Unified framework for information integration based on information geometry

Masafumi Oizumi[a,b,1], Naotsugu Tsuchiya (土谷　尚嗣)[b,c,d], and Shun-ichi Amari[a]

[a]RIKEN Brain Science Institute, Wako, Saitama 351-0198, Japan; [b]School of Psychological Sciences, Faculty of Biomedical and Psychological Sciences, Monash University, Melbourne, VIC 3800, Australia; [c]Monash Institute of Cognitive and Clinical Neuroscience, Monash University, Melbourne, VIC 3800, Australia; and [d]Advanced Telecommunications Research Institute International, Hikaridai, Keihanna Science City, Kyoto 619-0288, Japan

Edited by William Bialek, Princeton University, Princeton, NJ, and approved October 26, 2016 (received for review March 8, 2016)

Assessment of causal influences is a ubiquitous and important subject across diverse research fields. Drawn from consciousness studies, integrated information is a measure that defines integration as the degree of causal influences among elements. Whereas pairwise causal influences between elements can be quantified with existing methods, quantifying multiple influences among many elements poses two major mathematical difficulties. First, overestimation occurs due to interdependence among influences if each influence is separately quantified in a part-based manner and then simply summed over. Second, it is difficult to isolate causal influences while avoiding noncausal confounding influences. To resolve these difficulties, we propose a theoretical framework based on information geometry for the quantification of multiple causal influences with a holistic approach. We derive a measure of integrated information, which is geometrically interpreted as the divergence between the actual probability distribution of a system and an approximated probability distribution where causal influences among elements are statistically disconnected. This framework provides intuitive geometric interpretations harmonizing various information theoretic measures in a unified manner, including mutual information, transfer entropy, stochastic interaction, and integrated information, each of which is characterized by how causal influences are disconnected. In addition to the mathematical assessment of consciousness, our framework should help to analyze causal relationships in complex systems in a complete and hierarchical manner.

integrated information | mutual information | transfer entropy | information geometry | consciousness

Quantitative assessment of causal influences among elements in a complex system is a fundamental problem in many fields of science, including physics (1), economics (2), gene networks (3), social networks (4), ecosystems (5), and neuroscience (6). There have been many previous attempts to quantify causal influences between elements in stochastic systems. Information theory has played a pivotal role in these endeavors, leading to various measures, including predictive information (7), transfer entropy (8), and stochastic interaction (9). Drawn from consciousness studies involving measurement of integration of neural activity (10, 11), the mathematical concept of integrated information is also useful as a framework for analyzing causal relationships in complex systems with multiple elements.

Recent research suggests that the brain loses the ability to integrate information when consciousness is lost during dreamless sleep (12), general anesthesia (13), or vegetative states (14), suggesting that quantifying integration of information can serve as a neurophysiological marker of consciousness (10, 11, 15). The integrated information theory (IIT) of consciousness (16, 17) proposes a measure of integration called integrated information that quantifies multiple causal influences among elements of a system. Integrated information is theoretically motivated by the holistic property of consciousness experienced as a unified whole that is irreducible into separate parts or experiences. Whereas the original motivation for integrated information is intended to elucidate the neural substrate of consciousness, it can in principle be applied to many research fields.

Despite its broad potential impact, the application of integrated information (16, 18) to experimental data is severely limited (19, 20) due to the original measure's derivation under restricted conditions, wherein the probability distribution of past states in a system is assumed to be uniform, variable discrete (18). In an effort to broaden the applicability, several measures have been proposed under general conditions (9, 19, 21). However, these proposed measures are limited by mathematical problems. Quantification of a pairwise causal influence from one element to another can be achieved with existing measures, but to quantify multiple causal influences among many parts poses the problems of overestimation and confounding noncausal influences. To overcome these problems, we propose a unified framework for quantifying causal influences based on information geometry (22). The measure we propose, called "geometric integrated information" Φ_G, overcomes the described difficulties, provides geometric interpretations of existing measures, and elucidates the relationships among the measures in a hierarchical manner. The mathematical solution we derive should have broad utility in elucidating complex systems.

Three Postulates on Strength of Influences

We propose a unified theoretical framework for quantifying the strength of spatiotemporal influences based on three postulates. Let us consider a stochastic dynamical system in which the past and present states of the system are given by

Significance

Measuring the degree of causal influences among multiple elements of a system is a fundamental problem in physics and biology. We propose a unified framework for quantifying any combination of causal relationships between elements in a hierarchical manner based on information geometry. Our measure of integration, called geometrical integrated information, quantifies the strength of multiple causal influences among elements by projecting the probability distribution of a system onto a constrained manifold. This measure overcomes mathematical problems of existing measures and enables an intuitive understanding of the relationships between integrated information and other measures of causal influence such as transfer entropy. Inspired by the integration of neural activity in consciousness studies, our measure should have general utility in analyzing complex systems.

Author contributions: M.O. and S.A. designed research; M.O. and S.A. performed research; and M.O., N.T., and S.A. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]To whom correspondence should be addressed. Email: oizumi@brain.riken.jp.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1603583113/-/DCSupplemental.

STATISTICS

NEUROSCIENCE

$X = \{x_1, x_2, \cdots, x_N\}$ and $Y = \{y_1, y_2, \cdots, y_N\}$, respectively, where N is the number of elements in the system. Information about X is integrated by influences among elements and transmitted to Y. The spatiotemporal influences of the system are fully characterized by the joint probability distribution $p(X, Y)$. We call $p(X, Y)$ a "full model". In a dynamical system characterized by $p(X, Y)$, there are three different types of influences. Influences between elements at the same time (called equal-time influences) can be quantified by analyzing only the marginal distributions $p(X)$ or $p(Y)$. Influences across different time points (called across-time influences) can be further divided into those among different units (cross-influences) and those within the same unit (self-influences). The across-time influences can be quantified from the conditional probability distribution $p(Y|X)$. They are also known as causal influences (2, 8), in the sense of causality that is statistically inferred from conditional probability distributions although it does not necessarily mean actual physical causality (23). Here, we use the term causality in this context and focus on quantifying causal influences.

For quantifying causal influences (both self- and cross-influences) among elements of X and Y, consider approximating the probability distribution $p(X, Y)$ by another probability distribution $q(X, Y)$ in which the influences of interest are statistically disconnected. We call $q(X, Y)$ a "disconnected model." The strength of influences can be quantified by to what extent the corresponding disconnected model $q(X, Y)$ can approximate the full model $p(X, Y)$. The goodness of the approximation can be evaluated by the difference between the two probability distributions $p(X, Y)$ and $q(X, Y)$. Minimizing a difference between $p(X, Y)$ and $q(X, Y)$ corresponds to finding the best approximation of $p(X, Y)$ by a disconnected model $q(X, Y)$. From this reasoning, we propose the first postulate as follows.

Postulate 1. *Strength of influences is quantified by a minimized difference between the full model and a disconnected model.*

The second postulate is used to define a disconnected model. Consider partitioning the elements of a system into m parts, $X = \{X_1, X_2, \cdots, X_m\}$ and $Y = \{Y_1, Y_2, \cdots, Y_m\}$, where X_i and Y_i contain the same elements in a system. To avoid the confounds of noncausal influences, we should minimally disconnect only the influences of interest without affecting the rest. To define such a minimal operation of statistically disconnecting influences from X_i to Y_j, we propose the second postulate as follows.

Postulate 2. *A disconnected model, where influences from X_i to Y_j are disconnected, satisfies the Markov condition $X_i \rightarrow \tilde{X}_i \rightarrow Y_j$, where $\tilde{X}_i$ is the complement of X_i in X; that is, $\tilde{X}_i = X - X_i$.*

The Markov condition $X_i \rightarrow \tilde{X}_i \rightarrow Y_j$ means that X_i and Y_j are conditionally independent given $\tilde{X}_i$,

$$q(X_i, Y_j|\tilde{X}_i) = q(X_i|\tilde{X}_i)q(Y_j|\tilde{X}_i). \quad [1]$$

Under the Markov condition, there is no direct influence from X_i on Y_j given the states of the other elements $\tilde{X}_i$ being fixed.

The third postulate defines the measure of a difference between the full model and a disconnected model, which is denoted by $D[p : q]$. There are many possible ways to quantify the difference between two probability distributions (22, 24). We consider several theoretical requirements that the measure of difference should satisfy to have desirable mathematical properties (details in *Supporting Information*): (*i*) $D[p : q]$ should be nonnegative and becomes 0 if and only if $p = q$, (*ii*) $D[p : q]$ should be invariant under invertible transformations of random variables, (*iii*) $D[p : q]$ should be decomposable, and (*iv*) $D[p : q]$ should be flat. We can prove that the only measure that satisfies all of the theoretical requirements is the well-known Kullback–Leibler (KL) divergence (22). Thus, we propose the third postulate as follows.

Postulate 3. *A difference between the full model and a disconnected model is measured by KL divergence.*

Taken together, the strength of causal influences from X_i to Y_j, $ci[X_i \rightarrow Y_j]$, is quantified by the minimized KL divergence,

$$ci[X_i \rightarrow Y_j] = \min_{q(X,Y)} D_{KL}[p(X, Y)||q(X, Y)], \quad [2]$$

under the constraint of the Markov condition given by Eq. **1**.

A Unified Derivation of Existing Measures

In this section, we derive existing measures from the unified framework and provide the interpretations of them.

Total Causal Influences: Mutual Information. First, consider quantifying the total strength of causal influences between the past and present states. From the operation of disconnections given by Eq. **1**, the influences from all elements X to Y are disconnected by forcing X and Y to be independent,

$$q(X, Y) = q(X)q(Y). \quad [3]$$

The disconnected model is graphically represented in Fig. 1*A*. To introduce the perspective of information geometry, consider a manifold of probability distributions $\mathcal{M}_F$, where each point in the manifold represents a probability distribution $p(X, Y)$ (a full model). Consider also a manifold $\mathcal{M}_I$ where X and Y are independent, which means that there are no causal influences between X and Y. A probability distribution $q(X, Y)$ (a disconnected model) is represented as a point in the manifold $\mathcal{M}_I$. In general, the actual probability distribution $p(X, Y)$ is represented as a point outside the submanifold $\mathcal{M}_I$ (Fig. 2). The difference between the two probability distributions is quantified by KL divergence,

$$D_{KL}[p(X, Y)||q(X, Y)] = \sum_{X,Y} p(X, Y) \log \frac{p(X, Y)}{q(X, Y)}. \quad [4]$$

We consider finding the closest point q^* to p within the submanifold $\mathcal{M}_I$, which minimizes the KL divergence between $p(X, Y)$ and $q(X, Y) \in \mathcal{M}_I$ (Fig. 2). This corresponds to finding the best approximation of $p(X, Y)$. The minimizer of KL divergence is derived by orthogonally projecting the point $p(X, Y)$ to the manifold $\mathcal{M}_I$ according to the projection

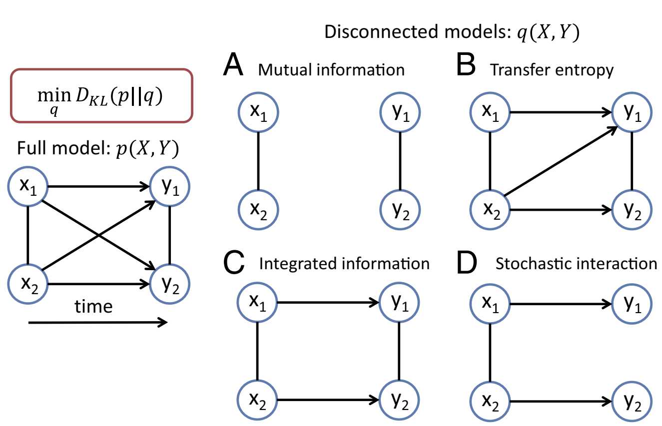

Fig. 1. (*A–D*) Minimizing the Kullback–Leibler (KL) divergence between the full and the disconnected model leads to various information theoretic quantities: (*A*) mutual information, (*B*) transfer entropy, (*C*) integrated information, and (*D*) stochastic interaction. Constraints imposed on the disconnected model are graphically shown.

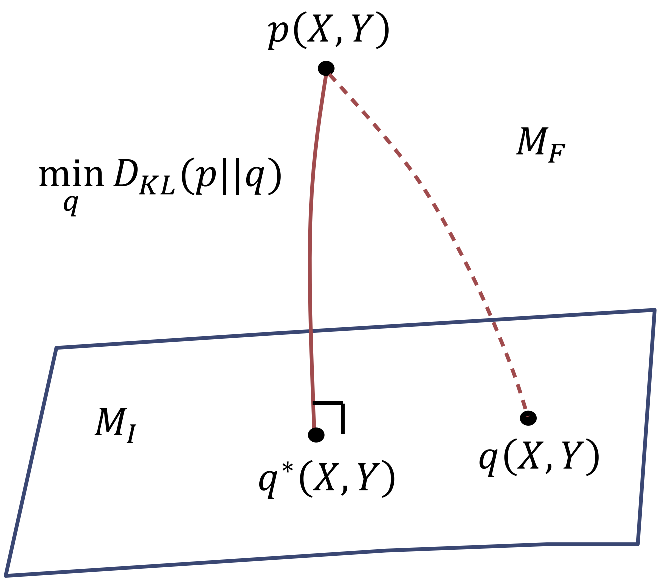

Fig. 2. Information geometric picture for minimizing the KL divergence between the full model $p(X, Y)$, which resides in the manifold $\mathcal{M}_F$, and the disconnected model $q(X, Y)$, which resides in the manifold $\mathcal{M}_I$. $q^*(X, Y)$ is the point in $\mathcal{M}_I$ that is closest to $p(X, Y)$.

theorem in information geometry (22) (*Supporting Information*). In the present case, p, the closest point q^*, and any point q in $\mathcal{M}_I$ form an orthogonal triangle. Thus, the following Pythagorean relation holds: $D(p||q) = D(p||q^*) + D(q^*||q)$. From the Pythagorean relation, we can find that the KL divergence is minimized when the marginal distributions of $q^*(X, Y)$ over X and Y are both equal to those of the actual distribution $p(X, Y)$; i.e., $q^*(X) = p(X)$ and $q^*(Y) = p(Y)$. The minimized KL divergence is given by

$$\min_q D_{KL}[p||q] = H(Y) - H(Y|X), \quad \textbf{[5]}$$

$$= I(X; Y) \quad \textbf{[6]}$$

where $H(Y)$ is the entropy of Y, $H(Y|X)$ is the conditional entropy of Y given X, and $I(X; Y)$ is the mutual information between X and Y. From the derivation, we can interpret the mutual information between X and Y as the total causal influences between X and Y. The mutual information between the present and past states can be also interpreted as the degree of predictability of the present states given the past states and has been termed as predictive information (7).

Partial Causal Influences: Conditional Transfer Entropy. Next, consider quantifying a partial causal influence from one element to another in the system. From the operation of disconnections in Eq. **1**, a partial causal influence from x_i to y_j is disconnected by q, satisfying

$$q(x_i, y_j|\tilde{x}_i) = q(x_i|\tilde{x}_i)q(y_j|\tilde{x}_i), \quad \textbf{[7]}$$

where $\tilde{x}_i$ is the past states of all of the variables other than x_i. Under the constraint, the KL divergence is minimized when $q(X) = p(X)$, $q(y_j|X) = p(y_j|\tilde{x}_i)$, and $q(\tilde{y}_j|X, y_j) = p(\tilde{y}_j|X, y_j)$ (*Supporting Information*). The minimized KL divergence is found to be equal to the conditional transfer entropy,

$$\min_q D_{KL}[p||q] = H(y_j|\tilde{x}_i) - H(y_j|X), \quad \textbf{[8]}$$

$$= TE(x_i \to y_j|\tilde{x}_i), \quad \textbf{[9]}$$

where $TE(x_i \to y_j|\tilde{x}_i)$ is the conditional transfer entropy from x_i to y_j given $\tilde{x}_i$. Thus, we can interpret the conditional transfer entropy as the strength of the partial causal influence from x_i to y_j.

A Measure of Integrated Information

Integrated information is defined as a measure to quantify the strength of all causal influences among parts of the system. In the case of two units, integrated information should quantify both of the causal influences from x_1 to y_2 and from x_2 to y_1. It aims to quantify the extent to which the whole system exerts synergistic influences on its future more than the parts of a system independently do and, thus, irreducibility of the whole system into independent parts (16). Accordingly, integrated information is theoretically required that it should be nonnegative and upper bounded by the total causal influences in the whole system, which is the mutual information between the past and present states $I(X; Y)$ in our framework as shown above (20). Based on *Postulates 1–3*, we uniquely derive a measure of integrated information by imposing the corresponding constraints, which naturally satisfies the theoretical requirement.

Consider again partitioning a system into m parts. By applying the operation in Eq. **1** for all pairs of i and j ($\neq i$), we can find that all causal influences among the parts are disconnected by the condition

$$q(Y_i|X) = q(Y_i|X_i) \ (\forall i). \quad \textbf{[10]}$$

To quantify integrated information, we consider a manifold $\mathcal{M}_G$ constrained by Eq. **10**. Note that within $\mathcal{M}_G$, the present states in a part Y_i directly depend only on the past states of itself, X_i, and thus the transfer entropies from one part X_i to all of the other parts Y_j ($j \neq i$) are 0. Now we propose a measure of integrated information, called geometric integrated information Φ_G, as the minimized KL divergence between the actual distribution $p(X, Y)$ and the disconnected distribution $q(X, Y)$ within $\mathcal{M}_G$:

$$\Phi_G = \min_{q \in \mathcal{M}_G} D_{KL}[p||q]. \quad \textbf{[11]}$$

The manifold $\mathcal{M}_G$ formed by the constraints for integrated information (Eq. **10**) includes the manifold $\mathcal{M}_I$ formed by the constraints for mutual information (Eq. **3**); i.e., $\mathcal{M}_I \subset \mathcal{M}_G$. Because minimizing the KL divergence in a larger space always leads to a smaller value, Φ_G is always smaller than or equal to the mutual information $I(X; Y)$:

$$0 \leq \Phi_G \leq I(X; Y). \quad \textbf{[12]}$$

Thus, Φ_G, uniquely derived from *Postulates 1–3*, naturally satisfies the theoretical requirements as integrated information.

Comparisons with Other Measures

The Sum of Transfer Entropies. For simplicity, consider a system consisting of two variables (Fig. 1). Conceptually, a measure of integrated information should be designed to quantify the strength of two causal influences from x_1 to y_2 and from x_2 to y_1 (Fig. 1*C*). Because each causal influence is quantified by the transfer entropy, $TE(x_1 \to y_2|x_2)$ or $TE(x_2 \to y_1|x_1)$, one may naively think that the sum of transfer entropies can be used as a valid measure of integrated information and may be the same as Φ_G. In contrast with this naive intuition, the sum of transfer entropies is not equal to Φ_G and moreover, it can exceed the mutual information between X and Y, which violates the important theoretical requirement as a measure of integrated information (Eq. **12**). When there is strong dependence between y_1 and y_2, simply taking the sum of transfer entropies leads to overestimation of the total strength of causal influences. An extreme case where such overestimation occurs is when y_1 and y_2 are copies of each other.

As a simple example, consider a system consisting of two binary units, each of which takes one of the two states, 0 or 1. Assume that the probability distribution of the past states of x_1 and x_2 is a uniform distribution; i.e., $p(x_1, x_2) = 1/4$. The present state of unit 1, y_1, is determined by the AND operation of the past state x_1 and x_2, that is, y_1 becomes 1 if both x_1 and x_2 are 1, and it becomes 0 otherwise. On the other hand, y_2 is determined by a "noisy" AND operation where the state of y_2 flips with certain probability r; i.e., $p(y_2 = 1) = 1 - r$ if $(x_1, x_2) = (1, 1)$ and $p(y_2 = 1) = r$ if $(x_1, x_2) = (0, 0), (0, 1), (1, 0)$, where r determines the noise level. As the noise level of the noisy AND operation decreases, the dependence between y_1 and y_2 gets stronger. When there is no noise, i.e., $r = 0$, y_1 and y_2 are completely equal. We varied the strength of dependence by changing the noise level and calculated transfer entropies and Φ_G (see *Supporting Information* for the computation of Φ_G in the binary case) (Fig. 3). As the noise level decreases, the transfer entropy from x_1 to y_2 increases but the mutual information stays the same because y_2, which is a noisy AND gate, does not add any additional information about the input X above the information already provided by y_1, which is the perfect AND gate. When the noise level is low and thus the dependence between y_1 and y_2 is strong, the sum of transfer entropies exceeds the amount of mutual information.

On the other hand, Φ_G never exceeds the amount of mutual information (Fig. 3). Φ_G avoids the overestimation by simultaneously evaluating the strength of multiple influences. In contrast, the sum of transfer entropies separately quantifies causal influences by considering only parts of the system. For example, when the transfer entropy from x_1 to y_2 is quantified, y_1 is not taken into consideration, which leads to the overestimation. To accurately evaluate the total strength of multiple influences, we need to take a holistic approach as we proposed to do with Φ_G. The flaw of the simple sum of transfer entropies illuminates the limitation of the part-based approach and the advantage of the holistic approach.

A related quantity with the sum of transfer entropies has been proposed as causal density (21). Originally, causal density was proposed as the normalized sum of the conditional Granger causality from one element to another (21). Because transfer entropy is equivalent to Granger causality for Gaussian variables

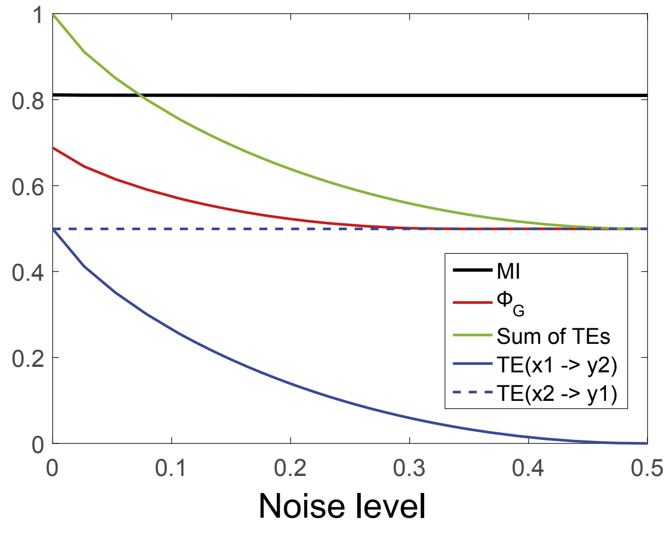

Fig. 3. Comparison between integrated information and the sum of transfer entropies (TE). A system consists of two binary units whose states are determined by an AND gate and a noisy AND gate. When the noise level of the noisy AND gate is low and thus the dependence between the units is strong, the sum of transfer entropies (green line) exceeds the mutual information (black line) whereas integrated information Φ_G (red line) is always less than the mutual information. Each transfer entropy (blue solid and dotted lines) is always less than or equal to Φ_G.

(25), the normalized sum of the conditional transfer entropies can be considered as a generalization of causal density. Although a simple sum of Granger causality or transfer entropies is easy to evaluate and would be useful for approximately evaluating the total strength of causal influences, we need to be careful about the problem of overestimation.

Stochastic Interaction. Another measure, called stochastic interaction (9), was proposed as a different measure of integrated information (19). In the derivation of stochastic interaction, Ay (9) considered a manifold $\mathcal{M}_S$ where the conditional probability distribution of Y given X is decomposed into the product of the conditional probability distributions of each part (Fig. 1*D*):

$$q(Y|X) = \prod_{i=1}^{m} q(Y_i|X_i). \qquad [13]$$

This constraint satisfies the constraint for the integrated information (Eq. **10**). Thus, $\mathcal{M}_S \subset \mathcal{M}_G$. In addition to that, this constraint further satisfies conditional independence among the present states of parts given the past states in the whole system X:

$$q(Y|X) = \prod_{i=1}^{m} q(Y_i|X). \qquad [14]$$

This constraint corresponds to disconnecting equal-time influences among the present states of the parts given the past states of the whole in addition to across-time influences (Fig. 1*D*). On the other hand, the constraint in Eq. **10** corresponds to disconnecting only across-time influences (Fig. 1*C*).

The KL divergence is minimized when $q(X) = p(X)$ and $q(Y_i|X_i) = p(Y_i|X_i)$ (9). The minimized KL divergence is equal to stochastic interaction $SI(X; Y)$:

$$\min_q D_{KL}[p||q] = \sum_i H(Y_i|X_i) - H(Y|X), \qquad [15]$$

$$= SI(X; Y). \qquad [16]$$

In contrast to the manifold $\mathcal{M}_G$ considered for Φ_G, the manifold $\mathcal{M}_S$ formed by the constraints for stochastic interaction (Eq. **13**) does not include the manifold $\mathcal{M}_I$ formed by the constraints for the mutual information between X and Y (Eq. **3**). This is because not only causal influences but also equal-time influences are disconnected in $\mathcal{M}_S$ (Fig. 1*D*). Stochastic interaction can therefore exceed the total strength of causal influences in the whole system, which violates the theoretical requirement as a measure of integrated information (Eq. **12**). Notably, stochastic interaction can be nonzero even when there are no causal influences, i.e., when the mutual information is 0 (20). To summarize, stochastic interaction does not purely quantify causal influences but rather quantifies the mixture of causal influences and simultaneous influences.

Analytical Calculation for Gaussian Variables

Although we cannot derive a simple analytical expression for Φ_G in general, it is possible to derive it for Gaussian variables. In this section, we analytically compute Φ_G when the probability distribution of a system $p(X, Y)$ is Gaussian. We also show a close relationship between the proposed measure of integrated information Φ_G and multivariate Granger causality. Consider the following multivariate autoregressive model,

$$Y = AX + E, \qquad [17]$$

where X and Y are the past and present states of a system, A is the connectivity matrix, and E is Gaussian random variables with mean 0 and covariance matrix $\Sigma(E)$, which are uncorrelated over time. The multivariate autoregressive model is the generative model of a multivariate Gaussian distribution. Regarding

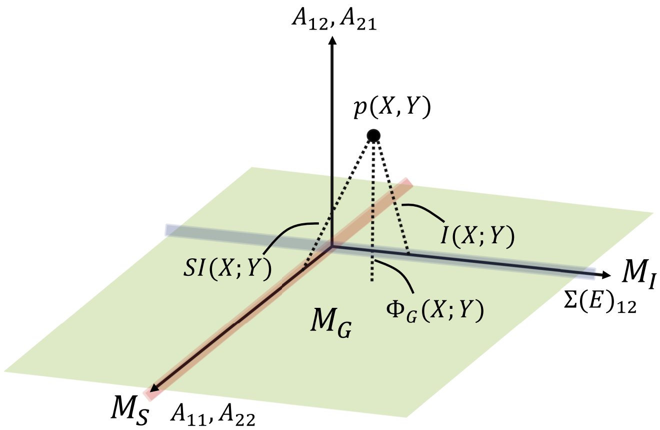

Fig. 4. Relationships between manifolds for mutual information $\mathcal{M}_I$ (gray line), stochastic interaction $\mathcal{M}_S$ (orange line), and integrated information $\mathcal{M}_G$ (green plane) in the Gaussian case. $\mathcal{M}_I$ is the line where $A = 0$, $\mathcal{M}_S$ is the line where $\Sigma(E)_{12}$ and A_{12}, A_{21} are 0, and $\mathcal{M}_G$ is the plane where A_{12}, A_{21} are 0.

Eq. **17** as a full model, we consider the following as a disconnected model:

$$Y = A'X + E'. \quad \textbf{[18]}$$

The constraints for Φ_G (Eq. **10**) correspond to setting the off-diagonal elements of A' to 0:

$$A'_{ij} = 0 \; (i \neq j). \quad \textbf{[19]}$$

It is instructive to compare this with the constraints for the other information theoretic quantities introduced above: the constraints for mutual information (Fig. 1*A*), transfer entropy from x_1 to y_2 (Fig. 1*B*), and stochastic interaction (Fig. 1*D*). They correspond to $A' = 0$, $A'_{21} = 0$, and the off-diagonal elements of A' and $\Sigma(E)'$ being 0, respectively. Fig. 4 shows the relationship between the manifolds formed by the constraints for mutual information $\mathcal{M}_I$, stochastic interaction $\mathcal{M}_S$, and integrated information $\mathcal{M}_G$. We can see that $\mathcal{M}_I$ and $\mathcal{M}_S$ are included in $\mathcal{M}_G$. Thus, Φ_G is smaller than $I(X;Y)$ or $SI(X;Y)$. On the other hand, there is no inclusion relation between $\mathcal{M}_I$ and $\mathcal{M}_S$.

By differentiating the KL divergence between the full model $p(X,Y)$ and a disconnected model $q(X,Y)$ with respect to $\Sigma(X)'^{-1}$, A', and $\Sigma(E)'^{-1}$, we can find the minimum of the KL divergence, using the following equations (details in *Supporting Information*):

$$\Sigma(X)' = \Sigma(X), \quad \textbf{[20]}$$

$$(\Sigma(X)(A - A')\Sigma(E)'^{-1})_{ii} = 0, \quad \textbf{[21]}$$

$$\Sigma(E)' = \Sigma(E) + (A - A')\Sigma(X)(A - A')^T. \quad \textbf{[22]}$$

By substituting Eqs. **20–22** into the KL divergence, we obtain

$$\Phi_G = \frac{1}{2}\log\frac{|\Sigma(E)'|}{|\Sigma(E)|}. \quad \textbf{[23]}$$

$|\Sigma(E)|$ is called the generalized variance, which is used as a measure of goodness of fit, i.e., the degree of prediction error, in multivariate Granger causality analysis (26, 27). In the Gaussian case, Φ_G can be interpreted as the difference in the prediction error on comparison of the full and the disconnected model, in which the off-diagonal elements of A' are set to 0. Thus, Φ_G is consistent with multivariate Granger causality based on the generalized variance. Φ_G can be rewritten as the difference between the conditional entropy in the full model and that in the disconnected model,

$$\Phi_G = H(q(Y|X)) - H(p(Y|X)). \quad \textbf{[24]}$$

For comparison, mutual information, transfer entropy, and stochastic interaction are given as $I(X;Y) = \frac{1}{2}\log\frac{|\Sigma(X)|}{|\Sigma(E)|}$, $TE(x_i \to y_j|x_j) = \frac{1}{2}\log\frac{\Sigma(E)^*_{jj}}{\Sigma(E)_{jj}}$, $SI(X;Y) = \frac{1}{2}\log\frac{\Sigma(E)^*_{11}\Sigma(E)^*_{22}}{|\Sigma(E)|}$, where $\Sigma(E)^*_{jj}$ $(j = 1,2)$ is the covariance of the conditional probability distribution $p(y_j|x_j)$.

Hierarchical Structure

We can construct a hierarchical structure of the disconnected models and then use it to systematically quantify all possible combinations of causal influences (28). For example, in a system consisting of two elements, there are four across-time influences, $x_1 \to y_1$, $x_1 \to y_2$, $x_2 \to y_1$, and $x_2 \to y_2$, which are denoted by T_{11}, T_{12}, T_{21}, and T_{22}, respectively. Although we consider only the cross-influences, T_{12} and T_{21} for transfer entropy and integrated information, we can also quantify self-influences T_{11} and T_{22} by imposing the corresponding constraints, such as $q(y_1|x_1,x_2) = q(y_1|x_2)$ and $q(y_2|x_1,x_2) = q(y_2|x_1)$, respectively. A set of all possible disconnected models forms a partially ordered set with respect to KL divergence between the full and the disconnected models (Fig. 5). If a given disconnected model is related to another one with a removal or an inclusion of an influence, the two models are connected by a line in Fig. 5. From *Bottom* to *Top* in Fig. 5, information loss increases as more influences are disconnected. Note that there is no ordering relationship between the disconnected models at the same level of the hierarchy. In Fig. 5, *Top*, all four influences are disconnected, and thus information loss is maximized, which corresponds to the mutual information $I(X;Y)$. The hierarchical structure generalizes related measures mentioned in this article and provides a clear perspective on the relationship among different measures.

Discussion

In this paper, we proposed a unified framework based on information geometry, which enables us to quantify multiple influences without overestimation and confounds of noncausal influences. With the framework, we uniquely derived the measure of integrated information, Φ_G. Moreover, our framework enables the complete description of causal relationships within a system by quantifying any combination of causal influences in a

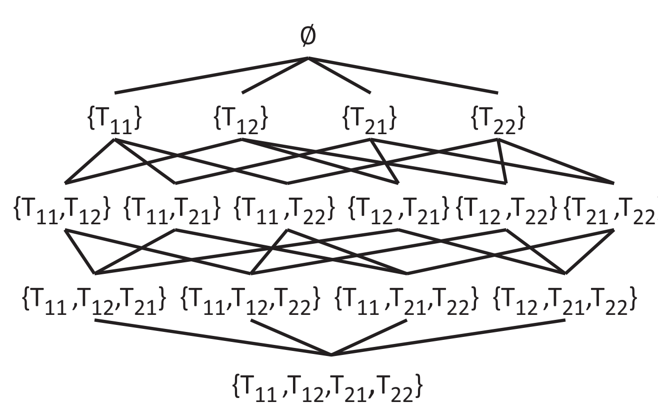

Fig. 5. A hierarchical structure of the disconnected models where across-time influences are broken in a system consisting of two units. All possible combinations of influences retained in the disconnected model are displayed. If two models are related with the addition or removal of one influence, they are connected by a line. The KL divergence between the full and the disconnected model increases from *Bottom* to *Top*.

hierarchical manner as shown in Fig. 5. We expect that our framework can be used in diverse research fields, including neuroscience (29, 30), where network connectivity analysis has been an active research topic (31), and in particular consciousness researchers (32–34) because information integration is considered to be a key prerequisite of conscious information processing in the brain (10, 11).

To apply the measure of integrated information in real data, we need to resolve several practical difficulties. First, the computational costs increase exponentially with the system size. Thus, some way of approximating data is necessary. As we showed in this paper, the Gaussian approximation enables us to analytically compute integrated information, allowing us to compute integrated information in a large system (Eqs. **20–23**). However, in real world systems, including brains, nonlinearity can be often significant and the Gaussian approximation may poorly fit to data. In such cases, transforming time series data into a sequence of discrete symbols can result in more accurate approximation (34, 35). Our measure of integrated information can be computed in such discrete distributions as shown in *Supporting Information*. Second, we need to find an appropriate partition of a system, which is an important problem in IIT (16). The computational costs for finding the optimal partition also exponentially increase. To overcome this difficulty, some effective optimization method needs to be used, possibly methods from discrete mathematics.

From a theoretical perspective, we could consider replacing *Postulates 2* and *3* with different ones as interesting future research. As for *Postulate 2*, which defines the operation of disconnecting causal influences, we can use the interventional formalism (23, 36), which quantifies causal influences based on mechanisms of a system rather than observation of the system. As for *Postulate 3*, which defines the difference between the full model and a disconnected model, we can replace the KL divergence with other measures (24), such as the optimal transport distance, a.k.a, earth mover's distance, which is considered to be important in IIT (17) and also has been shown to be useful in statistical machine learning (37). Our framework based on information geometry can be generally used for deriving different measures of causal influences from such different postulates and for analyzing the different geometric structures induced by them.

ACKNOWLEDGMENTS. We thank Charles Yokoyama, Matthew Davidson, and Dror Cohen for helpful comments on the manuscript. M.O. was supported by a Grant-in-Aid for Young Scientists (B) from the Ministry of Education, Culture, Sports, Science, and Technology of Japan (26870860). N.T. was supported by the Future Fellowship (FT120100619) and the Discovery Project (DP130100194) from the Australian Research Council. M.O. and N.T. were supported by CREST, Japan Science and Technology Agency.

1. Ito S, Sagawa T (2013) Information thermodynamics on causal networks. *Phys Rev Lett* 111(18):180603.
2. Granger CW (1988) Some recent development in a concept of causality. *J Econom* 39(1):199–211.
3. Bansal M, Belcastro V, Ambesi-Impiombato A, Di Bernardo D (2007) How to infer gene networks from expression profiles. *Mol Syst Biol* 3(1):78.
4. Xiang R, Neville J, Rogati M (2010) Modeling relationship strength in online social networks. *Proceedings of the 19th International Conference on World Wide Web* (Association for Computing Machinery, New York), pp 981–990.
5. Sugihara G, et al. (2012) Detecting causality in complex ecosystems. *Science* 338(6106):496–500.
6. Park HJ, Friston K (2013) Structural and functional brain networks: From connections to cognition. *Science* 342(6158):1238411.
7. Bialek W, Nemenman I, Tishby N (2001) Predictability, complexity, and learning. *Neural Comput* 13(11):2409–2463.
8. Schreiber T (2000) Measuring information transfer. *Phys Rev Lett* 85(2):461.
9. Ay N (2015) Information geometry on complexity and stochastic interaction. *Entropy* 17(4):2432–2458.
10. Koch C, Massimini M, Boly M, Tononi G (2016) Neural correlates of consciousness: Progress and problems. *Nat Rev Neurosci* 17(5):307–321.
11. Tononi G, Boly M, Massimini M, Koch C (2016) Integrated information theory: From consciousness to its physical substrate. *Nat Rev Neurosci* 17(7):450–461.
12. Massimini M, et al. (2005) Breakdown of cortical effective connectivity during sleep. *Science* 309(5744):2228–2232.
13. Alkire MT, Hudetz AG, Tononi G (2008) Consciousness and anesthesia. *Science* 322(5903):876–880.
14. Gosseries O, Di H, Laureys S, Boly M (2014) Measuring consciousness in severely damaged brains. *Annu Rev Neurosci* 37:457–478.
15. Casali AG, et al. (2013) A theoretically based index of consciousness independent of sensory processing and behavior. *Sci Transl Med* 5(198):198ra105.
16. Tononi G (2008) Consciousness as integrated information: A provisional manifesto. *Biol Bull* 215(3):216–242.
17. Oizumi M, Albantakis L, Tononi G (2014) From the phenomenology to the mechanisms of consciousness: Integrated information theory 3.0. *PLoS Comput Biol* 10(5):e1003588.
18. Balduzzi D, Tononi G (2008) Integrated information in discrete dynamical systems: Motivation and theoretical framework. *PLoS Comput Biol* 4(6):e1000091.
19. Barrett AB, Seth AK (2011) Practical measures of integrated information for time-series data. *PLoS Comput Biol* 7(1):e1001052.
20. Oizumi M, Amari S, Yanagawa T, Fujii N, Tsuchiya N (2016) Measuring integrated information from the decoding perspective. *PLoS Comput Biol* 12(1):e1004654.
21. Seth AK, Barrett AB, Barnett L (2011) Causal density and integrated information as measures of conscious level. *Philos Trans A Math Phys Eng Sci* 369(1952):3748–67.
22. Amari S (2016) *Information Geometry and Its Applications* (Springer, Tokyo).
23. Pearl J (2009) *Causality* (Cambridge Univ Press, Cambridge, UK).
24. Tegmark M (2016) Improved measures of integrated information. arXiv:1601.02626.
25. Barnett L, Barrett AB, Seth AK (2009) Granger causality and transfer entropy are equivalent for Gaussian variables. *Phys Rev Lett* 103(23):2–5.
26. Geweke J (1982) Measurement of linear dependence and feedback between multiple time series. *J Am Stat Assoc* 77(378):304–313.
27. Barrett AB, Barnett L, Seth AK (2010) Multivariate granger causality and generalized variance. *Phys Rev E* 81(4):041907.
28. Ay N, Olbrich E, Bertschinger N, Jost J (2011) A geometric approach to complexity. *Chaos* 21(3):037103.
29. Deco G, Tononi G, Boly M, Kringelbach ML (2015) Rethinking segregation and integration: Contributions of whole-brain modelling. *Nat Rev Neurosci* 16(7):430–439.
30. Boly M, et al. (2015) Stimulus set meaningfulness and neurophysiological differentiation: A functional magnetic resonance imaging study. *PLoS One* 10(5):e0125337.
31. Bullmore E, Sporns O (2009) Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nat Rev Neurosci* 10(3):186–198.
32. Lee U, Mashour GA, Kim S, Noh GJ, Choi BM (2009) Propofol induction reduces the capacity for neural information integration: Implications for the mechanism of consciousness and general anesthesia. *Conscious Cogn* 18(1):56–64.
33. Chang JY, et al. (2012) Multivariate autoregressive models with exogenous inputs for intracerebral responses to direct electrical stimulation of the human brain. *Front Hum Neurosci* 6:317.
34. King JR, et al. (2013) Information sharing in the brain indexes consciousness in noncommunicative patients. *Curr Biol* 23(19):1914–1919.
35. Bandt C, Pompe B (2002) Permutation entropy: A natural complexity measure for time series. *Phys Rev Lett* 88(17):174102.
36. Ay N, Polani D (2008) Information flows in causal networks. *Adv Complex Sys* 11(01):17–41.
37. Cuturi M (2013) Sinkhorn distances: Lightspeed computation of optimal transport. *Adv Neural Inform Process Syst* 26:2292–2300.