td>
<td>0.1303</td>
<td>0.135 ± 0.013 (10)</td>
</tr>
<tr>
<td>1(f)</td>
<td>0.1294</td>
<td>0.132 ± 0.008 (10)</td>
<td>0.133 ± 0.004 (10)</td>
<td>0.1383</td>
<td>0.1307</td>
<td>0.131 ± 0.012 (10)</td>
</tr>
<tr>
<td>1(g)</td>
<td>0.1266</td>
<td>0.128 ± 0.010 (6)</td>
<td>0.096 ± 0.006 (4)</td>
<td>0.162</td>
<td>0.1288</td>
<td>0.129 ± 0.013 (7)</td>
</tr>
<tr>
<td>2(a)</td>
<td>0.2502</td>
<td>0.244 ± 0.010 (10)</td>
<td>0.246 ± 0.004 (9)</td>
<td>0.2652</td>
<td>0.254</td>
<td>0.245 ± 0.009 (9)</td>
</tr>
<tr>
<td>2(b)</td>
<td>0.2965</td>
<td>0.291 ± 0.013 (10)</td>
<td>0.2902 ± 0.004 (10)</td>
<td>0.3012</td>
<td>0.2647</td>
<td>0.287 ± 0.010 (7)</td>
</tr>
</tbody>
</table>

Methods of computation are (i) $\Phi_E$ computed analytically; (ii) $\Phi_E$ computed numerically from 10 trials of 3000 data points each; (iii) $\Phi_{E_{\text{ext}}}$ computed numerically from 10 trials of 10,000 data points each; (iv) $\Phi_E$ computed analytically; (v) $\Phi_{E_{DM}}$ computed analytically, and (vi) $\Phi_{AR}$ 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

\[
\Sigma(X) = \begin{pmatrix} \Sigma(X)_{MM} & \Sigma(X)_{MN} \\ \Sigma(X)_{NM} & \Sigma(X)_{NN} \end{pmatrix},
\]

\[
\Gamma_1(X) = \begin{pmatrix} \Gamma_1(X)_{MM} \\ \Gamma_1(X)_{NM} \end{pmatrix} = \begin{pmatrix} \Sigma(X)_{MM} & -
\sum_{\tau=1}^{\infty} \Gamma_1(X)_{MM} \sum_{\tau=1}^{\infty} \Gamma_1(X)_{MM}^{-1} \Gamma_1(X)_{MM}^{-T}.
\]

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 $\Phi_E$ are obtained from Eqs. (0.29) and (0.30).

**$\Phi_E$ for Markovian Gaussian systems**

**Canonical examples.** We present results from computing $\Phi_E$, 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

\[ X_t = AX_{t-1} + E_t, \]

where $X_t$ contains 8 variables, $A$ is the connectivity matrix, and each component of $E_t$ is an independent Gaussian random variable of mean 0 and variance 1. We considered seven systems, with connectivity as shown in Fig. 1(a)–1(g); we refer to these systems ‘1(a),’ ‘1(b),’ and so on. The corresponding values of $\Phi_E$ are given in Fig. 1(h) and Table 1. For analytic computation, we performed the procedure described in the section ‘Computing $\Phi_E$ analytically for a Gaussian system.’ For simulated measurements, we first obtained time-series data from equation (0.43), and then computed $\Phi_E$ using the recipe described in the section ‘Computing $\Phi_E$ 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 $\Phi_E$ lay within $\approx 1$ 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 $\Phi_E$ value. Perhaps in contrast to expectations, adding sparse long-range ‘short-cut’ connections to a reciprocal ring (1(c)–1(g)), in the style of a so-called ‘small world’ network [23,24,25], does not increase $\Phi_E$ (compare with network 1(c)).

For values of $\Phi_E$ 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 $\Phi_E$ from time-series data is stable for these networks.

Network 1(g) exhibited instability when measuring $\Phi_E$ from simulation. As shown in Fig. 1(h), the corresponding values of $\Phi_E$ fell close to one of two values, one of which was the true
(analytically derived) value. For simulations of 3,000 data points 6/10 trials produced $\Phi_E$ 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 ($\varphi$) in identifying the MIB, but non-normalized $\varphi$ in specifying the corresponding value of $\Phi_E$. Given finite data, estimates of $\varphi$ cannot be guaranteed to be accurate. As a result, inter-trial variation in measuring $\Phi_E$ from data can arise when (i) there are two (or more) partitions with similar values of normalized $\varphi$ close to the true minimum (used to identify the MIB), and (ii) these partitions have substantially different values for non-normalized $\varphi$. The latter condition will typically hold when partitions with similar normalized $\varphi$ have significantly different sub-system sizes (see the section 'The previous measure, $\Phi_{DM}$'). 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 $\varphi$ is 0.0213.

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 $\Phi_E$ ($\tau = 1$) 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
However, there is an uneven bipartition, \( \{1,2,3,4,5\}, \{6,7,8\} \) for which the normalized \( \phi \) is 0.0218, i.e., very similar to the value of \( \phi \) for the true MIB. However, the non-normalized \( \phi \) for the MIB (i.e., \( \Phi_{\text{DM}} \)) is 0.1206, whereas the value for the uneven bipartition is 0.0966. Fig. 1(h) and Table 1 show that empirical measurements of \( \Phi_{\text{E}} \) cluster around these two values.

One may consider that this problem of instability could be avoided by using non-normalized \( \phi \) to identify the MIB. However, as discussed in the section ‘The previous measure, \( \Phi_{\text{DM}} \),’ in this case \( \Phi_{\text{E}} \) 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 \( \Phi_{\text{E}} \) in terms of normalized \( \phi \). However, in this case the meaning of \( \Phi_{\text{E}} \) 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 \( \Phi_{\text{E}} \). To examine whether network structures other than reciprocally connected rings could generate high levels of \( \Phi_{\text{E}} \), we performed numerical optimizations using a genetic algorithm (GA). Specifically, we used \( \Phi_{\text{E}} (r=1) \) 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 \( \Phi_{\text{E}} \). 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 \( \Phi_{\text{E}} \)) 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 \( \Phi_{\text{E}} = 0.2502 \), 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 \( \Phi_{\text{E}} = 0.2965 \), 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 \( \Phi_{\text{E}} \) 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 \( \Phi_{\text{E}} \) 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 \( \Phi_{\text{E}} \) to single mutations. Figs. 3(c) and 3(d) show the percentage decrease in \( \Phi_{\text{E}} \) 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 30% or more. Together, these observations show that the value of \( \Phi_{\text{E}} \) generated by a network is highly sensitive to small changes in topology and connection strength, further pointing to the ruggedness of the \( \Phi_{\text{E}} \) 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 \( \Phi_{\text{E}} \).
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 the effective information of 0.0424. We incrementally weakened the connection at the point at which the uneven bipartition becomes the MIB (see Fig. 3(c)).

**Comparison with \( \Phi_{DM} \), \( \Phi_E \), and full table of MVAR(1) results.** It is instructive to compare results obtained using \( \Phi_E \) with those obtained from the version of \( \Phi_{DM} \) extended to apply to stationary continuous (but still Markovian) systems (see sections ‘The previous measure, \( \Phi_{DM} \)’, and ‘Methods’). Table 1 shows (extended) \( \Phi_{DM} \) values for the various networks discussed above, as well as the corresponding \( \Phi_E \) 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 \( \Phi_E \) as a valid and useful measure of integrated information.

The network 2(a) has a value for \( \Phi_E \) that is approximately double that of the corresponding \( \Phi_{DM} \). 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 \( \Phi_{DM} \) 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 \( \Phi_E \)’.

We also compared results obtained using \( \Phi_E \) with those obtained using \( \Phi_{E} \), 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, \( \Phi_E \) 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 \( \Phi_E \) was computed analytically, via the method outlined in ‘Computing \( \Phi_E \) analytically for a Gaussian system’ and ‘Methods’. Fig. 4(a) shows \( \Phi_E \) measured for various values of \( \tau \), (where \( \tau \) 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 \( \Phi_E \) 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 \( \Phi_E \) as a function of \( \tau \) for the MVAR(3) process

\[
X_t = A_1 X_{t-1} + A_2 X_{t-2} + A_3 X_{t-3} + E_t,
\]

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, \( \Phi_E \) 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 \( \Phi_E \) landscape with respect to network connectivity.** (a) Histogram of \( \Phi_E \) for the 30 networks in the final population of a GA that yielded optimal network 2(a). (b) Histogram of \( \Phi_E \) for the 30 networks in the final population of a GA that yielded optimal network 2(b). (c) Histogram of percentage decrease in \( \Phi_E \) following single mutations of network 2(a) (200 evaluations). (d) Histogram of percentage decrease in \( \Phi_E \) following single mutations of network 2(b) (200 evaluations). (e) Discontinuity in \( \Phi_E \) 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
that $\Phi_e$ 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 $\Phi$ ($\Phi_{AR}$)**

We have presented a measure of integrated information, $\Phi_e$, that is practical to measure from time-series data under Gaussian assumptions. However, in the case of stationary, non-Gaussian distributed time-series, $\Phi_e$ 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 $\Phi_e$ under Gaussian assumptions can nonetheless lead to a meaningful quantity reflecting integrated information. We call this quantity $\Phi_{AR}$ (‘auto-regressive $\Phi$’). By construction, $\Phi_{AR}$ is equivalent to $\Phi_e$ 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 $\Phi_{AR}$ 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 $X$ and $Y$ be two multivariate random variables. Then the linear regression of $X$ on $Y$ is the expression

$$X = \alpha + A \cdot Y + E,$$  \hspace{1cm} (0.45)

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

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

$$\alpha = \bar{x} - A \cdot \bar{y}. $$ \hspace{1cm} (0.47)

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

$$\Sigma(E) = \Sigma(X|Y),$$ \hspace{1cm} (0.48)

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

$$H(X|Y=y) = \frac{1}{2} \log |\det \Sigma(E)| + \frac{1}{2} \log(2\pi e),$$ \hspace{1cm} (0.49)

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

$$\varphi[X;\tau;\{M_1, M_2\}] = \frac{1}{2} \log \left\{ \frac{\det \Sigma(X)}{\det \Sigma(E)} \right\} - \sum_{k=1}^{2} \frac{1}{2} \log \left\{ \frac{\det \Sigma(M_k)}{\det \Sigma(E_M^k)} \right\},$$ \hspace{1cm} (0.50)

where $E_{M_k}^k$, $k=1,2$, and $E_X$ are the residuals in the regressions

$$M_{t-1}^k = A^k - M_{t}^k + E_{t}^k,$$ \hspace{1cm} (0.51)

$$X_{t-1} = A^X \cdot X_t + E_{X}^t.$$ \hspace{1cm} (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 $\Phi_{AR}$, which we now define. Let $X$ be a stationary, not necessarily Gaussian, system, and let $\varphi_{AR}[X;\tau;\{M_1, M_2\}]$ be the RHS of Eq. (0.50), i.e.

$$\varphi_{AR}[X;\tau;\{M_1, M_2\}] = \frac{1}{2} \log \left\{ \frac{\det \Sigma(X)}{\det \Sigma(E)} \right\} - \sum_{k=1}^{2} \frac{1}{2} \log \left\{ \frac{\det \Sigma(M_k)}{\det \Sigma(E_M^k)} \right\},$$ \hspace{1cm} (0.53)

where $E_{M_k}^k$, $k=1,2$, and $E_X$ are the residuals in the regressions (0.51) and (0.52). Then $\Phi_{AR}$ is simply $\varphi_{AR}$ for the bipartition that minimizes $\varphi_{AR}$ divided by the normalization factor.
\[
L(\{M^1, M^2\}) = \frac{1}{2} \log \min \left\{ (2\pi e)^{\frac{k}{2}} \det \Sigma(k) \right\}.
\]

Thus,

\[
\Phi_{\text{AR}}[X; \tau] = : \frac{\varphi_{\text{AR}}[X; \tau, E^{\min}(\tau)]}{L(B)}.
\]

(0.55)

For Gaussian systems, \( \Phi_E \) and \( \Phi_{\text{AR}} \) 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 \( \Phi_E \) under Gaussian assumptions (as laid out in ‘Computing \( \Phi_E \) empirically under Gaussian assumptions’) yields precisely \( \Phi_{\text{AR}} \). Notably, this recipe implies that it is not necessary to explicitly carry out the linear regressions; rather, the equivalence (0.48) shows that \( \Phi_{\text{AR}} \) can be calculated using empirical covariance matrices.

\( \Phi_{\text{AR}} \) is meaningful as a measure of integrated information because of its formulation in terms of linear regression prediction error. \( \Phi_{\text{AR}} \) 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, \( \Phi_{\text{AR}} \) 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 \( \Phi_{\text{AR}} \) in terms of (backwards) prediction becomes exactly equivalent to the interpretation of \( \Phi_E \) in terms of Shannon information. Note that in fact, by the symmetry of mutual information (0.7), (0.28), \( \Phi_{\text{AR}} \) could also be expressed in terms of entirely analogous linear regressions in which the present is used to predict the future. Understood this way, \( \Phi_{\text{AR}} \) 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 \( \Phi_{\text{AR}} \) as distinct from \( \Phi_E \), 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 \( E_i \) 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 \( \Phi_{\text{AR}} \) empirically from 10 trials of 3000 data points each. The results, shown in Table 1, suggest that in each case \( \Phi_{\text{AR}} \) for the non-Gaussian dynamics is approximately equal to \( \Phi_E \) (\( \approx \Phi_{\text{AR}} \)) for the Gaussian dynamics. This finding provides support for \( \Phi_{\text{AR}} \) as a useful alternative to \( \Phi_E \), applicable to non-Gaussian dynamics.

**Discussion**

In this paper we have presented two new measures of integrated information, \( \Phi_E \) and \( \Phi_{\text{AR}} \). As with a previous measure, \( \Phi_{\text{DM}} \), 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 \( \Phi_{\text{DM}} \) 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 (\( \Phi_E \)), and (ii) to interpret integrated information in terms of predictive ability of the present of a system with respect to the past (\( \Phi_{\text{AR}} \)). Simulations showed that our measures conform to intuitions regarding conjoint dynamical integration and segregation; where comparisons could be made, in most cases our measures quantitatively aligned with \( \Phi_{\text{DM}} \). 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 \( \Phi \) 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 \( \Phi_{\text{DM}} \) 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 $\Phi_{DM}$, 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 $\Phi_{DM}$ 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 $\Phi_{DM}$ 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 $\Phi_E$ 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 $\Phi_{DM}$ but not $\Phi_E$ 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 $\Phi_E$ 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 $\Phi$ is computed. To compute $\Phi_{DM}$, 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 $\Phi_E$ 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**

$\Phi_E$ 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 $\Phi_E$ 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 $\Phi_E$ 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 $\Phi_E$ 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 $\Phi_{AR}$. This is constructed analogously to $\Phi_E$, but with information replaced by the reduction in the generalized covariance of the past state under prediction via linear regression on the current state. $\Phi_{AR}$ 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. $\Phi_{AR}$ and $\Phi_E$ are equivalent for Gaussian systems, but otherwise differ; (recall however that $\Phi_{AR}$ can be obtained for any system by using the recipe for computing $\Phi_E$ for a Gaussian system). In our examples, $\Phi_{AR}$ was in fact insensitive to a change from Gaussian noise to exponentially distributed noise, supporting its use as an alternative to $\Phi_E$.

**Normalization and instability**

All versions of $\Phi$ require a normalization step. Specifically, $\Phi$ is determined by the non-normalized effective information ($\phi$) across a minimum information bipartition (MIB) which is specified as the bipartition which minimizes the normalized $\phi$ (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 $\Phi$. On the other hand, it remains important to determine the value of $\Phi$ using the non-normalized $\phi$ in order to allow $\Phi$ to be interpreted as a quantity of information.

The use of normalization, as just described, leads to instabilities. Our simulations have shown that $\Phi_E$ 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 $\phi$ close to the true minimum and these partitions had substantially different values of non-normalized $\phi$. This instability does not arise for all systems, and indeed for most of our examples $\Phi_E$ is numerically stable. Nonetheless, the embedding of normalization within the definition of $\Phi$ challenges ascription of physical meaning to any measured value of $\Phi$. This is because the value of $\Phi$ 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 ($\Phi_C$ and $\Phi_{DM}$) 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 $\Phi_{DM}$, then one must interpret consciousness (integrated information) as a function of state transitions \cite{11}; accordingly, one cannot ask about the conscious level of a system \textit{per se}. By contrast, if one applies $\Phi_P$ or $\Phi_{AR}$ 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 \cite{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 \cite{33}, whereas epileptic absence seizures are characterized instead by increased synchrony in thalamocortical systems \cite{39}. As mentioned in the section ‘Empirical and maximum entropy distributions’, neural dynamics may be metastable \cite{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.) \cite{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’ \cite{41} and ‘neural complexity’ \cite{8} (see ‘Comparison with causal density and neural complexity’).

An important feature of the IITC as previously expressed is that consciousness \textit{qua} $\Phi$ is best considered as a capacity (equivalently a potential, or disposition), and not as an ‘object’ or a process \cite{14}. The original $\Phi_C$ operationalized the notion of capacity by subjecting a system to all possible perturbations and examining its responses. The recent $\Phi_{DM}$ 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 $\Phi_{DM}$ 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 $\Phi$ 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. $\Phi_P$ and $\Phi_{AR}$ can therefore be construed as measures of a process modulated by capacity, aligning more closely with the Jamesian intuition.

The notion that $\Phi$ exists as a ‘fundamental property’ deserves comment. As described in the section ‘Normalization and instability’, our results challenge the ascription of physical meaning to $\Phi$, in virtue of its exquisite sensitivity to the normalization process involved in specifying the MIB: this challenge pertains equally to the notion of $\Phi$ as a ‘fundamental quantity’. A further challenge to the ascription of physical meaning to $\Phi$ 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 $\Phi$ could be taken to define ‘natural’ coordinates, or macroscopic variables, for the system. In any case, it does not seem necessary to consider $\Phi$ 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 $\Phi$ as accounting for particular aspects of consciousness without the further step of claiming identity \cite{9}.

### Integrated information in other neurocognitive processes

Although $\Phi$ 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 ($\Phi_{DM}$, $\Phi_C$) 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 $\Phi$ 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, multi-modal binding and perceptual categorization \cite{20}, and action selection (decision making) \cite{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 \cite{6,41}. Our results now allow integrated information to be applied in similar situations, facilitating comparative analyses.

### Comparison with causal density and neural complexity

$\Phi$ 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’ \cite{41} and ‘neural complexity’ \cite{42}. Here, we briefly summarize the similarities and differences among these measures, in order to set $\Phi$ into a broader context.

Causal density, like $\Phi_{AR}$ and $\Phi_E$ (but in contrast to $\Phi_{DM}$ and $\Phi_C$), is a measure of process rather than capacity. In virtue of being based on ‘Granger causality’, it also shares with $\Phi$ 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 $X^{3..N}$) \cite{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 $\Phi_{AR}$. In a loose sense, integrated information,
as measured by $\Phi_{\text{AR}}$ or $\Phi_{\text{EE}}$, 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 $\Phi_{\text{AR}}$ and causal density.

Neural complexity is calculated as the sum of the average mutual information across all bipartitions of a system [42]. Unlike $\Phi$ 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 $\Phi_{\text{C}}$.) 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. $\Phi$ by contrast considers only one past state. An interesting further modification to $\Phi$ 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 $\Phi$. 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 $\Phi_{\text{EE}}$ and $\Phi_{\text{AR}}$, as well as functions which allow regeneration of some of the simulations we describe.

Extension and computation of $\Phi_{\text{DM}}$ for an MVAR(1) process

To extend $\Phi_{\text{DM}}$ 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 $X_0 \sim \mathcal{N}(0,\Sigma_{\text{DM}}(X))$, where

$$\Sigma_{\text{DM}}(X)_{ij} = \begin{cases} \Sigma(X)_{ij}, & i=j, \\ 0, & i \neq j. \end{cases} \quad (0.57)$$

Having defined a distribution for the initial state $X_0$, we explain how to compute the expected integrated information, $\Phi_{\text{DM}}$, 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, $\Omega = \Sigma(F)$. Alternatively, an estimate of $\Phi_{\text{DM}}$ from time-series data can be obtained by using estimates of $A$ and $\Omega$. The linear-regression formulae (0.46) and (0.48) yield the estimates

$$A = \tilde{\Sigma}(X_{1},X_{t-1}) \tilde{\Sigma}(X)^{-1}, \quad (0.58)$$
\[
\hat{\Omega} = \Sigma(X_{1}|X_{t-1}).
\]  
(0.59)

where the symbol \(\hat{\Omega}\) denotes empirical quantities.

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

\[
\Sigma(X) = A\Sigma(X)A^T + \Omega,
\]
(0.60)

and \(\Sigma^D\) from Eq. (0.57).

To compute the conditional probability \(P(X_0|X_{t-1})\) we first use the MVAR(1) dynamics (0.36) to obtain the distribution of \(X_1|X_0=x\) as

\[
X_1|\{X_0=x\} \sim \mathcal{N}(Ax, \Omega).
\]
(0.61)

Then we use Bayes' rule (0.15) to obtain

\[
P(X_0|X_{t-1} = x) \propto P(X_t|X_{t-1} = x)P(X_{t-1} = x).
\]
(0.62)

\[
\propto \exp\left\{ -\frac{1}{2} \left[ (x-Ax)^T \Omega^{-1} (x-Ax) + x^T (\Sigma^D)^{-1} x \right] \right\}. 
\]
(0.63)

From the term quadratic in \(x\) we can obtain the inverse of the covariance matrix of (the Gaussian distributed) conditional variable \(X_0|X_1=x\) as

\[
\Sigma(X_0|X_1=x)^{-1} = A^T \Omega^{-1} A + (\Sigma^D)^{-1},
\]
(0.64)

and hence express the conditional entropy \(H(X_0|X_1=x)\) in terms of the connectivity and stationary covariance matrices:

\[
H(X_0|X_1=x) = \frac{1}{2} |X| \log(2\pi e) - \frac{1}{2} \log \left\{ \det \left[ A^T \Omega^{-1} A + (\Sigma^D)^{-1} \right] \right\}.
\]
(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 \(X = (M_t, N_{jt})^T\) so that

\[
\Sigma^D = \begin{pmatrix} \Sigma_M^D & 0 \\ 0 & \Sigma_N^D \end{pmatrix}, \quad A = \begin{pmatrix} A_{MM} & A_{MN} \\ A_{NM} & A_{NN} \end{pmatrix},
\]
(0.66)

and similarly for \(\Omega\) and \(E\). To obtain the distribution for the conditional random variable \(M_1|M_0=m\), we express \(M_1\) in terms of \(M_0\) as

\[
M_1 = A_{MM}M_0 + A_{MN}N_0 + E_{M1},
\]
(0.67)

and note that \(N_0 \sim \mathcal{N}(0, \Sigma_N^D)\). Hence

\[
M_1|\{M_0=m\} \sim \mathcal{N}(A_{MM}m, \Omega_{MM} + A_{MN} \Sigma^D_N A_{MN}^T).
\]
(0.68)

From Bayes' rule, we can then calculate the inverse of the covariance matrix of (the Gaussian distributed) conditional variable \(M_0|M_1=m\) as

\[
\Sigma(M_0|M_1=m)^{-1} = A_{MM}^T (\Omega_{MM} + A_{MN} \Sigma^D_N A_{MN}^T)^{-1} A_{MM} + \Sigma_{M1}^{-1},
\]
(0.69)

and hence

\[
H(M_0|M_1=m) = \frac{1}{2} |M| \log(2\pi e) - \frac{1}{2} \log \left\{ \det \left[ A_{MM}^T (\Omega_{MM} + A_{MN} \Sigma^D_N A_{MN}^T)^{-1} A_{MM} + \Sigma_{M1}^{-1} \right] \right\}.
\]
(0.70)

The entropy formulae (0.65) and (0.70) furnish the sufficient quantities for computing \(\PhiDM\) as described in the section ‘The previous measure, \(\PhiDM\)’, using the expression (0.25) for the expected effective information. For present purposes, as with \(\PhiE\), we restrict attention to bipartitions only.

Analytical computation of \(\PhiE\) for a general Gaussian case

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

\[
X_t = A_1X_{t-1} + A_2X_{t-2} + \cdots + A_pX_{t-p} + E_t,
\]
(0.71)

where the \(A_i\), \(i=1,\ldots,p\), can be thought of as generalized connectivity matrices acting at different time-lags, and \(E_t\) is a stationary multivariate Gaussian ‘white noise’ source with zero mean and vanishing auto-covariance function, \(\Gamma_t(E) = 0\), \(\tau \neq 0\). (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

\[
det \left( I_{(p)} - \sum_{i=1}^{p} \lambda_i^t A_i \right) = 0
\]
(0.72)

lie outside the unit circle [33].

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

\[
\Xi_t = F\Xi_{t-1} + V_t,
\]
(0.73)

involving the block quantities \(\Xi = \{X_t, X_{t-1}, \ldots, X_{t-p}\}\), \(V_t = \{E_t, 0, 0, \ldots, 0\}\) and

\[
F = \begin{pmatrix}
A_1 & A_2 & \cdots & A_p & O_{[X]} & \cdots & O_{[X]} & O_{[X]} \\
I_{[X]} & O_{[X]} & \cdots & O_{[X]} & O_{[X]} & \cdots & O_{[X]} & O_{[X]} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\
O_{[X]} & O_{[X]} & \cdots & O_{[X]} & O_{[X]} & \cdots & I_{[X]} & O_{[X]} \\
\end{pmatrix},
\]
(0.74)

The stationary covariance matrix \(\Sigma(\Xi)\) for this process can be obtained from the Lyapunov equation, by analogy with \(\Sigma(X)\) for
the MVAR(1) case (0.37):

\[ \Sigma(\Xi) = F(\Sigma(\Xi))F^T + \Sigma(V), \]

(0.75)

where \( \Sigma(V) = \text{diag}[\Sigma(V_1), \Sigma(V_2), \ldots, \Sigma(V_N)] \). Then the stationary covariance \( \Sigma(X) \) and auto-covariance \( \Gamma_t(X) \) 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:

\[ \Sigma(X_{t-\tau}|X_t) = \Sigma(X) - \Gamma_t(X)\Sigma(X)^{-1}\Gamma_t(X)^T, \]

(0.76)

\[ \Sigma(X) = \left( \begin{array}{cc} \Sigma(X)_{MM} & \Sigma(X)_{MN} \\ \Sigma(X)_{NM} & \Sigma(X)_{NN} \end{array} \right), \]

\[ \Gamma_t(X) = \left( \begin{array}{cc} \Gamma_t(X)_{MM} & \Gamma_t(X)_{MN} \\ \Gamma_t(X)_{NM} & \Gamma_t(X)_{NN} \end{array} \right), \]

(0.77)

\[ \Sigma(M) = \Sigma(X)_{MM}, \quad \Gamma_t(M) = \Gamma_t(X)_{MM}. \]

(0.78)

The above expressions furnish the quantities needed to compute \( \Phi_E \) from equations (0.29), (0.30), (0.33) and (0.34).

### Supporting Information

**Text S1** Toolbox for computing integrated information as \( \Phi_E \) or \( \Phi_{AR} \). 'phiemvarp.m' computes \( \Phi_E \) from an MVAR(\( \beta \)) generative model, ‘ARphidata.m’ computes \( \Phi_{AR} \) (= \( \Phi_E \) if Gaussian), from stationary time-series data. ‘statdata.m’ creates time-series data from an MVAR(\( \beta \)) 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.


