PLOS | COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

Measuring Integrated Information from the Decoding Perspective

Masafumi Oizumi[1,2]*, Shun-ichi Amari[1], Toru Yanagawa[1], Naotaka Fujii[1], Naotsugu Tsuchiya[2,3,4]*

1 RIKEN Brain Science Institute, Wako, Saitama, Japan, **2** School of Psychological Sciences, Faculty of Biomedical and Psychological Sciences, Monash University, Clayton, Victoria, Australia, **3** Japan Science and Technology Agency, Kawaguchi, Saitama, Japan, **4** Monash Institute of Cognitive and Clinical Neurosciences, Monash University, Clayton, Victoria, Australia

* oizumi@brain.riken.jp (MO); naotsugu.tsuchiya@monash.edu (NT)

Abstract

Accumulating evidence indicates that the capacity to integrate information in the brain is a prerequisite for consciousness. Integrated Information Theory (IIT) of consciousness provides a mathematical approach to quantifying the information integrated in a system, called integrated information, Φ. Integrated information is defined theoretically as the amount of information a system generates as a whole, above and beyond the amount of information its parts independently generate. IIT predicts that the amount of integrated information in the brain should reflect levels of consciousness. Empirical evaluation of this theory requires computing integrated information from neural data acquired from experiments, although difficulties with using the original measure Φ precludes such computations. Although some practical measures have been previously proposed, we found that these measures fail to satisfy the theoretical requirements as a measure of integrated information. Measures of integrated information should satisfy the lower and upper bounds as follows: The lower bound of integrated information should be 0 and is equal to 0 when the system does not generate information (no information) or when the system comprises independent parts (no integration). The upper bound of integrated information is the amount of information generated by the whole system. Here we derive the novel practical measure Φ* by introducing a concept of mismatched decoding developed from information theory. We show that Φ* is properly bounded from below and above, as required, as a measure of integrated information. We derive the analytical expression of Φ* under the Gaussian assumption, which makes it readily applicable to experimental data. Our novel measure Φ* can generally be used as a measure of integrated information in research on consciousness, and also as a tool for network analysis on diverse areas of biology.

Author Summary

Integrated Information Theory (IIT) of consciousness attracts scientists who investigate consciousness owing to its explanatory and predictive powers for understanding the neural

Citation: Oizumi M, Amari S-i, Yanagawa T, Fujii N, Tsuchiya N (2016) Measuring Integrated Information from the Decoding Perspective. PLoS Comput Biol 12 (1): e1004654. doi:10.1371/journal.pcbi.1004654

Editor: Daniel Polani, University of Hertfordshire, UNITED KINGDOM

Received: May 9, 2015

Accepted: November 13, 2015

Published: January 21, 2016

Data Availability Statement: The details on the recording apparatus and the dataset is available online (http://neurotycho.org).

Funding: MO was supported by a Grant-in-Aid for Young Scientists (B) from the Ministry of Education, Culture, Sports, Science, and Technology of Japan (26870860) (https://www.jsps.go.jp/english/). NT was supported by Precursory Research for Embryonic Science and Technology from Japan Science and Technology Agency (3630) (http://www.jst.go.jp/EN/), Future Fellowship (FT120100619) and Discovery Project (DP130100194) from Australian Research Council (http://www.arc.gov.au/). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

properties of consciousness. IIT predicts that the levels of consciousness are related to the quantity of information integrated in the brain, which is called integrated information Φ. Integrated information measures excess information generated by a system as a whole above and beyond the amount of information independently generated by its parts. Although IIT predictions are indirectly supported by numerous experiments, validation is required through quantifying integrated information directly from experimental neural data. Practical difficulties account for the absence of direct, quantitative support. To resolve these difficulties, several practical measures of integrated information have been proposed. However, we found that these measures do not satisfy the theoretical requirements of integrated information: First, integrated information should not be below 0; and second, integrated information should not exceed the quantity of information generated by the whole system. Here, we propose a novel practical measure of integrated information, designated as Φ^* that satisfies these theoretical requirements by introducing the concept of mismatched decoding developed from information theory. Φ^* creates the possibility of empirical and quantitative validations of IIT to gain novel insights into the neural basis of consciousness.

Introduction

Although its neurobiological basis remains unclear, consciousness may be related to certain aspects of information processing [1, 2]. In particular, Integrated Information Theory of consciousness (IIT) developed by Tononi and colleagues [2–9] predicts that the amount of information integrated among the components of a system, called integrated information Φ, is related to the level of consciousness of the system. The level of consciousness in the brain varies from a very high level, as in full wakefulness, to a very low level, as in deeply anesthetized states or dreamless sleep. When consciousness changes from high to low, IIT predicts that the amount of integrated information changes from high to low, accordingly. This prediction is indirectly supported by recent neuroimaging experiments that combine noninvasive magnetic stimulation of the brain (transcranial magnetic stimulation, TMS) with electrophysiological recordings of stimulation-evoked activity (electroencephalography) [10–14]. Such evidence implies that if there is a practical method to estimate the amount of integrated information from neural activities, we may be able to measure levels of consciousness using integrated information.

IIT provides several versions of mathematical formulations to calculate integrated information [2–8]. Although the detailed mathematical formulations are different, the central philosophy of integrated information does not vary among different versions of IIT. Integrated information is mathematically defined as the amount of information generated by a system as a whole above and beyond the amount of information generated independently by its parts. If the parts are independent, no integrated information should exist.

Despite its potential importance, the empirical calculation of integrated information is difficult. For example, one difficulty involves making an assumption when integrated information is calculated according to the informational relationship between the past and present states of a system. The distribution of the past states is assumed to maximize entropy, which is called the maximum entropy distribution. The assumption of the maximum entropy distribution severely limits the applicability of the original integrated information measure Φ as indicated by [15]. First, the concept of the maximum entropy distribution cannot be applied to a system that comprises elements whose states are continuous, because there is no unique maximum entropy distribution for continuous variables [15, 16]. Second, information under the assumption of the maximum entropy distribution can be computed only when there is complete knowledge about the transition probability matrix that describes how the system transits between states. However, the transition probability matrix for actual neuronal systems is practically impossible to estimate.

To overcome these problems, Barrett and Seth [15] proposed using the empirical distribution estimated from experimental data, thereby removing the requirement to rely on the assumption of the maximum entropy distribution. Although we believe that their approach does lead to practical computation of integrated information, we found that their proposed measures based on the empirical distribution [15] do not satisfy key theoretical requirements as a measure of integrated information. Two theoretical requirements should be satisfied as a measure of integrated information. First, the amount of integrated information should not be negative. Second, the amount of integrated information should never exceed information generated by the whole system. These theoretical requirements, which are satisfied by the original measure Φ, are required so that a measure of integrated information is interpretable in accordance with the original philosophy of integrated information.

Here, we propose a novel practical measure of integrated information, Φ^*, by introducing the concept of mismatched decoding developed from information theory [17–20]. Φ^* represents the difference between "actual" and "hypothetical" mutual information between the past and present states of the system. The actual mutual information corresponds to the amount of information that can be extracted about the past states by knowing the present states (or vice versa) when the actual probability distribution of a system is used for decoding. In contrast, hypothetical mutual information corresponds to the amount of information that can be extracted about the past states by knowing the present states when the "mismatched" probability distribution is used for decoding where a system is partitioned into hypothetical independent parts. Decoding with a mismatched probability distribution is called mismatched decoding. Φ^* quantifies the amount of loss of information caused by the mismatched decoding where interactions between the parts are ignored. We show here that Φ^* satisfies the theoretical requirements as a measure of integrated information. Further, we derive the analytical expression of Φ^* under the Gaussian assumption and make this measure feasible for practical computation. We also compute Φ^* and the previously proposed measures in electrocorticogram (ECoG) data recorded in monkeys to demonstrate that the previous measures violate the theoretical requirements even in real brain recordings.

Results

While its central ideas are unchanged, IIT updated measures of integrated information. The original formulation, IIT 1.0 [2], underwent major developments leading to IIT 2.0 [6] and the latest version IIT 3.0 [8]. In the present study, we focus on the version in IIT 2.0 [3, 6], because the measure of integrated information proposed in IIT 2.0 is simpler and more feasible to calculate compared with that in IIT 3.0 [5, 8].

Here, we briefly review the original measure of integrated information, Φ, in IIT 2.0 [3, 6] and describe its limitations for practical application [15]. From the concept of the original measure, we point out the lower and upper bounds that a measure of integrated information should satisfy. We introduce next two practical measures of integrated information, Φ_I and Φ_H, proposed by [15] and show that Φ_I and Φ_H fail to satisfy the lower and upper bounds of integrated information. Finally, we derive a novel measure of integrated information, Φ^*, from the decoding perspective, which is properly bounded from below and above.

Intrinsic information and extrinsic information

In IIT, information refers to intrinsic information as opposed to extrinsic information (See S1 Text for details). Intrinsic information is quantified from the intrinsic perspective of a system itself and only depends on internal variables of the system. On the other hand, extrinsic information is quantified from the extrinsic perspective of an external observer and depends on external variables. For example, in neuroscience, extrinsic information is quantified as mutual information between neural states X and external stimuli S, $I(X;S)$ [21–24]. In contrast, intrinsic information can be quantified by the mutual information between the past states $X^{t-\tau}$ and the present states X^t of the system, $I(X^{t-\tau};X^t)$. The mutual information, $I(X^{t-\tau};X^t)$, is expressed by

$$I(X^{t-\tau};X^t) = H(X^{t-\tau}) - H(X^{t-\tau}|X^t), \tag{1}$$

where $H(X^{t-\tau})$ is the entropy of the past states and $H(X^{t-\tau}|X^t)$ is the conditional entropy of the past states given the present states. In IIT, the distribution of the past states is assumed to be the maximum entropy distribution so that the entropy of the past states is maximized, i.e., the past states are maximally uncertain. We can interpret that intrinsic information, $I(X^{t-\tau};X^t)$, quantifies to what extent uncertainty of the past states can be reduced by knowing the present states from the system's intrinsic point of view. IIT considers such quantity as the amount of information intrinsically generated by the system.

Measure of integrated information with the maximum entropy distribution

Consider partitioning a system into m parts such as $M_1, M_2, \cdots$, and M_m and computing the quantity of information that is integrated across the m parts of a system. As detailed in S1 Text, the measure of integrated information proposed in IIT 2.0 can be expressed as follows:

$$\Phi = I({}^{\max}X^{t-\tau};X^t) - \sum_{i=1}^{m} I({}^{\max}M_i^{t-\tau};M_i^t), \tag{2}$$

where the superscript$^{\max}$ indicates that the distribution of the past states is the maximum entropy distribution. The first term of Eq 2, $I({}^{\max}X^{t-\tau};X^t)$, represents the mutual information between the past and present states in the whole system, and the second term represents the sum of the mutual information between the past and present states in the i-th part of the system $I({}^{\max}M_i^{t-\tau};M_i^t)$. Thus, Φ, the difference between them, gives the information generated by the whole system above and beyond the information generated independently by its parts. If the parts are independent, no extra information is generated, and the integrated information is 0. We can rewrite Eq 2 in terms of entropy H as follows:

$$\Phi = \sum_{i=1}^{m} H({}^{\max}M_i^{t-\tau}|M_i^t) - H({}^{\max}X^{t-\tau}|X^t). \tag{3}$$

To derive the above expression, we use the fact that the entropy of the whole system $H({}^{\max}X^{t-\tau})$ equals the sum of the entropy of the subsystems $\sum_{i=1}^{m} H({}^{\max}M_i^{t-\tau})$ when the maximum entropy distribution is assumed.

Theoretical requirements as a measure of integrated information

To interpret a measure of integrated information as the "extra" information generated by a system as a whole above and beyond its parts, it should satisfy theoretical requirements, as follows: First, integrated information should not be negative because information independently generated by the parts should never exceed information generated by the whole. Integrated information should equal 0 when the amount of information generated by the whole system equals 0 (no information) or when the amount of information generated by the whole is equal to that generated by its parts (no integration). Second, integrated information should not exceed the amount of information generated by the whole system because the information generated by the parts should not be negative. In short, integrated information should be lower-bounded by 0 and upper-bounded by the information generated by the whole system.

One can check the original measure Φ satisfies the lower and upper bounds.

$$0 \leq \Phi \leq I({}^{\max}X^{t-\tau}; X^{t}). \tag{4}$$

As shown in S1 Text, Φ can be written as the Kullback-Leibler divergence. Thus, Φ is positive or equal to 0. Further, as can be seen from Eq 2, the upper bound of Φ is the mutual information in the entire system, because the sum of mutual information in the parts is larger than or equal to 0.

Practical measures of integrated information with empirical distribution. The original measure Φ assumes the distribution of the past states to be the maximum entropy distribution, which limits the practical application of Φ for two reasons. First, the maximum entropy distribution can be applied only when the states of a system are discrete. If the states are represented by discrete variables, the maximum entropy distribution is the uniform distribution over all possible states of $X^{t-\tau}$. When the states of a system are described by continuous variables, the maximum entropy distribution cannot be uniquely defined [15, 16]. Second, the transition probability matrix of a system, $p(X^{t}|X^{t-\tau})$ must be known for all possible past states $X^{t-\tau}$ for obtaining the mutual information $I({}^{\max} X^{t-\tau};X^{t})$. However, it is nearly impossible to estimate such a complete transition probability matrix experimentally in an actual neural system, because some states may not occur during a reasonable period of observation.

A simple remedy for the limitations of the original measure Φ is not to impose the maximum entropy distribution on the past states but instead to use the probability distributions obtained from empirical observations of the system. Barrett and Seth [15] adopted this strategy to derive two practical measures of integrated information from Eqs 2 and 3 by substituting the maximum entropy distribution with the empirical distribution as follows:

$$\Phi_{I} = I(X^{t-\tau}; X^{t}) - \sum_{i=1}^{m} I(M_{i}^{t-\tau}; M_{i}^{t}), \tag{5}$$

$$\Phi_{H} = \sum_{i=1}^{m} H(M_{i}^{t-\tau}|M_{i}^{t}) - H(X^{t-\tau}|X^{t}). \tag{6}$$

Note that Φ_I and Φ_H are not equal when the empirical distribution is used for the past states, because the entropy of the whole system $H(X^{t-\tau})$ is not equal to the sum of the entropy of the subsystems, $\sum_{i} H(M_{i}^{t-\tau})$. Φ_H was also derived from a different perspective from IIT, i.e. the perspective of information geometry, as a measure of spatio-temporal interdependencies and is termed "stochastic interaction" [25, 26].

Although these two measures appear as natural modifications of the original measure, they do not satisfy the theoretical requirements as a measure of integrated information. We discuss the problems of Φ_I and Φ_H in detail later.

Integrated information measure based on mismatched decoding. Here, we propose an alternative practical measure of integrated information that satisfies the theoretical requirements which we call Φ^{*} (phi star) (Fig 1). Φ^{*}, which uses the empirical distribution, can be applied to actual neuronal recordings. Similar to Φ_I, we will derive Φ^{*} based on the original measure Φ in Eq 2 based on mutual information. Given the problem of Φ_I in Eq 5, we should refine the second term of Eq 5, while the first term, the mutual information in the whole

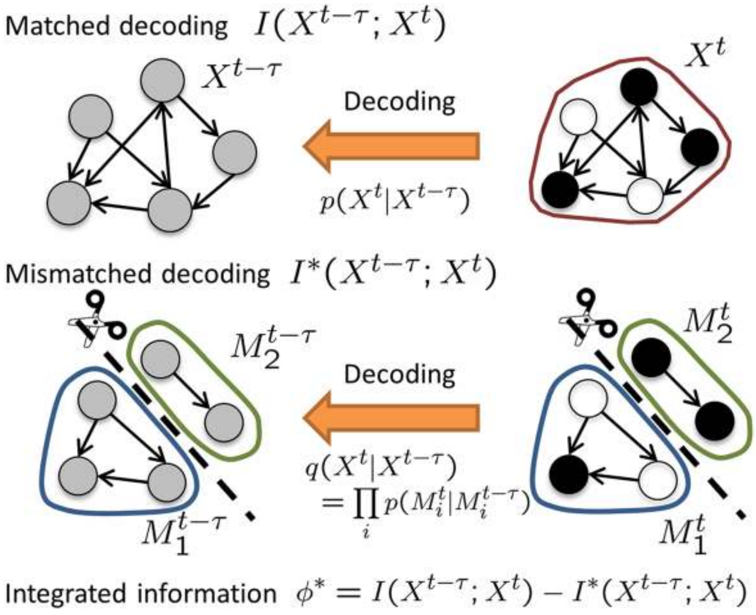
Matched decoding $I(X^{t-\tau};X^t)$
$X^{t-\tau}$
Decoding
$p(X^t|X^{t-\tau})$
X^t
Mismatched decoding $I^*(X^{t-\tau};X^t)$
$M_2^{t-\tau}$
$M_1^{t-\tau}$
Decoding
$q(X^t|X^{t-\tau})$
$=\prod_i p(M_i^t|M_i^{t-\tau})$
M_2^t
M_1^t
Integrated information $\phi^* = I(X^{t-\tau};X^t) - I^*(X^{t-\tau};X^t)$

Fig 1. Integrated information based on the concept of mismatched decoding. The figure shows a system with five neurons in which the arrows represent directed connectivity and the colors represent the states of the neurons (black: silence, white: firing, gray: unknown). The past states $X^{t-\tau}$ are decoded given the present states X^t. The "true" conditional distribution $p(X^t|X^{t-\tau})$ is used for matched decoding, while a "false" conditional distribution $q(X^t|X^{t-\tau})$ is used for mismatched decoding where the parts of a system M_1 and M_2 are assumed independent. The amount of information about the past states that can be extracted from the present states using matched and mismatched decoding is quantified by the mutual information $I(X^{t-\tau};X^t)$ and the "hypothetical" mutual information $I^*(X^{t-\tau};X^t)$ for mismatched decoding, respectively. In this framework, integrated information, $\Phi^*(X^{t-\tau};X^t)$, is defined as the difference between $I(X^{t-\tau};X^t)$ and $I^*(X^{t-\tau};X^t)$.

doi:10.1371/journal.pcbi.1004654.g001

system, is unchanged. The second term should be a quantity that can be interpreted as information generated independently by the parts of a system and should be less than information generated by the system as a whole.

To derive a proper second term in Eq 5, we interpret the mutual information from a decoding perspective and introduce the concept of "mismatched decoding", which was developed by information theory [17] (see S1 Text for details). Consider that the past states $X^{t-\tau}$ are decoded given the present states X^t. From the decoding perspective, the mutual information can be interpreted as the maximum information about the past states that can be obtained knowing the present states. To extract the maximum information, the decoding must be performed optimally using the "true" conditional distribution,

$$p(X^t|X^{t-\tau}) = p(M_1^t, \cdots, M_m^t|M_1^{t-\tau}, \cdots, M_m^{t-\tau}). \tag{7}$$

Note that the expression on the right explicitly accounts for interactions among all the parts. The optimal decoding can be performed using the maximum likelihood estimation. In the above setting, the maximum likelihood estimation chooses the past state that maximizes $p(X^t|X^{t-\tau})$ given a present state. Decoding that uses the true distribution, $p(X^t|X^{t-\tau})$, is called "matched decoding" because the probability distribution used for decoding matches the actual probability distribution.

Decoding that uses a "false" conditional distribution, $q(X^t|X^{t-\tau})$, is called "mismatched" decoding. To quantify integrated information, we consider specifically the mismatched decoding that uses the "partitioned" probability distribution $q(X^t|X^{t-\tau})$,

$$q(X^t|X^{t-\tau}) = \prod_{i=1}^{m} p(M_i^t|M_i^{t-\tau}), \tag{8}$$

where a system is partitioned into parts and the parts M_i are assumed to be independent. $q(X^t|X^{t-\tau})$ is the product of the conditional probability distribution in each part $p(M_i^t|M_i^{t-\tau})$. The distribution, $q(X^t|X^{t-\tau})$, is "mismatched" with the actual probability distribution, because parts are generally not independent in reality. As is matched decoding, mismatched decoding is also performed using the maximum likelihood estimation, wherein the past state that maximizes $q(X^t|X^{t-\tau})$ is selected. The amount of information obtained from mismatched decoding is necessarily degraded compared with that obtained from matched decoding. The best decoding performance can be achieved only when matched decoding is used with the actual probability distribution $p(X^t|X^{t-\tau})$.

We consider the amount of information that can be obtained from mismatched decoding, $I^*(X^{t-\tau};X^t)$, as a proper second term of Eq 5 (see Methods for the mathematical expression of I^*). The difference between $I(X^{t-\tau};X^t)$ and $I^*(X^{t-\tau};X^t)$ provides a new practical measure of integrated information (Fig 1),

$$\Phi^*(X^{t-\tau};X^t) = I(X^{t-\tau};X^t) - I^*(X^{t-\tau};X^t). \tag{9}$$

Φ^* quantifies the information loss caused by mismatched decoding where a system is partitioned into independent parts, and the interactions between the parts are ignored. Φ^* satisfies the theoretical requirements, because I^* is greater than or equal to 0 and is less than or equal to the information in the whole system I. Φ^* is equivalent to the original measure Φ if the maximum entropy distribution is imposed on the past states instead of an empirical distribution (see S1 Text for the proof). Thus, we can consider Φ^* as a natural extension of Φ to the case when the empirical distribution is used.

Analytical computation of Φ* using Gaussian approximation

Although using an empirical distribution instead of the maximum entropy distribution makes integrated information feasible to calculate, it is still difficult to compute Φ^* in a large system, because the summation over all possible states must be calculated. The number of all possible states grows exponentially with the size of the system and therefore, computational costs for computing Φ^* also grow exponentially. Thus, for practical calculation of Φ^*, we need to approximate Φ^* in some way such as approximating the probability distribution of neural states using the Gaussian distribution [15]. Φ^* can be analytically computed using the Gaussian approximation (see Methods). The Gaussian approximation significantly reduces the computational costs and makes Φ^* practically computable even in a large system.

Theoretical requirements are not satisfied by previously proposed measures

In this section, by considering two extreme cases, we demonstrate that the previously proposed measures Φ_H and Φ_I[15] do not satisfy either the lower or upper bound.

When there is no information. First, we consider the case where there is no information between the past and present states of a system, i.e. $I(X^{t-\tau};X^t) = 0$. In this case, integrated information should be 0. As expected, Φ^* and Φ_I are 0, because the amount of information for mismatched decoding, $I^*(X^{t-\tau};X^t)$, and the mutual information in each part, $I(M_i^{t-\tau};M_i^t)$, are both

0 when $I(X^{t-\tau};X^t) = 0$;

$$\Phi^* = 0, \tag{10}$$

$$\Phi_I = 0. \tag{11}$$

However, Φ_H is not 0. Φ_H can be written as

$$\Phi_H = \sum_i H(M_i^{t-\tau}) - H(X^{t-\tau}). \tag{12}$$

Φ_H is not 0 even when the information $I(X^{t-\tau};X^t)$ is 0 because Φ_H is not based on the mutual information but on the conditional entropy (see Eq 6). Therefore, Φ_H does not necessarily reflect the amount of information in a system.

As a simple example that shows the above problem of Φ_H, consider the following linear regression model,

$$X^t = AX^{t-1} + E^t. \tag{13}$$

Here, X is the state of units, A is a connectivity matrix, and E^t is multivariate Gaussian noise with zero mean and covariance $\Sigma(E)$. E^t is uncorrelated over time. For simplicity, consider a system composed of two units (the following argument can be easily generalized to a system with more than two units). We set the connectivity matrix A and the covariance matrix of noise $\Sigma(E)$ as follows:

$$A = a\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \tag{14}$$

$$\Sigma(E) = \begin{pmatrix} 1 & c \\ c & 1 \end{pmatrix}, \tag{15}$$

where a and c are parameters that control the strengths of connections and noise correlation, respectively. We compute measures of integrated information using the above model. The time difference τ is set to 1. We assume that the prior distribution of the system is the steady state distribution, where the covariance of the past states, $\Sigma(X^{t-1})$, and that of the present states, $\Sigma(X^t)$, are equal, i.e. $\Sigma(X^{t-1}) = \Sigma(X^t) = \Sigma(X)$. The covariance of the steady state distribution $\Sigma(X)$ can be calculated by taking the covariance of both sides of Eq 13,

$$\Sigma(X) = A\Sigma(X)A^T + \Sigma(E). \tag{16}$$

We consider a case where the connection strength a is 0. Fig 2 shows an exemplar time series when the strength of noise correlation c is 0.9. Because there are no connections, including self-connections within each unit, each unit has no information between the past and present states, i.e., $I_1 = I_2 = 0$. As can be seen from Fig 2, however, the two time series correlate at each moment because of the high noise correlation.

We varied the degree of noise correlation, c, from 0 to 1 while keeping the connection strength a as 0 (Fig 3(A)). Φ^* and Φ_I stay 0 independent of noise correlation. However, an entropy-based measure, Φ_H, increases monotonically with c, irrespective of the amount of information in the whole system (Fig 3(A)). As shown in Eq 12, Φ_H is the difference between the sum of entropy within each part and entropy in the whole system. When the parts correlate, the entropy in the whole system decreases. In contrast, the sum of entropy of each part does not change, because the degree of noise within each part (the diagonal elements of E^t) is fixed.

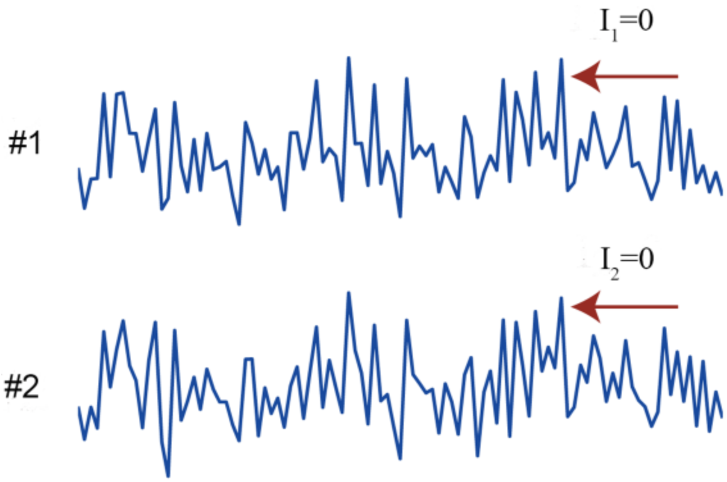

Fig 2. Exemplar time series when there is no information between the past and present states. The connection strength *a* and the strength of noise correlation *c* are set to 0 and 0.9, respectively in the linear regression model (Eq 13). I_1 and I_2 represent the mutual information in units 1 and 2. Because there is no connection, there is no information between the past and present states of the system: I_1 and I_2 are both 0. In this case, Φ* and Φ_I are 0 as they should be, yet Φ_H is positive.

doi:10.1371/journal.pcbi.1004654.g002

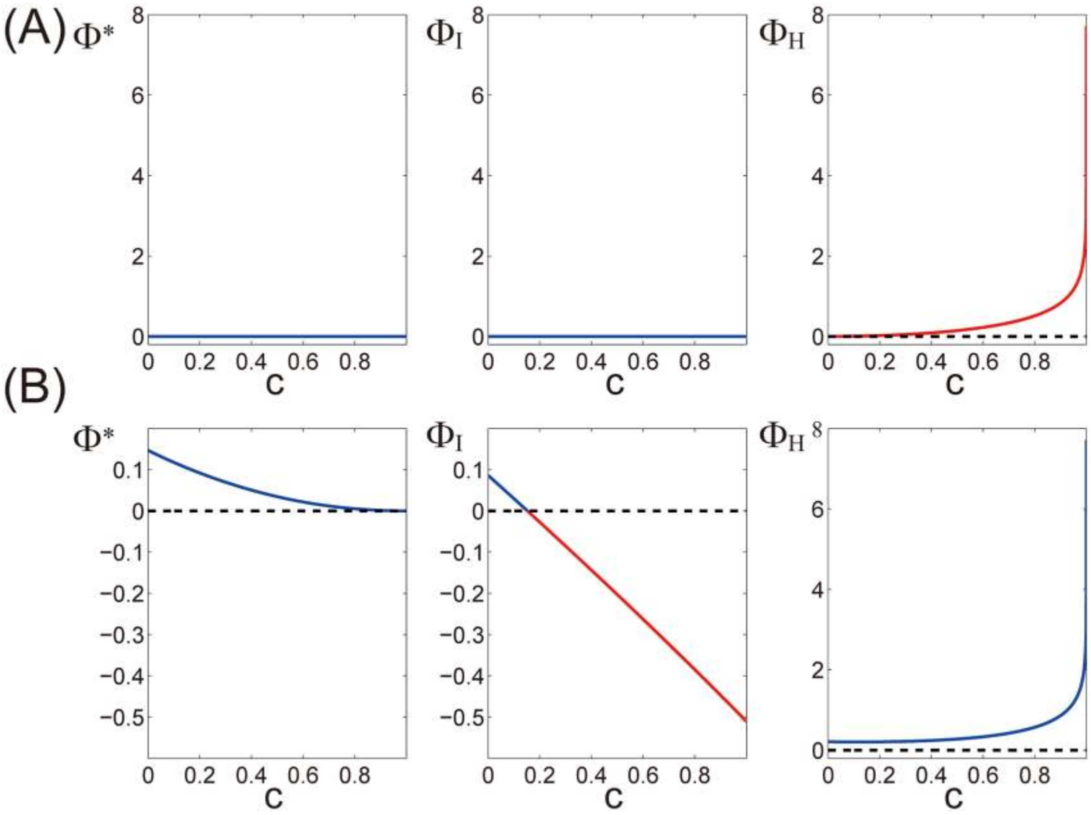

Fig 3. Violation of theoretical requirements as a measure of integrated information. The behaviors of Φ*, Φ_I, and Φ_H are shown in the left, middle, and right panels, respectively, when the strength of noise correlation *c* is varied in a linear regression model (Eq 13). Red lines indicate the regime where the theoretical requirements are violated, and the blue lines indicate that the theoretical requirements are satisfied. Dotted black lines are drawn at 0. (A) Violation of the upper bound. The strength of connections *a* is set to 0. In this case, there is no information between the past and present states of the system but Φ_H is not 0, i.e., Φ_H violates the upper bound. (B) Violation of the lower bound. The strength of connections *a* is set to 0.4. At the right ends of the figures where *c* is 1, the two units in the system are perfectly correlated. Φ_I is negative, i.e., Φ_I violates the lower bound when the degree of correlation is high.

doi:10.1371/journal.pcbi.1004654.g003

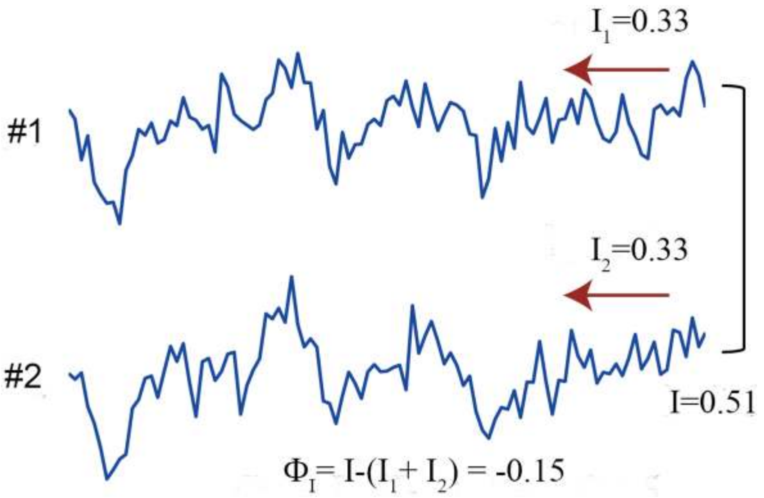

Fig 4. Exemplar time series when correlation is high. The strength of noise correlation *c* and the connection strength *a* are set to both 0.4 in the linear regression model (Eq 13). I_1 and I_2 represent the mutual information in unit 1 and 2, and *I* represents the mutual information in the whole system. In this case, the sum of the mutual information in the parts exceeds the mutual information in the whole system and Φ_I is negative.

doi:10.1371/journal.pcbi.1004654.g004

Thus, Φ_H increases as the degree of noise correlation *c* increases without reflecting the amount of information in the system.

When parts are perfectly correlated. Next, we consider the case where the parts are perfectly correlated. More specifically, consider the case where the two parts M_1 and M_2 are equal at every time, i.e. $M_1^{t-\tau} = M_2^{t-\tau} = M^{t-\tau}$ and $M_1^t = M_2^t = M^t$. Here, Φ^* is 0 because the amount of information extracted by mismatched decoding would not degrade even if the other part is ignored for decoding (see S1 Text for the mathematical proof).

$$\Phi^* = 0. \tag{17}$$

Regarding Φ_I, the mutual information of each part is equal to each other, $I(M_1^{t-\tau}; M_1^t) = I(M_2^{t-\tau}; M_2^t) = I(M^{t-\tau}; M^t)$ and the mutual information in the whole system is equal to the mutual information of each part, $I(X^{t-\tau};X^t) = I(M^{t-\tau};M^t)$. Thus, the second term in Eq 5 is twice the value of the first, and Φ_I is the negative value of the mutual information in one part,

$$\Phi_I = -I(M^{t-\tau}; M^t). \tag{18}$$

Thus, Φ_I does not satisfy the lower bound as a measure of integrated information. Φ_H is given by

$$\Phi_H = H(X^{t-\tau}|X^t) - 2H(M^{t-\tau}|M^t), \tag{19}$$

which is larger than or equal to 0 (Φ_H is always larger than or equal to 0 because it can be written as the Kullback-Leibler divergence.).

We considered again the same linear regression model presented in the previous section (Eq 13). We varied the degree of noise correlation, *c*, from 0 to 1 while keeping connection strength *a* as 0.4. When *c* is 1, the two units correlate perfectly. Fig 4 shows an exemplar time series when *c* is 0.4 and *a* is 0.4. Φ_I takes positive values when *c* is less than ~ 0.2 but takes negative values when *c* is greater (Fig 3(B)). Φ^* decreases monotonically with *c* and becomes 0 when *c* is

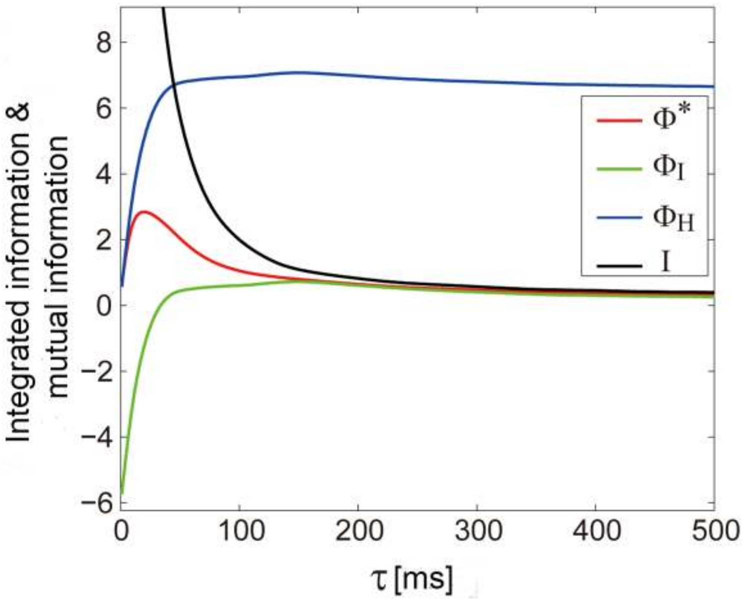

Fig 5. Measures of integrated information and mutual information computed in monkey ECoG data. Time lag τ is varied from 1 to 500 ms. The behaviors of Φ* (red line), Φ_I (green line), Φ_H (blue line), and mutual information *I* (black line) are shown. Φ_I and Φ_H violate the theoretical requirements.

doi:10.1371/journal.pcbi.1004654.g005

1. Φ_H increases monotonically with c reflecting the degree of correlation between the units. The detailed behaviors of Φ^*, Φ_I and Φ_H when a and c are both varied are shown in S1 Fig.

Electrocorticogram data analysis. The problems of Φ_H and Φ_I can manifest in their application to real neural recordings from the brain. Fig 5 shows the measures of integrated information, Φ^*, Φ_I, Φ_H, and the mutual information I computed from the electrocorticogram (ECoG) recordings in an awake monkey as a function of the time lag τ (See Methods for details).

As we can see, the mutual information between X^t and $X^{t-\tau}$ monotonically decreases as τ increases. Φ^* is positive, peaks around $\tau = 20$ ms, and less than the mutual information, always satisfying the theoretical requirements. However, Φ_I is negative when τ is small and Φ_H remains large even when I approaches 0 with increasing τ, both violating the theoretical requirements.

Discussion

In this study, we consider the two theoretical requirements that a measure of integrated information should satisfy, as follows: The lower and upper bounds of integrated information should be 0 and the amount of information generated by the whole system, respectively. The theoretical requirements are naturally derived from the original philosophy of integrated information [3, 6], which states that integrated information is the information generated by a system as a whole above and beyond its parts. The original measure of integrated information Φ satisfies the theoretical requirements so that we can interpret a measure of integrated information according to the original philosophy. To derive a practical measure of integrated information that satisfies the required lower and upper bounds, we introduced a concept of mismatched decoding. We defined our measure of integrated information Φ^* as the amount of information lost when a mismatched probability distribution, where a system is partitioned into "independent" parts, is used for decoding instead of the actual probability distribution. In this framework, Φ^* quantifies the amount of information loss associated with mismatched decoding where interactions between the parts of a system are ignored and therefore quantifies the amount of information integrated by the interactions. We show that Φ^* satisfies the lower and upper bounds, that Φ_I does not satisfy the lower bound, and that Φ_H does not satisfy the upper bound. We consider Φ^* a proper measure of integrated information that can be generally used for practical applications.

Here, we briefly note a potential reason why the previous study [15] failed to identify these problems of Φ_I and Φ_H. Although they calculated their measures in small networks by using the autoregressive model in Eq 12, they did not extensively vary the connectivity matrix A and the Gaussian noise E. In particular, they fixed the covariance of the Gaussian noise E to 0. As we can clearly see in Fig 3 and S1 Fig, both connectivity strength a and the covariance of the noise c strongly affect the amount of integrated information. In particular, when the covariance of E is large, Φ_I and Φ_H violate the theoretical requirements. For future investigations of calculating integrated information in networks described by autoregressive model, we should note that it is very important to take account of not only the effects of connectivity matrix A but also the effects of covariance of E on the amount of integrated information.

The basic concept of Integrated Information Theory (IIT) was tested by conducting empirical experiments, and the evidence accumulated supports the conclusion that when consciousness is lost, integration of information is lost [10–14]. In particular, Casali and colleagues [14] found that a complexity measure, motivated by IIT, successfully separates conscious awake states from various unconscious states due to deep sleep, anesthesia, and traumatic brain injuries. Although their measure is inspired by the concept of integrated information, it measures the complexity of averaged neural responses to one particular type of external perturbation (e.g. a TMS pulse to a target region) and does not directly measure integrated information.

There are few studies that directly estimate integrated information in the brain [27, 28] using the measure introduced in IIT 1.0 [2] or Φ_H. Our new measure of integrated information, Φ^*, will contribute to experiments designed to test whether integrated information is a key to distinguishing conscious states from unconscious states [29–31].

We considered the measure of integrated information proposed in IIT 2.0 [3, 6], because its computations are feasible. There are several updates in the latest version, IIT 3.0 [8]. In IIT 2.0, integrated information is quantified by measuring how the distribution of the past states differs when a present state is given (see S1 Text for details) whereas in IIT 3.0, it is quantified by measuring how the distribution of the past and future states differs when a present state is given. In other words, IIT 2.0 considers only the information flow from the present to the past while IIT 3.0 additionally considers the information flow from the present to the future. Our measure Φ^* does not asymmetrically quantify integrated information from the present to the past or from the present to the future, because the mutual information is a symmetric measure for the time points $t - \tau$ and t. An unanswered question is how integrated information should be practically calculated taking account of the both directions of information flow, using an empirical distribution.

An unresolved difficulty that impedes practical calculation of integrated information is how to partition a system. In the present study, we considered only the quantification of integrated information when a partition of a system is given. IIT requires that integrated information should be quantified using the partition where information is least integrated, called the minimum information partition (MIP) [3, 6]. To find the MIP, every possible partition must be examined, yet the number of possible partitions grows exponentially with the size of the system. One way to work around this difficulty would be to develop optimization algorithms to quickly find a partition that well approximates the MIP.

Besides the practical problem of finding the MIP, there remains a theoretical problem of how to compare integrated information across different partitions. Integrated information increases as the number of parts gets larger, because more information is lost by partitioning the system. Further, integrated information is expected to be larger in a symmetric partition where a system is partitioned into two parts of equal size than in an asymmetric partition. IIT 2.0 [6] proposes a normalization factor, which considers these issues. However, there might be other possible ways to perform normalization. It is unclear whether there is a reasonable theoretical foundation that adjudicates the best normalization scheme. Moreover, it is unclear if the normalization factor, which is proposed for systems whose states are represented by discrete variables, is appropriate for systems whose states are represented by continuous variables. The normalization factor, which is based on the entropies of the parts of a system, can be negative because entropy can be negative for continuous variables. Thus, we need a different normalization factor when we deal with continuous variables. Further investigations are required to resolve the practical and theoretical issues related to the MIP.

Although we derived Φ^*, because we were motivated by IIT and its potential relevance to consciousness, Φ^* has unique meaning from the perspective of information theory, which is independent of IIT. Thus, it can be applied to research fields other than research on consciousness [32]. Φ^* quantifies the loss of information when interactions or connections between the units in a system are ignored. Thus, Φ^* is expected to be related to connectivity measures such as Granger causality [33] or transfer entropy [34]. It will be interesting to clarify mathematical relationships between Φ^* and the other connectivity measures. We expect that information geometry [25, 26, 35, 36] plays an important role for studying the properties of these quantities. Here, we indicate only an apparent difference between them as follows: Φ^* intends to measure global integrations in a system as a whole, while traditional bivariate measures such as Granger causality or transfer entropy intends to measure local interactions between elements of the system. Consider that we divide a system into parts *A*, *B*, and *C*. Using integrated information, our goal is to quantify the information integrated among *A*, *B*, and *C* as a whole. In contrast, what we quantify using Granger causality or transfer entropy is the influence of *A* on *B*, *B* on *C*, *C* on *A* and the reverse. It is not obvious how a measure of global interactions in the whole system should be defined and derived theoretically from measures of the local interactions. As an example, one possibility is simply summing up all local interactions and considering the sum as a global measure [37]. Yet, more research is required to determine whether such an approach is a valid method to define global interactions [36]. Φ^*, in contrast, is not derived from the local interaction measures but is derived directly by comparing the total mutual information in the whole system with hypothetical mutual information when the system is assumed to be partitioned into independent parts. Thus, the interpretation of Φ^* is straightforward from an information theoretical viewpoint. Our measure, which we consider a measure of the global interaction, may provide new insights into diverse research subjects as a novel tool for network analysis.

Methods

Mathematical expression of *I**

The amount of information for mismatched decoding can be evaluated using the following equation,

$$I^*(X^{t-\tau};X^t) = -\sum_{X^t} p(X^t)\log\sum_{X^{t-\tau}} p(X^{t-\tau})q(X^t|X^{t-\tau})^{\beta} + \sum_{X^{t-\tau},X^t} p(X^{t-\tau},X^t)\log q(X^t|X^{t-\tau})^{\beta}, \tag{20}$$

where β is the value that maximizes I^*. The maximization of I^* with respect to β is performed by differentiating I^* and solving the equation, $dI^*(\beta)/d\beta = 0$. In general, the solution of the equation can be found using the standard gradient ascent method, because I^* is a convex function with respect to β[17, 18].

For comparison, the mutual information is given by

$$I(X^{t-\tau};X^t) = -\sum_{X^t} p(X^t)\log p(X^t) + \sum_{X^{t-\tau},X^t} p(X^{t-\tau},X^t)\log p(X^t|X^{t-\tau}). \tag{21}$$

If a mismatched probability distribution $q(X^t|X^{t-\tau})$ is replaced by the actual distribution $p(X^t|X^{t-\tau})$ in Eq 20, the derivative of I^* becomes 0 when $\beta = 1$. By substituting $q = p$ and $\beta = 1$ into Eq 20, one can check that I^* is equal to I in Eq 21, as it should be. The amount of information for mismatched decoding, I^*, was first derived in the field of information theory as an extension of the mutual information in the case of mismatched decoding [17]. I^* was first introduced into neuroscience in [18] and was first applied to the analysis of neural data by [19]. However, I^* in the prior neuroscience application [18, 19] was quantified between stimuli and neural states, not between the past and present states of a system, as described in the present study.

Analytical computation of Φ* under the Gaussian assumption

Assume that the probability distribution of neural states X is the Gaussian distribution,

$$p(\mathrm{X}) = \frac{1}{\left((2\pi)^N|\Sigma(X)|\right)^{1/2}}\exp\left(-\frac{1}{2}(\mathrm{X}-\bar{\mathrm{X}})^T\Sigma(X)^{-1}(\mathrm{X}-\bar{\mathrm{X}})\right). \tag{22}$$

where N is the number of variables in X, $\bar{\mathrm{X}}$ is the mean value of X, and $\Sigma(X)$ is the covariance matrix of X. The Gaussian assumption allows us to analytically compute Φ^*, which substantially reduces the costs for computing Φ^*. When $X^{t-\tau}$ and X^t are both multivariate Gaussian variables, the mutual information between $X^{t-\tau}$ and X^t, $I(X^{t-\tau};X^t)$, can be analytically computed as

$$I(X^{t-\tau};X^t) = \frac{1}{2}\log\frac{|\Sigma(X^{t-\tau})|}{|\Sigma(X^{t-\tau}|X^t)|}, \tag{23}$$

where $\Sigma(X^{t-\tau}|X^t)$ is the covariance matrix of the conditional distribution, $p(X^{t-\tau}|X^t)$, which is expressed as

$$\Sigma(X^{t-\tau}|X^t) = \Sigma(X^{t-\tau}) - \Sigma(X^{t-\tau},X^t)\Sigma(X^t)^{-1}\Sigma(X^{t-\tau},X^t)^T, \tag{24}$$

where $\Sigma(X^{t-\tau}, X^t)$ is the cross covariance matrix between $X^{t-\tau}$ and X^t, whose element $\Sigma(X^{t-\tau}, X^t)_{ij}$ is given by $\mathrm{cov}(X_i^{t-\tau}, X_j^t)$.

Similarly, we can obtain the analytical expression of I^* as follows:

$$I^*(\beta) = \frac{1}{2}\mathrm{Tr}\left(\Sigma(X^t)R\right) + \frac{1}{2}\log\left(|Q||\Sigma(X^{t-\tau})|\right) - \frac{\beta N}{2}, \tag{25}$$

where Tr stands for trace. Q and R are given by

$$Q = \Sigma(X^{t-\tau})^{-1} + \beta\Sigma_D(X^{t-\tau})^{-1}\Sigma_D(X^t,X^{t-\tau})^T\Sigma_D(X^t|X^{t-\tau})^{-1}\Sigma_D(X^t,X^{t-\tau})\Sigma_D(X^{t-\tau})^{-1}, \tag{26}$$

$$\begin{aligned} R &= \beta\Sigma_D(X^t|X^{t-\tau})^{-1} \\ &-\beta^2\Sigma_D(X^t|X^{t-\tau})^{-1T}\Sigma_D(X^t,X^{t-\tau})\Sigma_D(X^{t-\tau})^{-1}Q^{-1}\Sigma_D(X^{t-\tau})^{-1}\Sigma_D(X^t,X^{t-\tau})^T\Sigma_D(X^t|X^{t-\tau})^{-1}, \end{aligned} \tag{27}$$

where $\Sigma_D(X^{t-\tau})$, $\Sigma_D(X^t, X^{t-\tau})$ and $\Sigma_D(X^t|X^{t-\tau})$ are diagonal block matrices. Each block matrix is a covariance matrix of each part, $\Sigma(M_i^{t-\tau})$, $\Sigma(M_i^t, M_i^{t-\tau})$, and $\Sigma(M_i^t|M_i^{t-\tau})$ where M_i is a subsystem. For example, $\Sigma_D(X^{t-\tau})$ is given by

$$\Sigma_D(X^{t-\tau}) = \begin{pmatrix} \Sigma(M_1^{t-\tau}) & & & \\ & \Sigma(M_2^{t-\tau}) & & 0 \\ 0 & & \ddots & \\ & & & \Sigma(M_m^{t-\tau}) \end{pmatrix}. \tag{28}$$

The maximization of I^* with respect to β is performed by solving the equation $dI^*(\beta)/d\beta = 0$. The derivative of $I^*(\beta)$ with respect to β is given by

$$\frac{dI^*(\beta)}{d\beta} = \frac{1}{2}\mathrm{Tr}\left(\Sigma(X^t)\frac{dR}{d\beta}\right) + \frac{1}{2}\mathrm{Tr}\left(Q^{-1}\frac{dQ}{d\beta}\right) - \frac{N}{2}, \tag{29}$$

where

$$\begin{aligned}\frac{dR}{d\beta} &= \Sigma_D(X^t|X^{t-\tau})^{-1}\\ &-2\beta\Sigma_D(X^t|X^{t-\tau})^{-1T}\Sigma_D(X^t,X^{t-\tau})\Sigma_D(X^{t-\tau})^{-1}Q^{-1}\Sigma_D(X^{t-\tau})^{-1}\Sigma_D(X^t,X^{t-\tau})^T\Sigma_D(X^t|X^{t-\tau})^{-1}\\ &-\beta^2\Sigma_D(X^t|X^{t-\tau})^{-1T}\Sigma_D(X^t,X^{t-\tau})\Sigma_D(X^{t-\tau})^{-1}\frac{dQ^{-1}}{d\beta}\Sigma_D(X^{t-\tau})^{-1}\Sigma_D(X^t,X^{t-\tau})^T\Sigma_D(X^t|X^{t-\tau})^{-1},\end{aligned} \tag{30}$$

$$\frac{dQ}{d\beta} = \Sigma_D(X^{t-\tau})^{-1}\Sigma_D(X^t,X^{t-\tau})^T\Sigma_D(X^t|X^{t-\tau})^{-1}\Sigma_D(X^t,X^{t-\tau})\Sigma_D(X^{t-\tau})^{-1}, \tag{31}$$

and

$$\frac{dQ^{-1}}{d\beta} = -Q^{-1}\frac{dQ}{d\beta}Q^{-1}, \tag{32}$$

$$= -Q^{-1}\Sigma_D(X^{t-\tau})^{-1}\Sigma_D(X^t,X^{t-\tau})^T\Sigma_D(X^t|X^{t-\tau})^{-1}\Sigma_D(X^t,X^{t-\tau})\Sigma_D(X^{t-\tau})^{-1}Q^{-1}. \tag{33}$$

Inspection of the above equations reveals that $dI^*(\beta)/d\beta = 0$ is a quadratic equation with respect to β. Thus, β can be analytically computed without resorting to numerical optimization such as gradient ascent.

Electrocorticogram (ECoG) recording

The detailed recording protocols were described in [38]. Here, we briefly describe the aspects of the protocols that are relevant for our analysis. We used customized multichannel ECoG electrode arrays. An array of ECoG electrodes was embedded in an insulating silicone sheet. The surface of the sheet was dimpled to expose the surface of ECoG electrodes with the diameter of 1 mm. The electrodes were made of platinum discs, and inter-electrode distance was 5 mm. We implanted 128 ECoG electrodes in the subdural space in four adult macaque monkeys. The ECoG electrodes covered the left hemisphere over the frontal, parietal, temporal, and occipital lobes. ECoG signal was recorded at a sampling rate of 1 kHz. All experimental and surgical procedures were performed in accordance with the protocols approved by the RIKEN ethics committee. During the experiments, the monkeys were seated in a primate chair with both arms and head restrained. We analyzed the data recorded when the monkeys were awake.

Data processing and calculation of integrated information Φ*

To remove line noise and reduce artifacts in the ECoG data, we computed bipolar re-referenced signals between two neighboring electrodes. We calculated integrated information Φ^* using all the bipolar re-referenced signals (64 in total). We considered the simplest partition scheme, "atomic partition" [39], in which the system is partitioned into its individual elements. For this data set, it meant that we computed Φ^* assuming that all the 64 channels are independent. The atomic partition gives the upper bound of Φ^* among all the possible partitions because it quantifies the amount of information loss when all the interactions in the system are ignored for decoding.

We approximated the probability distributions of the continuous ECoG signals with the Gaussian distribution. Under the Gaussian assumption, we analytically computed Φ^* by using the equations derived in Methods. We estimated the covariance matrices of the data with a time window of 2s and a time step of 2s. Then, we averaged the covariance matrices over 600s and used the average of the covariance matrices for computation of Φ^*.

Supporting Information

S1 Text. Mathematical details of integrated information.
(PDF)

S1 Fig. Theoretical requirements are not satisfied by previously proposed measures.
(PDF)

Author Contributions

Conceived and designed the experiments: MO SiA NT. Performed the experiments: MO TY NF. Analyzed the data: MO. Contributed reagents/materials/analysis tools: MO SiA NT. Wrote the paper: MO NT.

References

1. Chalmers DJ. Facing up to the problem of consciousness. J Conscious Stud. 1995; 2: 200–219.
2. Tononi G. An information integration theory of consciousness. BMC Neurosci 2004; 5: 42. doi: 10.1186/1471-2202-5-42 PMID: 15522121
3. Tononi G. Consciousness as integrated information: a provisional manifesto. Biol Bull. 2008; 215: 216–242. doi: 10.2307/25470707 PMID: 19098144
4. Tononi G. Information integration: its relevance to brain function and consciousness. Arch Ital Biol. 2010; 148: 299–322. PMID: 21175016
5. Tononi G. Integrated information theory of consciousness: an updated account. Arch Ital Biol. 2012; 150: 56–90. PMID: 23165867
6. Balduzzi D, Tononi G. Integrated information in discrete dynamical systems: Motivation and theoretical framework. PLoS Comput Biol 2008; 4: e1000091. doi: 10.1371/journal.pcbi.1000091 PMID: 18551165
7. Balduzzi D, Tononi G. Qualia: the geometry of integrated information. PLoS Comput Biol. 2009; 5: e1000462. doi: 10.1371/journal.pcbi.1000462 PMID: 19680424
8. Oizumi M, Albantakis L, Tononi G. From the phenomenology to the mechanisms of consciousness: Integrated Information Theory 3.0. PLoS Comp Biol. 2014; 10: e1003588. doi: 10.1371/journal.pcbi.1003588
9. Tononi G, Koch C. Consciousness: here, there and everywhere? Phil Trans R Soc B. 2015; 19: 370.

10. Massimini M, Ferrarelli F, Huber R, Esser SK, Singh H, Tononi G. Breakdown of cortical effective connectivity during sleep. Science. 2005; 309: 2228–2232. doi: 10.1126/science.1117256 PMID: 16195466

11. Massimini M, Ferrarelli F, Esser SK, Riedner BA, Huber R, Murphy M, Peterson MJ, Tononi G. Triggering sleep slow waves by transcranial magnetic stimulation. Proc Natl Acad Sci USA. 2007; 104: 8496–8501. doi: 10.1073/pnas.0702495104 PMID: 17483481

12. Ferrarelli F, Massimini M, Sarasso S, Casali A, Riedner BA, Angelini G, Tononi G, Pearce RA. Breakdown in cortical effective connectivity during midazolam-induced loss of consciousness. Proc Natl Acad Sci USA. 2010; 107: 2681–2686. doi: 10.1073/pnas.0913008107 PMID: 20133802

13. Rosanova M, Gosseries O, Casarotto S, Boly M, Casali AG, Bruno MA, Mariotti M, Boveroux P, Tononi G, Laureys S, Massimini M. Recovery of cortical effective connectivity and recovery of consciousness in vegetative patients. Brain. 2012; 135. 1308–1320. doi: 10.1093/brain/awr340 PMID: 22226806

14. Casali AG, Gosseries O, Rosanova M, Boly M, Sarasso S, et al. A theoretically based index of consciousness independent of sensory processing and behavior. Sci Transl Med. 2013; 5: 198ra105. PMID: 23946194

15. Barrett AB, Seth AK. Practical measures of integrated information for time-series data. PLoS Comput Biol 2011; 7. e1001052. doi: 10.1371/journal.pcbi.1001052 PMID: 21283779

16. Cover TM, Thomas JA. Elements of information theory. New York: Wiley; 1991.

17. Merhav N, Kaplan G, Lapidoth A, Shamai Shitz S. On information rates for mismatched decoders. IEEE Trans Inform Theory. 1994; 40: 1953–1967. doi: 10.1109/18.340469

18. Latham PE, Nirenberg S. Synergy, redundancy, and independence in population codes, revisited. J Neurosci. 2005; 25: 5195–5206. doi: 10.1523/JNEUROSCI.5319-04.2005 PMID: 15917459

19. Oizumi M, Ishii T, Ishibashi K, Hosoya T, Okada M. Mismatched decoding in the brain. J Neurosci. 2010; 30: 4815–4826. doi: 10.1523/JNEUROSCI.4360-09.2010 PMID: 20357132

20. Oizumi M, Okada M, Amari S. Information loss associated with imperfect observation and mismatched decoding. Front Comput Neurosci. 2011; 5: 9. doi: 10.3389/fncom.2011.00009 PMID: 21629857

21. Rieke F, Warland D, de Ruyter van Steveninck R, Bialek W. Spikes: Exploring the neural code. Cambridge, MA: MIT Press; 1997.

22. Dayan P, Abbott LF. Theoretical Neuroscience. Computational and Mathematical Modeling of Neural Systems. Cambridge, MA: MIT Press; 2001.

23. Averbeck BB, Latham PE, Pouget A. Neural correlations, population coding and computation. Nat Rev Neurosci. 2006; 7: 358–366. doi: 10.1038/nrn1888 PMID: 16760916

24. Quian Quiroga R, Panzeri S. Extracting information from neuronal populations: information theory and decoding approaches. Nat Rev Neurosci. 2009; 10: 173–185. doi: 10.1038/nrn2578 PMID: 19229240

25. Ay N. Information geometry on complexity and stochastic interaction. 2001. MPI MIS Preprint 95. Available: http://www.mis.mpg.de/publications/preprints/2001/prepr2001-95.html.

26. Ay N. Information geometry on complexity and stochastic interaction. Entropy. 2015; 17: 2432–2458. doi: 10.3390/e17042432

27. Lee U, Mashour GA, Kim S, Noh GJ, Choi BM. Propofol induction reduces the capacity for neural information integration: Implications for the mechanism of consciousness and general anesthesia. Conscious Cogn. 2009; 18: 56–64. doi: 10.1016/j.concog.2008.10.005 PMID: 19054696

28. Chang JY, et al. Multivariate autoregressive models with exogenous inputs for intracerebral responses to direct electrical stimulation of the human brain. Front Hum Neurosci. 2012; 6: 317. doi: 10.3389/fnhum.2012.00317 PMID: 23226122

29. Alkire MT, Hudetz AG, Tononi G. Consciousness and anesthesia. Science. 2008; 322: 876–880. doi: 10.1126/science.1149213 PMID: 18988836

30. Boly M. Measuring the fading consciousness in the human brain. Curr Opin Neurol. 2011; 24: 394–400. PMID: 21577107

31. Sanders RD, Tononi G, Laureys S, Sleigh J. Unresponsiveness $\neq$ unconsciousness. Anesthesiology. 2012; 116: 946–959. doi: 10.1097/ALN.0b013e318249d0a7 PMID: 22314293

32. Boly M, Sasai S, Gosseries O, Oizumi M, Casali A, Massimini M, et al. Stimulus set meaningfulness and neurophysiological differentiation: A functional magnetic resonance imaging study. PLoS ONE. 2015; 10: e0125337. doi: 10.1371/journal.pone.0125337 PMID: 25970444

33. Ding M, Chen Y, Bressler SL. Granger causality: Basic theory and application to neuroscience. In: Schelter S, Winterhalder N, Timmer J, editors. Handbook of time series analysis. Wienheim: Wiley; 2006. pp. 438–460.

34. Vicente R, Wibral M, Lindner M, Pipa G. Transfer entropy–a model-free measure of effective connectivity for the neurosciences. J Comput Neurosci. 2011; 30: 45–67. doi: 10.1007/s10827-010-0262-3 PMID: 20706781

35. Amari S, Nagaoka H. Methods of information geometry. AMS and Oxford University Press; 2000.

36. Oizumi M, Tsuchiya N, Amari S. A unified framework for information integration based on information geometry. 2015. Preprint. Available: arXiv:1510.04455.

37. Seth AK, Barrett AB, Barnett L. Causal density and integrated information as measures of conscious level. Philos Transact A Math Phys Eng Sci. 2011; 369: 3748–3767. doi: 10.1098/rsta.2011.0079

38. Yanagawa T, Chao ZC, Hasegawa N, Fujii N. Large-scale information flow in conscious and unconscious states: an ECoG study in monkeys. PLoS One. 2013; 8: e80845. doi: 10.1371/journal.pone.0080845 PMID: 24260491

39. Edlund J, Chaumont N, Hintze A, Koch C, Tononi G, Adami C. Integrated information increases with fitness in the evolution of animats. PLoS Comput Biol. 2011; 7: e1002236. doi: 10.1371/journal.pcbi.1002236 PMID: 22028639