Article

Evaluating Approximations and Heuristic Measures of Integrated Information

André Sevenius Nilsen [1*], Bjørn Erik Juel[1], William Marshall[2,3], and Johan Frederik Storm[1]

[1] Brain Signalling Group, Department of Physiology, Institute of Basic Medicine, University of Oslo, Norway; *sevenius.nilsen@gmail.com, b.e.juel@medisin.uio.no, j.f.storm@medisin.uio.no.
2 Department of Psychiatry, University of Wisconsin, Madison WI, USA; wmarshall3@wisc.edu
3 Department of Mathematics and Statistics, Brock University, St. Catharines, ON, Canada.
* Correspondence: sevenius.nilsen@gmail.com; Tel.: +47-908-07-044.

Abstract

Integrated information theory (IIT) proposes a measure of integrated information (Φ) to capture the level of consciousness for a physical system in a given state. Unfortunately, calculating Φ itself is currently only possible for very small model systems, and far from computable for the kinds of systems typically associated with consciousness (brains). Here, we consider several proposed measures and computational approximations, some of which can be applied to larger systems, and test if they correlate well with Φ. While these measures and approximations capture intuitions underlying IIT and some have had success in practical applications, it has not been shown that they actually quantify the type of integrated information specified by the latest version of IIT.

In this study, we evaluated these approximations and heuristic measures, based not on practical or clinical considerations, but rather based on how well they estimate the Φ values of model systems. To do this, we simulated networks consisting of 3-6 binary linear threshold nodes randomly connected with excitatory and inhibitory connections. For each system, we then constructed the system's state transition probability matrix (TPM), as well as its state transition matrix (STM) over time for all possible initial states. From these matrices, we calculated Φ, approximations to Φ, and measures based on state differentiation, state entropy, state uniqueness, and integrated information. All measures were correlated with Φ in a state dependent and state independent manner.

Our findings suggest that Φ can be approximated closely in small binary systems by using one or more of the readily available approximations ($r > 0.95$), but without major reductions in computational demands. Furthermore, Φ correlated strongly with measures of signal complexity (LZ, $r_s = 0.722$), decoder based integrated information (Φ*, $r_s = 0.816$), and state differentiation (D1, $r_s = 0.827$), on the system level (state independent). These measures could allow for efficient estimation of Φ on a group level, or as accurate predictors of low, but not high, Φ systems. While it's uncertain whether the results extend to larger systems or systems with other dynamics, we stress the importance that measures aimed at being practical alternatives to Φ are at a minimum rigorously tested in an environment where the ground truth can be established.

Keywords:

integrated information theory; differentiation; integration; complexity; consciousness; computational; IIT; Phi

1. Introduction

The nature of consciousness, defined as subjective experience, has been a philosophical topic for centuries, but has only recently become incorporated into mainstream neuroscience [1]. However, as consciousness is a subjective phenomenon, and thus not directly measurable, it must be operationalized to allow for empirical investigation of its nature and underlying mechanisms [2]. In other words, the scientific study of consciousness requires an objective measure. One such measure has been developed within the framework of the Integrated information theory (IIT), introduced and elaborated by Giulio Tononi and colleagues [3–5]. The theory has attracted much interest because of its axiomatic, quantitative approach towards illuminating fundamental aspects of consciousness. The theory proposes that consciousness is identical to a particular type of integrated information and can be quantified by Phi (Φ). Φ is defined within the theory as a measure of a system's informational irreducibility, or how much information a system in a specific state specifies about its own past and future above and beyond its parts.

A major practical limitation of IIT is that the computational cost of calculating Φ, which according to the current formulation (version 3.0 [5]; here referred to as $\Phi_{3.0}$, implemented through PyPhi [6]) grows as $O(n53^n)$ [6], for binary systems where n is the number of elements in the system. In addition, computing $\Phi_{3.0}$ requires full knowledge of a system's transition probabilities (the probability of the system transitioning from any state to any other state). Taken together, these knowledge and computational requirements place strong constraints on both the system size and level of precision possible for which $\Phi_{3.0}$ can be calculated. Therefore, the exact value of $\Phi_{3.0}$ is intractable for most biological or artificial systems of interest. Currently, the largest systems being investigated are in the order of 20-30 binary elements [7,8], with a practical limit of ~10-12 elements unless special assumptions are made about the system under investigation (e.g. see [9]).

As $\Phi_{3.0}$ quickly becomes computationally intractable as a function of network size, one approach is to implement computational shortcuts, or approximations, that reduce the computational cost. Another approach is to use heuristic measures that capture aspects of IIT such as information differentiation and integration via more tractable methods [10–15]. While many heuristics have been applied to electrophysiological data (e.g. [10,13,14,16–18]), simulated time series of continuous variables (e.g. [11,19]), and discrete variables (e.g. [15,20]), none of the proposed approximations and heuristics have been rigorously validated with respect to $\Phi_{3.0}$ (except [15] which tested several measures in evolved logic gate based animats).

Given this lack in direct comparisons, we claim that if an approximation or heuristic is to be a tractable alternative to the full $\Phi_{3.0}$, then it's important to validate them in systems in which $\Phi_{3.0}$ can be exactly calculated. Therefore, this paper aims to evaluate the accuracy of (1) approximations that speed up parts of the $\Phi_{3.0}$ calculations, and, (2) heuristic measures of integrated information in deterministic isolated discrete networks of binary logic gates of similar type as employed in IIT [5].

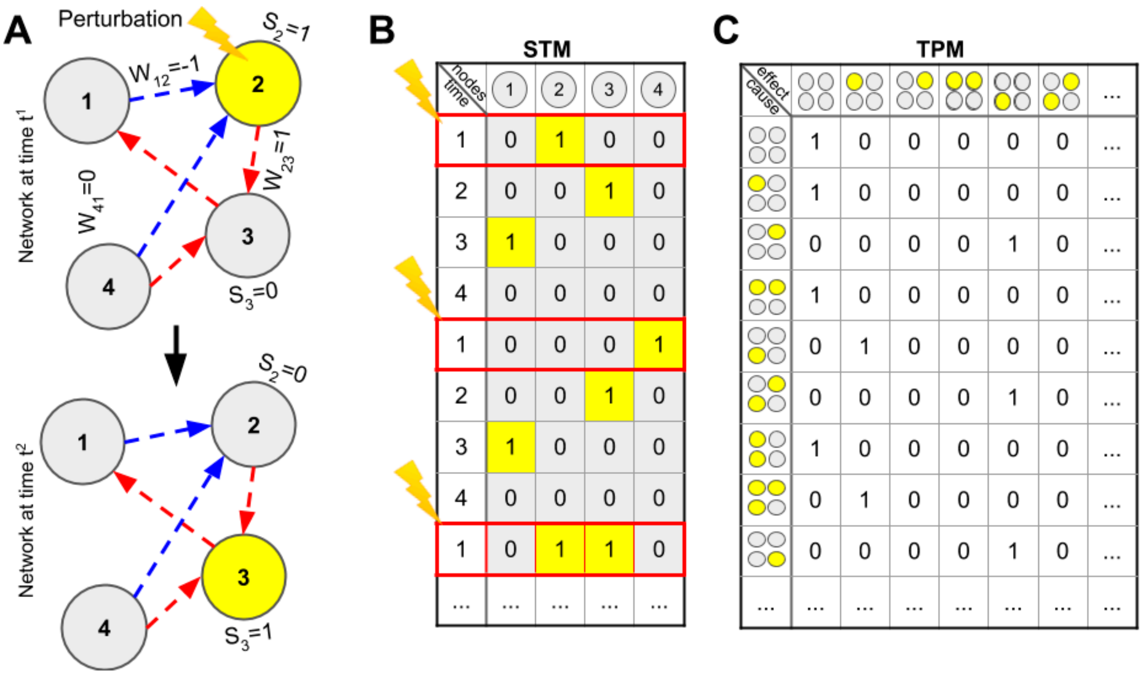
2. Materials and Methods

2.1 Networks

We randomly generated networks consisting of n ∈ {3, ..., 6} binary linear threshold nodes (state S ∈ {0,1}), with fixed threshold ($\theta = 1$) and weighted connections between nodes (W_{ij}

$\in$ {1,0,-1}, for i,j = 1, ..., n). There were no self-connections, $W_{ii} = 0$. Connections were generated as follows: First, for all i ≠ j, we set $W_{ij} = 1$ with a probability $p(W_{ij} = 1) \in \{0.2, 0.3, \ldots, 1.0\}$, a parameter that was fixed for each network. Second, we changed the sign of non-zero connections to $W_{ij} = -1$ with probability $p(W_{ij}=-1) \in \{0.0, 0.1, \ldots, 0.8\}$, this parameter was also fixed for each network. Remaining weights were kept at $W_{ij} = 0$, i.e. no connection. To avoid duplicate network architectures, all networks were checked for uniqueness up to an isomorphism of nodes, i.e. two networks were considered equal if they could be mapped to each other by a relabeling of nodes (using a brute force algorithm). The networks were isolated (no external inputs or modulators). In sum, we generated networks with nodes that could take one of two states ($S_t = 0, 1$) and would be activated ($S_{t+1} = 1$) if the weighted sum of the inputs to the node were equal or larger than its threshold ($\theta = 1$). If a node was activated, it would then output to other nodes according to its outgoing connection weights. Importantly, this allowed for networks with both excitatory ($W_{ij} = 1$), inhibitory ($W_{ij} = -1$), and no ($W_{ij} = 0$) connection between any given pair of nodes. See Figure 1a.

To investigate various measures and approximations we needed functional information about the networks in the form of time series data, i.e. a state transition matrix (STM), and probabilistic descriptions of the transitions from one given state to any other, i.e. a transition probability matrix (TPM). For STMs, each network was periodically set/perturbed to each possible state (full entropy distribution), and simulated for $\alpha(n)2^6$ timesteps, where $\alpha(n)$ is an adjustment factor which depends on the number of nodes in the network (see Appendix A1). The adjustment factor is required to ensure comparability and avoid normalization issues between systems of different number of nodes. The simulated time series (sequence of states) resulted in a $\alpha(n)n2^{6+n}$ STM where one row in the STM reflects the current state of each network node (on/off). In other words, the STM is the observed time series data of the system simulated from all possible initial conditions/states. For TPMs, we require the probability of any given system state leading to any other given system state under the maximum entropy assumption, resulting in a 2^n-by-2^n matrix. In other words, the STM is the observed time series data of the system from all possible initial conditions/states, while the TPM is a general probabilistic description of the STM. As the networks were fully deterministic with no noise, both the STM and the TPM contain the same information and can be generated from each other. See Figure 1b,c.



STM

nodes \ time	①	②	③	④
1	0	1	0	0
2	0	0	1	0
3	1	0	0	0
4	0	0	0	0
1	0	0	0	1
2	0	0	1	0
3	1	0	0	0
4	0	0	0	0
1	0	1	1	0
...	...	...	...	...

TPM

effect \ cause	○○ ○○	●○ ○○	○● ○○	●● ○○	○○ ●○	○● ●○	...
○○ ○○	1	0	0	0	0	0	...
●○ ○○	1	0	0	0	0	0	...
○● ○○	0	0	0	0	1	0	...
●● ○○	1	0	0	0	0	0	...
○○ ●○	0	1	0	0	0	0	...
○● ●○	0	0	0	0	1	0	...
●○ ●○	1	0	0	0	0	0	...
●● ●○	0	1	0	0	0	0	...
○○ ○●	0	0	0	0	1	0	...
...	...	...	...	...	...	...	...

Figure 1. (**A**) Networks were randomly generated with *n* binary linear threshold nodes ($S_i \in \{0, 1\}$, $\Theta \geq 1.0$) and connections ($W_{ij} \in \{-1, 0, 1\}$). Each network was perturbed into each possible initial state and the following state transitions recorded. (**B**) The sequences of state transitions were collected in a state transition matrix (STM) with S_i of each node over time. (**C**) The STM was then used to create a state to state transition probability matrix (TPM) with the probability of a past (cause) state leading to a future (effect) state.

2.2 Integrated Information

For the networks defined above, we calculated $\Phi_{3.0}$ as implemented through PyPhi v1.0 [6]. Here, we just give a brief summary of how $\Phi_{3.0}$ is defined and calculated, but see [5] for a more detailed account. Generally, IIT proposes that a physical system's degree of consciousness is identical to its level of state dependent causal irreducibility (Φ^{max}), i.e. how much information a system in a specific state generates above and beyond that of the system's parts. For a given candidate system (subset of a network) in a state we identify all possible *mechanisms* (subsets of system nodes in a state that irreducibly constrains the past and future state of the system). For each *mechanism*, we find its *cause-effect repertoire* (CER; a probability distribution specifying how the *mechanism* causally constrains the past and future states of the candidate system). The elements that are constrained by a *mechanism* in a given CER comprise the *mechanism's purview*. To find the irreducibility of the CER, connections between the *purview* and the *mechanisms* are cut in all possible ways, and the cut that reduces the *mechanisms* constraint the least is called the *minimum information partition* (MIP). Irreducibility, or integrated information, φ, is quantified by the earth's mover distance (EMD) between the CER of the uncut *mechanism* and CER of the mechanism partitioned by the MIP. A *mechanism* together with its maximally irreducible CER and the associated φ value specifies a *concept*, which expresses the causal role played by the *mechanism* within the system. The set of all concepts is called the cause-effect structure of the candidate system.

Once all irreducible mechanisms of a candidate system have been found, a similar set of operations is done on the system level to understand whether the set of mechanisms specified by the system are reducible to the mechanisms specified by its parts. The irreducibility of the candidate system is quantified by its conceptual integrated information, Φ. This process is repeated for all candidate systems, and the candidate system that is maximally irreducible among all candidate systems is termed a *major complex* (MC). According to IIT then, the MC is the substrate that specifies a particular conscious experience for the (physical) system in a state, and $\Phi_{3.0}$ quantifies the irreducibility of the cause-effect structure it specifies in that state. As such, $\Phi_{3.0}$ is calculated for every reachable state of the system, and the maximum $\Phi_{3.0}$ over all reachable states of the system ($\Phi_{3.0}^{peak}$) is taken as a state independent measure for $\Phi_{3.0}$ of the system.

2.3 Approximations and heuristics

To speed up the calculation of $\Phi_{3.0}$, one can implement several shortcuts or approximations based on assumptions about the system under consideration. Here, we aimed to test six specific approximations; three approximations that are already implemented in the toolbox for calculating $\Phi_{3.0}$ (PyPhi; [6]) that reduces complexity of evaluating information lost during partitioning of a network; two shortcuts based on making an educated guess about the elements included in the MC rather than explicitly testing every candidate subsystem; and one estimation of a system's $\Phi_{3.0}^{peak}$ from the Φ of a few states rather than taking the maximum over all possible states. All approximations are likely to compare well against $\Phi_{3.0}$, but are unlikely to yield significant savings in computational demand.

Another approach is to use heuristics that capture aspects of $\Phi_{3.0}$. These heuristics can be separated into two classes; those that require the full TPM and discrete dynamics (heuristics on discrete networks requiring perturbational data), and those that require time series data (heuristics from observed data). While these measures may reduce computational demands, the heuristics based on discrete dynamics still require full structural and functional knowledge of the system reducing their applicability. On the other hand, measures based on observed data significantly broadens potential applicability at the cost of estimating the underlying causal structure by using observed time series.

All approximations and heuristics that were tested are listed in Table 1, together with an identifier (from "A" to "O") that will be used in the text tor ease of reading, as well as a reference and brief description.

Table 1. Overview of measures

#	S.D. Measure	S.I. Measure	Description	Ref.
	$\Phi_{3.0}$	$\Phi_{3.0}^{peak}$	Integrated information according to IIT 3.0	[5]
A	CO $\Phi_{3.0}$	CO $\Phi_{3.0}^{peak}$	Cut one connection when making partitions	[6]
B	NN $\Phi_{3.0}$	NN$\Phi_{3.0}^{peak}$	No new concepts after partitioning	[6]
C	AP $\Phi_{3.0}$	AP$\Phi_{3.0}^{peak}$	All-partitioning (vs. default bi-partitioning)	[5]
D	WS $\Phi_{3.0}$	WS$\Phi_{3.0}^{peak}$	Whole system as MC	
E	IC $\Phi_{3.0}$	IC$\Phi_{3.0}^{peak}$	Elements with recurrent connections as MC	
F		Est.$n\Phi_{3.0}^{peak}$	Estimate $\Phi_{3.0}^{peak}$ from n states (n=1,2,...,15)	
G	$\Phi_{2.0}$	$\Phi_{2.0}^{peak}$	Integrated information according to IIT 2.0	[3]
H	$\Phi_{2.5}$	$\Phi_{2.5}^{peak}$	$\Phi_{2.0}$ / $\Phi_{3.0}$ hybrid	[12]
I		D1	Reachable states	[15]
J		D2	Cumulative variance of elements	[15]
K	*S*	*S*	Coalition sample entropy	[13]
L	LZ	LZ	Functional complexity	[13]
M	Φ*	Φ*	Decoder based integrated information	[10]
N	SI	SI	Integrated stochastic interaction	[11]
O	MI	MI	Mutual information	[21]

Abbreviations: S.D.: state dependent; S.I.: state independent; Ref: reference; IIT: integrated information theory; Φ: integrated information; Φ^{peak}: maximum Φ over system states; CO: cut one approximation; NN: no new concepts approximation; AP: all partitions; WS; whole system approximation; IC: iterative cut approximation; Est.n: $\Phi_{3.0}^{peak}$ estimated from n sample states; Dx: state differentiation variant x; *S*: state entropy; LZ: Lempel Ziv complexity; SI: stochastic interaction; MI: mutual information.

2.3.1 Approximations to $\Phi_{3.0}$

We calculated several approximations to $\Phi_{3.0}$. A) the cut one approximation (CO) reduces the number of partitions considered when searching for the MIP. The approximation assumes that the MIP is achieved by only cutting a single node out of the candidate system. B) the no new concepts approximation (NN) eliminates the need to rebuild the entire cause effect structure for every partition under the assumption that when a partition is made it doesn't give rise to new concepts. Thus, one only needs to check for changes to existing mechanisms, rather than reevaluating the entire powerset of potential mechanisms. C) bipartitioning (BP) when finding the MIP is the only partitioning scheme considered in IIT$_{3.0}$. However, it's not clear how partitioning a mechanism in more than two parts would affect $\Phi_{3.0}$. While IIT$_{3.0}$ is defined using BP, a criticism against the theory is that BP is arbitrary and one could use tripartitioning, or more. As such, we tested the default BP versus that of all possible partitions (AP) to investigate how well they correspond. While this causes a massive increase in the number of network element sets to consider, the results might be informative in terms of other more expedient partitioning schemes such as atomic partitioning used by for example Φ* [10].

We also tested two approximations based on *a priori* assumptions about which nodes are included in the MC. These approximations assumed the MC contained D) all the nodes in the system taken as a whole (whole system; WS), or E) the subsystem of the system where all nodes with no recursive connectivity (no input and/or output connections) or only inhibitory inputs (unreachable "on" state) have been removed, iteratively (iterative cut; IC). While IIT$_{3.0}$ already considers pure input or output nodes as outside of a candidate set, the IC approximation is more drastic and also removes nodes that can never be activated, i.e. their "on" state has no cause.

As with $\Phi_{3.0}$, these measures were calculated in a state dependent and state independent manner. Finally, we tested F) if the state independent $\Phi_{3.0}^{peak}$ could be estimated by randomly sampling state dependent $\Phi_{3.0}$, termed here "Est.$n\Phi_{3.0}^{peak}$", where n refers to the number of samples (n=1,2,...15).

2.3.2 Heuristics on discrete networks

To estimate $\Phi_{3.0}$, we investigated several heuristic measures defined for discrete networks. While the latest iteration of IIT takes steps to make the mathematical formalism more in tune with the intended interpretation of its axioms and postulates, IIT$_{3.0}$ is more computationally intractable than previous versions (see S1 of [5]). To compare the results of the two newest versions of the theory, we tested G) Φ based on IIT$_{2.0}$, $\Phi_{2.0}$ [3], and H) $\Phi_{2.0}$ incorporating minimization over both cause-effect and not only cause, $\Phi_{2.5}$ [12]. These measures are, however, still limited by exponential growth in computational time, and are included here because IIT$_{2.0}$ has been used as inspiration for other measures and their validity depends on the correspondence between IIT$_{2.0}$ and IIT$_{3.0}$.

As $\Phi_{3.0}$ is sensitive to a large state repertoire, i.e. divergent and convergent behavior weakening cause/effect constraints (assuming irreducibility), we also included two measures that capture dynamical differentiation of states in the system; I) The number of reachable states, D1, quantifying the system's available repertoire of states, and J) cumulative variance of system elements, D2, indicating the degree of difference between system states [15]. For D1, we calculate the number of states that are reachable, i.e. states that have a valid precursor state. Accordingly, D1 is inversely related to a system's degeneracy of state transitions. D2 calculates the cumulative variance of activity in each system node given the maximum entropy distribution of initial conditions. As such, D2 reflects how different the system's reachable states are from each other. See [15] for a more thorough account.

Both $\Phi_{2.0}$ and $\Phi_{2.5}$ were calculated in a state dependent and state independent manner ($\Phi_{2.0}^{peak}/\Phi_{2.5}^{peak}$), while both D1 and D2 are only defined state independently. All the heuristics on discrete systems were calculated using the system TPM. As such, while these measures are faster to calculate and flexible in terms of network size, they still require full knowledge of the functional dynamics of the system (i.e. the full TPM).

2.3.3 Heuristics from observed data

To alleviate the full knowledge requirement, we consider heuristic measures that are defined for observed (time series) data. Given their relative success in distinguishing conscious from unconscious states in experiments and clinical populations (e.g. [13,22,23]), and their apparent similarity to central IIT intuitions, we focus on measures of signal diversity. There are many candidates to choose from, but here we include K) signal entropy (S), measured by the entropy of the observed state distribution indicating a systems average diversity of visited states [22], and L) signal complexity measured by algorithmic compressibility through Lempel Ziv compression (LZ), indicating a the degree of order or patterns in the observed state sequences of a system [22]. Both entropy and complexity measures have been used in EEG to distinguish between states of consciousness [13,24].

In addition, several measures has been developed that share many of IITs underlying intuitions such as capturing integrated information of a system above and beyond its parts while staying computationally tractable [10,11,19,21,25]. Although these measures can be applied to continuous data in the time domain such as EEG, here we focus on a selection of these measures that can be applied to discrete, binary data. Specifically, we tested: M) decoder based integrated information (Φ^*) based on IIT$_{2.0}$ [21], N) integrated stochastic interaction (SI) based on IIT$_{2.0}$ [11], and O) mutual information (MI) based on IIT$_{1.0}$ [21]. The integrated information measures were implemented using the "Practical PHI toolbox for integrated information analysis" [26].

All heuristics on observed data were calculated on the STM of the systems. As the STM consisted of series of simulated state sequences starting from a perturbation into every possible state of the system, 2^n, state dependent heuristics were based on the sequence of states starting from one perturbed state (state dependent) up to the next, as if it was an epoched time series of size n-by-$\alpha(n)2^6$. For the state independent variants, the whole STM (all perturbations and state sequences) were used as a basis as if it was a single time series of size n-by-$\alpha(n)2^{6+n}$.

2.4. Analysis

Comparisons between $\Phi_{3.0}$ and approximate measures (CO, NN, BP, WS, IC) were analysed using Pearson correlations (r) and separate ordinary least squares linear regression models as the approximations are essentially the same as $\Phi_{3.0}$. Statistics of linear fits are reported. For comparisons between $\Phi_{3.0}$ and all other measures we used Spearman's correlation (r_s) to investigate the monotonicity of the relationship as a linear relationship is not necessarily expected. As $\Phi_{3.0}$ is state dependent, all measures except Est.$n\Phi_{3.0}^{peak}$, D1 and D2, were compared in both a state dependent (vs. $\Phi_{3.0}$) and state independent manner (vs. $\Phi_{3.0}^{peak}$).

Significance values are not reported as the high n (1981) means $r_s > 0.065$ is significant to the order of $p < 10^{-54}$. As we are here focused on high correspondence, we instead defined correlations as weak, $0.5 < r < 0.7$, medium $0.7 < r < 0.8$, strong $0.8 < r < 0.9$, and very strong, $r > 0.9$ (for both r and r_s).

2.5 Setup

Calculation of measurements were performed in Python (v3.6) with PyPHI (v1.0) [6] for $\Phi_{3.0}$, CO, NN, WS, and IC; Matlab (v2016b) with "Practical PHI toolbox for integrated information analysis" (v1.0) [26] for Φ^*, SI, MI; custom code in Python (v3.6) for $\Phi_{2.0}$, $\Phi_{2.5}$, D1, D2; and Python (v3.6) with scripts from [13] for LZ, and *S*. Statistics were done with custom code in Python (v3.6) and Statsmodels (v.0.8.0). Everything else was done with custom code in Python (v3.6), Numpy (v1.13.1), SciPy (v0.19.1), and Pandas (v0.20.3).

3. Results

We analyzed 2032 randomly generated networks, with 131 three node, 675 four node, 866 five node, and 360 six node networks. In total, 61224 states were analyzed. Not all measures were applied to the full array of networks; e.g. the comparison of bi-partitioning vs. all-partitions was not applied to six node networks due to time constraints. See Table 2 for an overview of the main results, and Figure 2 for four example networks.

3.1. Descriptive Statistics

Mean state dependent $\Phi_{3.0}$ grew as a function of network elements ($n = 3$: M = 0.015 ± 0.121SD to $n = 6$: M = 0.386 ± 0.487SD). As systems increased in size, the fraction of $\Phi_{3.0}^{peak}= 0$ networks (indicating a completely reducible system, e.g. a feed forward network) decreased, $\Phi_{3.0}^{peak}= 1$ networks (indicating the system is reducible to or consist of a stereotyped "loop"/"circular" network) stayed relatively stable, while the fraction of networks with $\Phi_{3.0}^{peak}> 1$ increased. See Figure 3.

3.2 Approximations

Both, the no-new concepts (NN) and cut-one (CO) approximations, were nearly perfectly correlated with state dependent (S.D.) $\Phi_{3.0}$ and state independent (S.I.) $\Phi_{3.0}^{peak}$ ($r > 0.996$). Regression analysis showed that no-new concepts was a stronger linear predictor (S.I.: F(1, 1979) = 5.9x10^7, $p < 0.0001$, $R^2 > 0.999$, $\Phi_{3.0}^{peak} = 0.00 + 1.00NN\Phi_{3.0}^{peak}$. S.D.: F(1, 57987) = 7.3x10^8, $p < 0.0001$, $R^2 > 0.999$, $\Phi_{3.0} = 1.00NN\Phi_{3.0}$.), than cut-one (S.I.: F(1, 1979) = 3.1x10^5, $p < 0.0001$, $R^2= 0.994$, $\Phi_{3.0}^{peak} = 0.00 + 1.04CO\Phi_{3.0}^{peak}$). S.D.: F(1, 57987) = 1.2x10^7, $p < 0.0001$, $R^2 = 0.995$, $\Phi_{3.0} = 1.02CO\Phi_{3.0}$). See Figure 4.

As for partitioning, bi-partitioning was a very strong linear predictor of all-partitioning both with S.I. $\Phi_{3.0}^{peak}$ (F(1, 1670) = 4.77x10^4, $p < 0.0001$, $R^2 = 0.966$, $AP\Phi_{3.0}^{peak}= -0.134 + 1.438BP\Phi_{3.0}^{peak}$), and with S.D. $\Phi_{3.0}$ (F(1, 38407) = 4.49x10^5, $p < 0.0001$, $R^2 = 0.921$, $\Phi_{3.0} = -0.033 + 1.541BP\Phi_{3.0}$). We also observed significantly higher $\Phi_{3.0}$ values for all-partitioning with relative increase of S.D. (M = 32.29 ± 150.11%, $t = -5.77$, $p < 0.0001$) and S.I.: (M = 16.21 ± 28.60, $t = -21.16$, $p < 0.0001$). See Figure 5a,b.

In regards to estimating $\Phi_{3.0}^{peak}$, we took samples from $n = 1,2,...,15$ states with results ranging from weak correlation ($n = 1$, $r = 0.688$) to strong correlation ($n = 15$, $r = 0.893$) as the number of samples increased (for n=5; F(1, 2030) = 5090.0, $p < 0.001$, $R^2 = 0.738$,

$\Phi_{3.0}^{peak}$= 0.097 + 0.262$SS\Phi_{3.0}$). This was in accordance with a very strong correlation between $\Phi_{3.0}^{peak}$and $\Phi_{3.0}^{mean}$ (F(1, 2030) = 1.12x10^4, p < 0.001, R^2 > 0.846, $\Phi_{3.0}^{peak}$ = 0.087 + 0.274$\Phi_{3.0}^{mean}$). See Figure 5c,d.

Finally, we tested whether a priori defined MCs could predict $\Phi_{3.0}$. WS$\Phi_{3.0}^{peak}$ was very strongly correlated with S.I.$\Phi_{3.0}^{peak}$ (F(1, 2030) = 4.2x10^4, p < 0.001, R^2 > 0.954, with $\Phi_{3.0}^{peak}$ = -0.255 + 0.986$WS\Phi_{3.0}^{peak}$) and with S.D. $\Phi_{3.0}$ (F(1, 61222) = 4.3x10^5. p < 0.001, R^2 > 0.876, with $\Phi_{3.0}$ = -0.163 + 0.899$WS\Phi_{3.0}$). IC$\Phi_{3.0}$ was very strongly correlated with S.I.$\Phi_{3.0}^{peak}$ (F(1, 1979) = 7.53x10^4, p < 0.001, R^2 > 0.974, with $\Phi_{3.0}^{peak}$= -0.167 + 0.995$IC\Phi_{3.0}^{peak}$) and very strongly correlated with $\Phi_{3.0}$ (F(1, 50859) = 5.3x10^5. p < 0.001, R^2 > 0.912, with $\Phi_{3.0}$ = -0.119 + 0.927$IC\Phi_{3.0}$). See Figure 6.

Together, these results suggest that the tested approximations can be used as strong predictors of Φ, however, these approximation still need knowledge of the systems TPM and their computational cost grows exponentially, leading to only a marginal increase in the size of networks that can be analyzed. See Appendix A4 for an overview of estimates of computational time for each of the measures.

3.3 Heuristics

State differentiation measures D1 and D2 showed strong (r_s = 0.827) and medium (r_s = 0.718) rank order correlations with S.I.$\Phi_{3.0}^{peak}$ respectively. See Figure 6e,f.

S.D. $\Phi_{2.0}$ and $\Phi_{2.5}$ were weakly or less correlated with $\Phi_{3.0}$ (r_s = 0.622 and r_s = 0.473, respectively), while S.I. variants of $\Phi_{2.0}$ and $\Phi_{2.5}$ were strongly rank order correlated with $\Phi_{3.0}^{peak}$(r_s = 0.838 and r_s = 0.832, respectively). See Figure 7a,b.

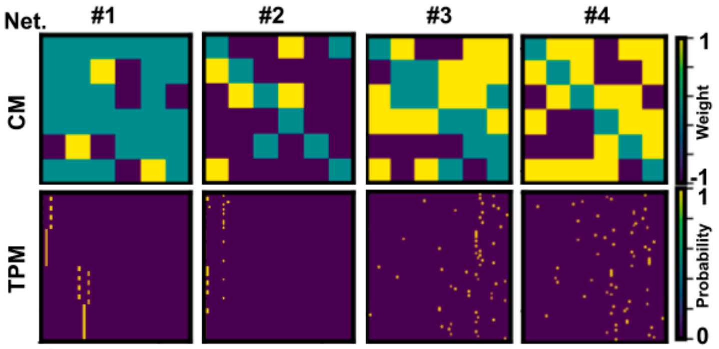
S.D. LZ complexity and signal entropy (S) were less than weakly correlated with $\Phi_{3.0}$ (r_s < 0.5). However, S.I. LZ and S were medium correlated with $\Phi_{3.0}^{peak}$ (0.71 < r_s < 0.72). See Figure 7c (only LZ shown).

For the alternative measures Φ*, SI, and MI, the S.D. variants were not correlated with $\Phi_{3.0}$ (r_s < 0.16), S.I. variants were weakly or less correlated with $\Phi_{3.0}^{peak}$ (r_s < 0.54), except for S.I. Φ* being strongly rank order correlated with $\Phi_{3.0}^{peak}$, (r_s = 0.82). See Figure 7d (only Φ* shown). For Φ*, evident in the results is two clusters of values, one seemingly linearly related to $\Phi_{3.0}^{peak}$, and one non-correlated cluster consisting of low $\Phi_{3.0}^{peak}$/high Φ* outliers. A post-hoc analysis removing outliers above two standard deviations of the mean negligibly influenced results (see Appendix A2).

Together, these results suggest that the tested heuristics might be accurate predictors of $\Phi_{3.0}^{peak}$on a group level, however not necessarily on an individual level. In addition, all heuristics show an increased variance of $\Phi_{3.0}^{peak}$with higher values, suggesting reduced correspondence for higher values. These heuristics can drastically reduce computational demands, see Appendix A4.

3.4 Post-hoc tests

For all measures, removing non-integrated ($\Phi_{3.0}^{peak}$=0) or irreducible circular networks ($\Phi_{3.0}^{peak}$=1), reduced correlational values. This was true for all heuristics, while the approximations were minimally affected. After this adjustment, S.I. D1 and Φ* were the heuristics highest correlated with $\Phi_{3.0}^{peak}$ (r_s = 0.703 and r_s=0.698, respectively), with LZ the third (r_s = 0.616). This indicates that the results are influenced by a large cluster of non-integrated and circular networks, and that they are sensitive to the difference between them. See Appendix A3.



	#1	#2	#3	#4
$\Phi_{3.0}^{peak}$	0.00	1.00	18.67	30.85
$\Phi_{2.5}^{peak}$	2.00	4.77	6.00	6.00
$\Phi*$	0.00	0.01	0.05	0.07
LZ	561	508	936	1021
D1	6	5	25	31
D2	0.69	0.33	1.22	1.40

Figure 2. Four example networks with connection matrices (CM) and transition probability matrices (TPM), with $\Phi_{3.0}^{peak}$ and corresponding values for selected state independent heuristics.

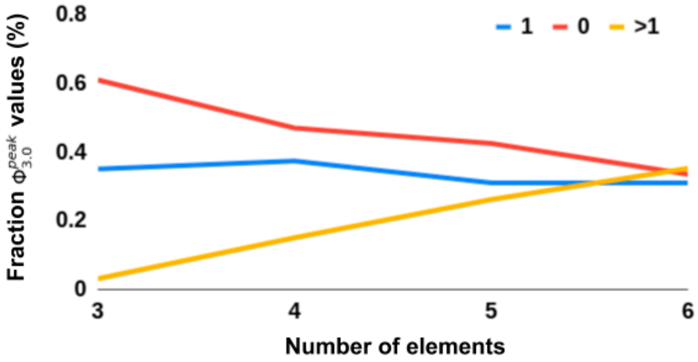


Figure 3. Overview of fraction of networks with $\Phi_{3.0}^{peak} \in \{1, 0, >1\}$.

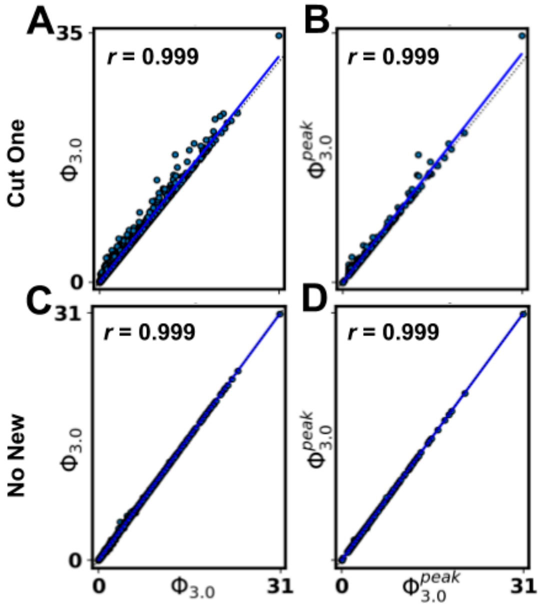


Figure 4. Results of comparison between $\Phi_{3.0}$ and approximations, with plotted linear fit (blue) and one-to-one relationship (dotted); (**A**) $\Phi_{3.0}$ of the state dependent Cut One approximation (CO), (**B**) $\Phi^{peak}_{3.0}$ of the state independent CO, (**C**) $\Phi_{3.0}$ of the state dependent No New concepts approximation (NN), (**D**) $\Phi^{peak}_{3.0}$ of the state independent NN.

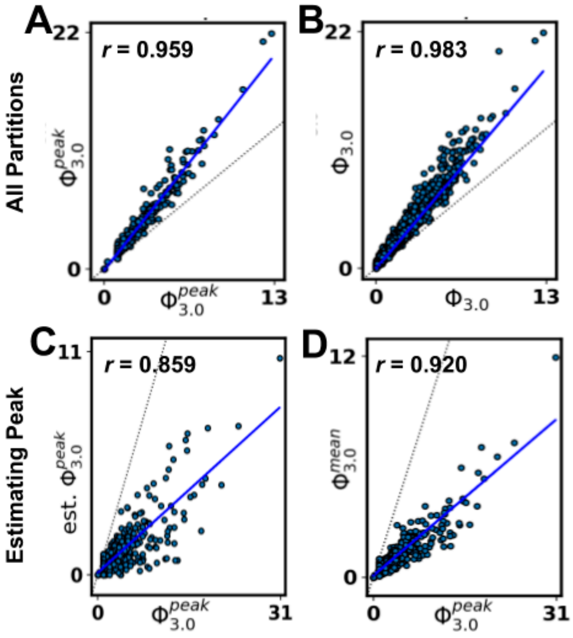


Figure 5. Results of comparison between $\Phi_{3.0}$ and approximations, with plotted linear fit (blue) and one-to-one relationship (dotted); (**A**) $\Phi^{peak}_{3.0}$ of the state independent All Partitioning (AP) approximation, (**B**) $\Phi_{3.0}$ of the state dependent AP, (**C**) the estimated state independent $\Phi^{peak}_{3.0}$using 5 randomly sampled states (**D**) state independent $\Phi^{mean}_{3.0}$.

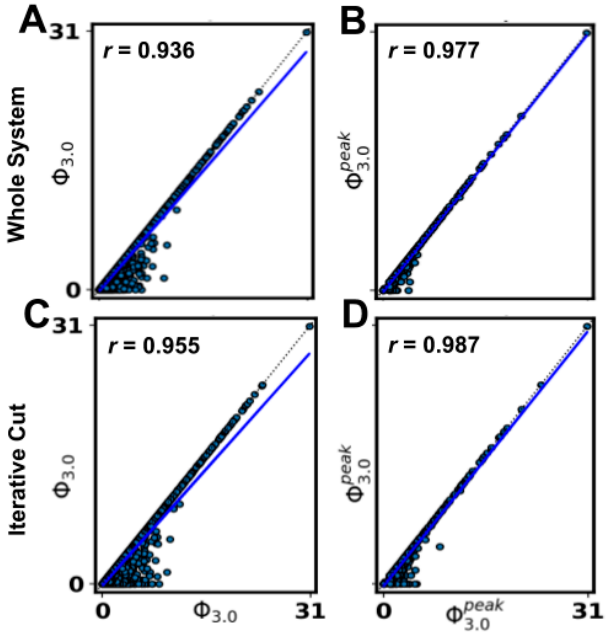


Figure 6. Results of comparison between Φ3.0 and approximations, with plotted linear fit (blue) and one-to-one relationship (dotted); (**A**) Φ3.0 of the state dependent Whole System (WS) estimated main complex, (**B**) $\Phi_{3.0}^{peak}$ of the state independent WS, (**C**) Φ3.0 of the state dependent Iterative Cut (IC) estimated main complex, (**D**) $\Phi_{3.0}^{peak}$ of the state independent IC.

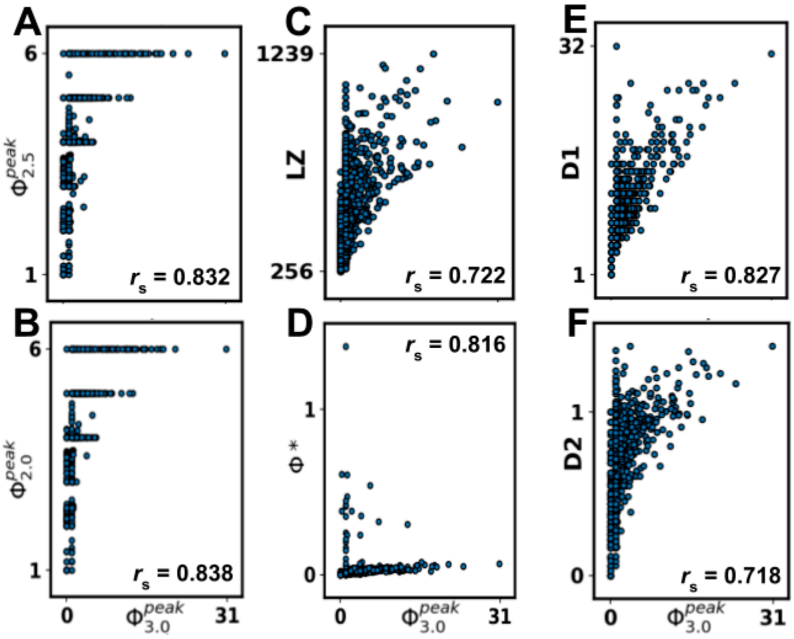


Figure 7. Results of comparison between state independent $\Phi_{3.0}^{peak}$ and heuristics of $\Phi_{3.0}^{peak}$; (**A**) Φ2.5 modified from Φ2.0, (**B**) Φ2.0 based on IIT2.0, (**C**) Lempel Ziv (LZ) complexity (non-normalized), (**D**) decoder based Φ*, based on Φ2.0, (**E**) state differentiation D1, (**F**) cumulative variance of system elements D2.

Table 2. Overview of results

#	S.D. Measure	*r*	S.I. Measure	*r*
	$\Phi_{3.0}$		$\Phi_{3.0}^{peak}$	
A	CO $\Phi_{3.0}$	.999	CO $\Phi_{3.0}^{peak}$	.999
B	NN $\Phi_{3.0}$	.999	NN$\Phi_{3.0}^{peak}$	.999
C	AP $\Phi_{3.0}$	.983	AP$\Phi_{3.0}^{peak}$	.959
D	WS $\Phi_{3.0}$	.936	WS$\Phi_{3.0}^{peak}$	.977
E	IC $\Phi_{3.0}$	.955	IC$\Phi_{3.0}^{peak}$	.987
F			Est$_5\Phi_{3.0}$	.859
G	$\Phi_{2.0}$	.622	$\Phi_{2.0}^{peak}$	.838
H	$\Phi_{2.5}$	.473	$\Phi_{2.5}^{peak}$	.832
I			D1	.827
J			D2	.718
K	S	-.063	S	.711
L	C	.411	C	.722
M	Φ*	-.076	Φ*	.816
N	SI	.156	SI	.537
O	MI	.118	MI	.306

Abbreviations: *r*: correlation values with measures A-F using Pearson's *r* and G-O using Spearman's r_s; S.D.: state dependent; S.I.: state independent; Φ: integrated information; Φ^{peak}: maximum Φ over system states; CO: cut one approximation; NN: no new concepts approximation; AP: all partitions; WS; whole system approximation; IC: iterative cut approximation; Est$_5$: $\Phi_{3.0}^{peak}$ estimated from 5 sample states; Dx: state differentiation variant x; *S*: state entropy; LZ: Lempel Ziv complexity; SI: stochastic interaction; MI: mutual information.

4. Discussion

In summary, we observed that all the computational approximations that we tested were very strong predictors of both the state dependent $\Phi_{3.0}$ and the state independent $\Phi_{3.0}^{peak}$. However, while they slightly decrease computational time, they still incur exponential growth in computational time as a function of system size, and are therefore also only applicable to smaller model systems. On the other hand, all heuristic measures were only weakly correlated or less with $\Phi_{3.0}$ when compared in a state dependent manner.

However, doing the comparison on a state independent level, showed a medium or stronger correlation with $\Phi_{3.0}^{peak}$ for all measures (except for SI and MI). However, all heuristic measures were negatively impacted by removing networks with $\Phi_{3.0}^{peak}$= 0||1, indicating that reducible ($\Phi_{3.0}^{peak}$= 0) or circular ($\Phi_{3.0}^{peak}$= 1) networks can confound comparisons as a majority of networks fall in this range. The heuristics that showed the strongest correlation after removal of $\Phi_{3.0}^{peak}$=0||1 networks were measures of state differentiation (D1), complexity (LZ), and integrated information (Φ*). Together, this suggests that both D1, Φ*, and to a lesser degree LZ, could be useful heuristics for $\Phi_{3.0}^{peak}$ at the group level, although unreliable at an individual level.

The heuristic D1 measures the amount of accessible states of a system [15], and the strong correlation we observed indicates that systems with a large repertoire of available states are also likely to have high $\Phi_{3.0}^{peak}$ (assuming the systems are irreducible, i.e. $\Phi_{3.0}^{peak}$> 0). This finding is interesting because clinical results also corroborate state differentiation as a factor in unconsciousness where it has been observed that the state repertoire of the brain is reduced during anesthesia [27]. While D1 is computationally tractable, it requires full knowledge of the system (i.e. a TPM with 2^{2n} bits of information), that the system is integrated, and that transitions are relatively noise free. As such D1 can unfortunately not be applied to larger artificial or biological systems of interest (such as the brain). The second measure that correlated well with $\Phi_{3.0}^{peak}$ can also be seen to quantify state differentiation to some extent. LZ is a measure of signal complexity [28], offering a concrete algorithm to quantify the number of unique patterns in a signal. While LZ has been used to differentiate conscious and unconscious states [13,29], it cannot distinguish between a noisy system and an integrated but complex one from observed data alone. Thus, some knowledge of the structure of the system in question is required for its interpretation. In addition, while LZ allows for analysis of real systems based on time series data, it's also the measure that is the furthest removed from IIT (but see [14]). It also requires binary (or binarized) data, is highly dependent on the size of the input, and is hard to interpret without normalization which makes it difficult to compare systems of varying size. Finally, the measure Φ* is aimed at providing a tractable measure of integrated information using mismatched decoding, and is applicable to time series data, both discrete and continuous [10]. Φ* is both relatively fast to compute and can be applied to continuous time series like EEG. However, while we observed a high correlation with $\Phi_{3.0}^{peak}$, a cluster of high Φ* values with corresponding low $\Phi_{3.0}^{peak}$ values limit interpretation. This suggests that Φ* might not be reliable for low$\Phi_{3.0}^{peak}$networks, but analysis of larger networks is needed to draw a conclusion. While the results do not suggest a clear tractable alternative to $\Phi_{3.0}$, several of the measures could be useful in statistical comparisons of groups of networks.

In addition to the heuristics, there were two approximations that could be indicative of future lines of research; a) investigating a full partitioning scheme vs. bipartitioning as employed by $IIT_{3.0}$, and, b) trying to suggest *a priori* MC candidates rather than searching through all possible candidate subsystems rigorously. For a), we observed a very high correspondence between $\Phi_{3.0}$ when employing full and bipartioning schemes. This could indicate that even if bipartitioning may be an arbitrary limitation in the formalism, it's not a limitation that strongly affects the rank order of $\Phi_{3.0}$ values. In addition, this might also indicate that future research focusing on more efficient partitioning schemes or MIP searches could be fruitful (e.g. see [30–32]). For b), we showed that the WS approximation ($\Phi_{3.0}$ of an MC produced by only considering the whole system) was highly predictive of $\Phi_{3.0}$. Implementing such an approximation would eliminate the requirement for knowing the full structural connectivity of a system, as $\Phi_{3.0}$ can be calculated under the assumption of a fully connected network (at an increased computational cost) [6]. The IC approximation employing the rule of "cut pure input/output nodes or nodes with only inhibitory incoming connections, iteratively" eliminates the chance of including pure input/output nodes while reducing the size of a candidate set beyond that already implemented in PyPhi [6], however, at the cost of requiring more prior structural knowledge. The IC approximation was highly predictive of $\Phi_{3.0}$ of the actual MC, butthe computational savings are only as big as the reduction in number of nodes, and might thus only offer minimal savings if the size of the estimated MC is close to the size of the full network. Generally, regarding *a priori* estimates of the MC, any estimated $\Phi_{3.0}$ value is a lower bound on $\Phi_{3.0}$ of the actual MC. Therefore, the higher the estimated value is (relative to an upper bound), the more likely it is to be accurate. However, as $\Phi_{3.0}$ grows as a function of network size, so does the computational cost. Any MC estimate close to the size of the whole system will provide relatively little savings. Therefore, finding methods for maximally constraining a likely candidate set, would be most efficient for reducing computational demands. While this limits the usability of *a priori* MC estimates for highly integrated systems, such methods could be used to investigate questions regarding which part of a system is conscious (e.g. cortical location of consciousness [33]). The two methods used here to define *a priori* MC candidates are by no means exhaustive, but are suggestive that further work could be profitable.

Prior work directly comparing $\Phi_{3.0}$ with candidate alternatives reported lower correlations (complexity, state entropy, and differentiation) than observed here [15]. There are at least four possible reasons for this, (a) the current work considers a wider variety of networks in terms of size and structural parameters, (b) we compared against $\Phi_{3.0}^{peak}$and not $\Phi_{3.0}^{mean}$, (c) we considered only the whole system as a basis for the heuristics and not only the elements of the MC, and, (d) the networks used here did not ensure integration or connectedness of the systems, which could produce results that are strictly not representative for a measure. For (a), it suggests the need to test measures in a wider range of systems as done here. For (b) we reran the analysis replacing $\Phi_{3.0}^{peak}$ with $\Phi_{3.0}^{mean}$ producing negligible deviances in results (not shown). For (c) and (d), using the whole system as a basis for the heuristics (except for $\Phi_{2.0}$ and $\Phi_{2.5}$), could influence results that would be valid if applied to the MC. However, this would entail first finding the MC which would increase computational cost, although, a priori estimates could be beneficial here. Follow up analysis using the WS$\Phi_{3.0}^{peak}$approximation instead of $\Phi_{3.0}^{peak}$ showed no improvements in rank order correlations among the heuristics (results not shown), suggesting that it's not necessarily the candidate set considered in the heuristics that is most relevant.

Further, removing all non-integrated networks ($\Phi_{3.0}^{peak}$= 0) and circular networks or reducible to circular subnetworks ($\Phi_{3.0}^{peak}$= 1) reduced correlations between all heuristics and $\Phi_{3.0}^{peak}$(see Appendix A3) closer to that observed in [15]. This suggests that such networks are indeed relevant to consider as both that finding a tractable measure that seperates $\Phi_{3.0}^{peak}$= 0 and $\Phi_{3.0}^{peak}$= > 0 networks would be useful in its own right, and that the employed heuristics might not be as accurate for higher $\Phi_{3.0}^{peak}$. Evident in the results was that all heuristics except S, SI, and MI, showed an inverse predictability with $\Phi_{3.0}^{peak}$, i.e. low scores on a given heuristic corresponded to a low score on $\Phi_{3.0}^{peak}$but the higher the scores the larger the spread of $\Phi_{3.0}^{peak}$ (see Figure 7). This could explain why correlations drop when removing networks with $\Phi_{3.0}^{peak} \leq 1$. This inverse predictability indicates two things. First, that the tested measures could be useful as negative markers, that is, low scores on measures can indicate low $\Phi_{3.0}^{peak}$ networks, but not the converse. Secondly, it suggests $\Phi_{3.0}^{peak}$ has dependencies on aspects of the underlying network that is not captured by any of the heuristic measures.

While there are several conceptual differences between $\Phi_{3.0}$ and the other measures employed here (outside the scope of the current paper), there are some methodological aspects worth considering. One is that only the approximations and heuristics of discrete networks are calculated on the full state TPM, while the heuristics on observed data employed here use the STM. While in deterministic networks, the STM and TPM contain the same information, given that the system is initialized to all possible states at least once, the STM might be "insufficient" as a time series, e.g. systems often converge on a few cyclical states. Conversely, it might contain irrelevant information if, as discussed above, the heuristics should have been applied to only the information from nodes that are integrated in a connected system, e.g. high values on measures such as D1 and LZ doesn't entail irreducibility.

Other studies comparing measures of information integration, differentiation, and complexity, have also observed both qualitative and quantitative differences between the measures even for simple systems [19,20]. Thus, there might be a large number of networks where the tested heuristics would correspond to $\Phi_{3.0}$ if only certain prerequisites are met, such as a certain degree of irreducibility or small-worldedness. One could for example imagine systems that have evolved to become both highly integrated and interact with an environment [34]. Such evolved networks might have further qualities than merely integratedness, such as meaningful state differentiation that serves distinctive roles for the system, i.e. differences that makes a difference to an organism, which is an important concept in IIT (although used differently in the theory) [5]. Given such properties, it could be probable that such a system would also score high on measures capturing integrated information. While it is still an open question what $\Phi_{3.0}$ captures of the underlying network above that of the heuristics considered here, investigation into structural and functional aspects that lead to systems with high $\Phi_{3.0}$ could further point to avenues for developing new measures inspired by IIT.

While estimates of the upper bound of $\Phi_{3.0}$ given system size have been proposed (e.g. see [15]), not much is known about the actual distribution of $\Phi_{3.0}$ over different network types and topologies. Here we explored a varietyof network topologies, but the system properties, such as weights, noise, thresholds, element types, and so on, were omitted due to the limited scope of the paper. Investigating the relation between such network properties and $\Phi_{3.0}$ would be an interesting research project moving forward. This could be useful as a testbed for future IIT inspired measure and be informative about what kind of properties could be important for high $\Phi_{3.0}$ in biological systems, and properties to aim for in artificial systems, to produce "awareness". In addition, there are several approximations and heuristics not included in the present study [11,12,19,30,35–40], some of which are specifically applicable to time series data [10–12,19,21,30,40]. Accordingly, the present work should not be considered an exhaustive exploration of $\Phi_{3.0}$ correlates.

In sum, while many of the measures employed here have shown promising results for tracking altered states of consciousness, in accordance with IIT predictions [13,16,17,29,41], the goal in this paper was not to relate the proposed measures to states of consciousness but to relate the proposed measures to $\Phi_{3.0}$ in tractable situations. Based on the results there are two main findings: 1) most of the approximations employed here were highly predictive of $\Phi_{3.0}$ but only granting a minor increase in the size of systems available for study, and 2) several of the heuristics employed (notably LZ, D1, and Φ*) could drastically reduce computational time, but might only be useful at a group level or as an indicator of low $\Phi_{3.0}$ networks. However, as the current project is limited in the scope of networks considered, we cannot rule out a higher correspondence between the applied heuristics and $\Phi_{3.0}$ in different classes of systems. Finally, we note that as the ground truth with respect to $\Phi_{3.0}$ cannot currently be established in larger or continuous (temporal/dynamical) systems, we argue that for a heuristic to truly capture underlying aspects of IIT, it should as a minimal requirement be validated in smaller deterministic binary networks as considered here.

Author Contributions: Conceptualization, A.S.N.; methodology, A.S.N.; software, A.S.N.; validation, A.S.N.; formal analysis, A.S.N.; investigation, A.S.N., B.E.J., and W.M.; writing—original draft preparation, A.S.N.; writing—review and editing, A.S.N., B.E.J., W.M., and J.F.S.; visualization, A.S.N; supervision, W.M. and J.F.S.; project administration, J.F.S; funding acquisition, J.F.S..

Funding: This study was supported by the European Union's Horizon 2020 research and innovation programme under grant agreement 7202070 (Human Brain Project (HBP)) and the Norwegian Research Council (NRC: 262950/F20 and 214079/F20)

Acknowledgments: We appreciate the discussions and theoretical contributions of Larissa Albantakis, PhD, Benjamin Thürer, PhD, and Arnfinn Aamodt.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Appendix

A1. State transition matrix size

Each network were perturbed into each possible initial condition (2^n states) and simulated for $\alpha(n)2^6$ timesteps, resulting in a node (n) by time ($\alpha(n)n2^{6n}$) state transition matrix (STM). The adjustment factor a is required to force the same number of bits in the STM to ensure that measures employing the STM (Φ*, LZ, S, SI, MI) is calculate on the same amount of data. The adjustment factor a is a function of *n:*

$$\alpha(n) = \frac{s_{max}}{n^2 * 2^{7n}} \tag{1}$$

$$s_{max} = n_{max} 2^{2n_{max}} \tag{2}$$

where n_{max} is nodes in the biggest network generated (here six nodes), and n is number of nodes in current network.

A2. Φ post-hoc analysis*

To investigate the distribution of Φ* relative to $\Phi_{3.0}^{peak}$ after removing a cluster of high Φ* and low $\Phi_{3.0}^{peak}$ values, we removed outliers above two standard deviations of the mean. This did not improve results drastically as the bulk of observations lie within a narrow band of low Φ* values. See Figure A1.

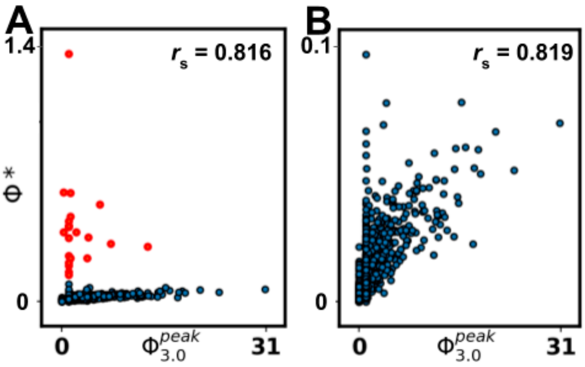


Figure A1. Results of comparison between state independent Φ* and $\Phi_{3.0}^{peak}$ with and without outliers; r_s=.816 and r_s=.819, respectively. (**A**) scatter with outliers > two standard deviations marked in red, (**B**) scatter with outliers > two standard deviations removed.

A3. Post-hoc analysis of networks not totally reducible or reducible to circular systems

Systems that are completely reducible ($\Phi_{3.0}^{peak}$= 0) or reducible to a circular or ring shaped (sub)network ($\Phi_{3.0}^{peak}$= 1), might not be representative for candidate heuristics as these networks can be considered "special" cases in terms of $IIT_{3.0}$. The absolute difference in correlation values can be seen in Table A1 and the corresponding scatter plots of some select measures in Figure A2. Most measures drop in correlational values, while those that increase are low to begin with. Only measures A to F have an *r*>.8, while measure I and M stay close to r_s=.7. The other measures have r_s<.65. This suggests that the reported correlational values for most heuristics (G to O) are primarily driven by a cluster of non or trivially integrated networks ($\Phi_{3.0}^{peak}$= 0||1). For measures G to O, Spearman's rank order correlation has been used, Pearson's correlation otherwise.

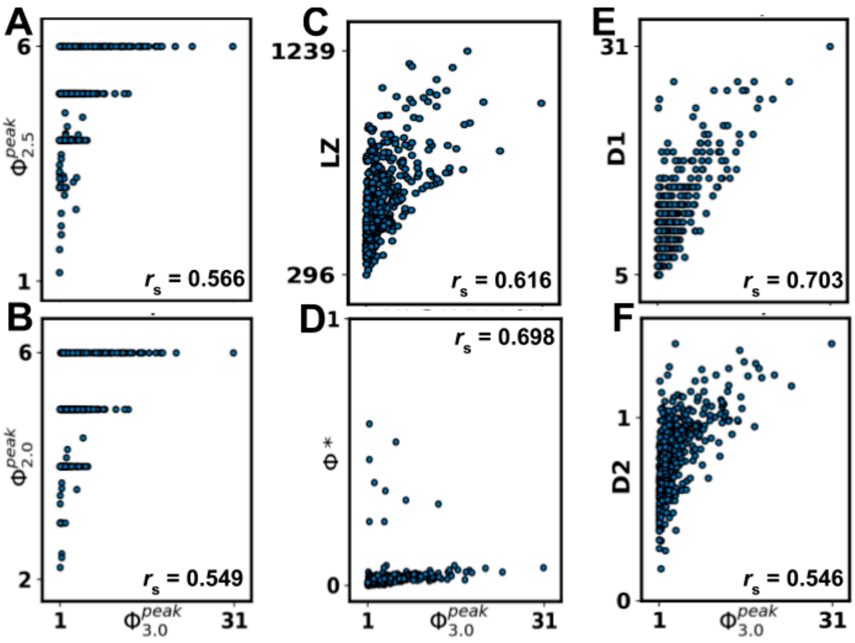


Figure A2. Results of comparison between state independent $\Phi_{3.0}^{peak}$ and heuristics of $\Phi_{3.0}^{peak}$ after all networks with $\Phi_{3.0}$ =$\Phi_{3.0}^{peak}$ ≤ 1 were removed; (**A**) Φ based on $\Phi_{2.5}$, (**B**) Φ based on $\Phi_{2.0}$, (**C**) Lempel Ziv (LZ) complexity (non-normalized), (**D**) Φ*, e) state differentiation D1, f) cumulative variance of system elements D2.

Table A1. Difference in results after removing networks and states with $\Phi_{3.0} = \Phi_{3.0}^{peak} \leq 1$

#	S.D. Measure	*r*	S.I. Measure	*r*
	$\Phi_{3.0}$		$\Phi_{3.0}^{peak}$	
A	CO $\Phi_{3.0}$	-.004	CO $\Phi_{3.0}^{peak}$	-.004
B	NN $\Phi_{3.0}$	.000	NN$\Phi_{3.0}^{peak}$	.000
C	AP $\Phi_{3.0}$	-.029	AP$\Phi_{3.0}^{peak}$	.014
D	WS $\Phi_{3.0}$	.027	WS$\Phi_{3.0}^{peak}$	.006
E	IC $\Phi_{3.0}$	.021	IC$\Phi_{3.0}^{peak}$	.006
F			Est$_5\Phi_{3.0}$	-.080
G	$\Phi_{2.0}$	-.396	$\Phi_{2.0}^{peak}$	-.289
H	$\Phi_{2.5}$	-.491	$\Phi_{2.5}^{peak}$	-.266
I			D1	-.124
J			D2	-.172
K	S	.153	S	-.405
L	LZ	-.376	LZ	-.106
M	Φ*	.015	Φ*	-.118
N	SI	-.086	SI	.063
O	MI	-.061	MI	.010

Abbreviations: *r*: correlation values with measures A-F using Pearson's *r* and G-O using Spearman's r_s; S.D.: state dependent; S.I.: state independent; Φ: integrated information; Φ^{peak}: maximum Φ over system states; CO: cut one approximation; NN: no new concepts approximation; AP: all partitions; WS; whole system approximation; IC: iterative cut approximation; Est$_5$: $\Phi_{3.0}^{peak}$ estimated from 5 sample states; Dx: state differentiation variant x; *S*: state entropy; LZ: Lempel Ziv complexity; SI: stochastic interaction; MI: mutual information.

A4. Estimated computational demands

To estimate computational demands, seven networks of each $n \in \{3,4,5,6\}$ were randomly generated, with $p(W_{ij}=1) \in \{0.7,0.8,0.9,1.0\}$, and $p(W_{ij}=-1) \in \{0.3,0.4,0.5\}$. The average times was recorder for each measure, then fitted to a logarithmic regression with reported exponent *x*, in the form of $time = bn^x$, where *b* is a constant, and *n* is system size in nodes. In

Peer-reviewed version available at Entropy 2019, 21, 525; doi:10.3390/e21050525

essence, x>1 indicates exponential (more than linear) increase, while x<1 indicates less than linear increase. The reported exponents, especially for the measures of Φ, are likely underestimated. However, these estimates are highly dependent on underlying computational power, parallelization, efficiency of algorithmic implementation, as well as utilization of shortcuts. As such, estimated computational demands are guiding at best. See Table A2 for average time taken to compute measure (in seconds) for each network size, and fitted logarithmic regression, and Figure A3 for an overview of the relationship between computational time and correlation with $\Phi_{3.0}^{peak}$.

Table A2. Estimated computational demands

#	**Measure**	***t(n=3)***	***t(n=4)***	***t(n=5)***	***t(n=6)***	***x***
	$\Phi_{3.0}^{peak}$	0.40	1.51	102.67	9397.08	31.13
A	CO $\Phi_{3.0}^{peak}$	0.39	1.22	26.61	874.54	13.74
B	NN$\Phi_{3.0}^{peak}$	0.35	1.27	68.61	7070.00	29.06
C	AP$\Phi_{3.0}^{peak}$	0.44	3.98	2021.79	...	67.73
D	WS$\Phi_{3.0}^{peak}$	0.32	1.41	91.57	8379.66	32.33
E	IC$\Phi_{3.0}^{peak}$	0.31	1.18	74.64	8175.50	31.56
F	Est$_5\Phi_{3.0}$	0.08	0.30	20.53	1879.41	31.13
G	$\Phi_{2.0}^{peak}$	0.39	1.49	27.60	691.91	12.59
H	$\Phi_{2.5}^{peak}$	0.37	1.90	33.12	850.35	13.47
I	D1	0.00003	0.00003	0.00004	0.0001	1.43
J	D2	0.000	0.001	0.001	0.002	1.64
K	S	0.005	0.005	0.006	0.007	1.14
L	LZ	0.03	0.02	0.02	0.02	0.99
M	Φ*	0.23	0.29	0.43	0.60	1.38
N	SI	0.20	0.29	0.38	0.38	1.24
O	MI	0.17	0.22	0.18	0.06	0.71

Abbreviations: *t(n=i)*: time in seconds to calculate the relevant measure for a system of size n=3,4,5,6; x: exponent in a logarithmic regression fit of the form *time*=bn^x, where n is system size in nodes, and b is a constant (not reported); Φ: integrated information; Φ^{peak}: maximum Φ over system states; CO: cut one approximation; NN: no new concepts approximation; AP: all partitions; WS; whole system approximation; IC: iterative cut approximation; Est$_5$: $\Phi_{3.0}^{peak}$ estimated from 5 sample states; D1/2: state differentiation measure 1/2; *S*: state entropy; LZ: Lempel Ziv complexity; SI: stochastic interaction; MI: mutual information; ^{all}: measure applied to the whole state transition matrix (STM).

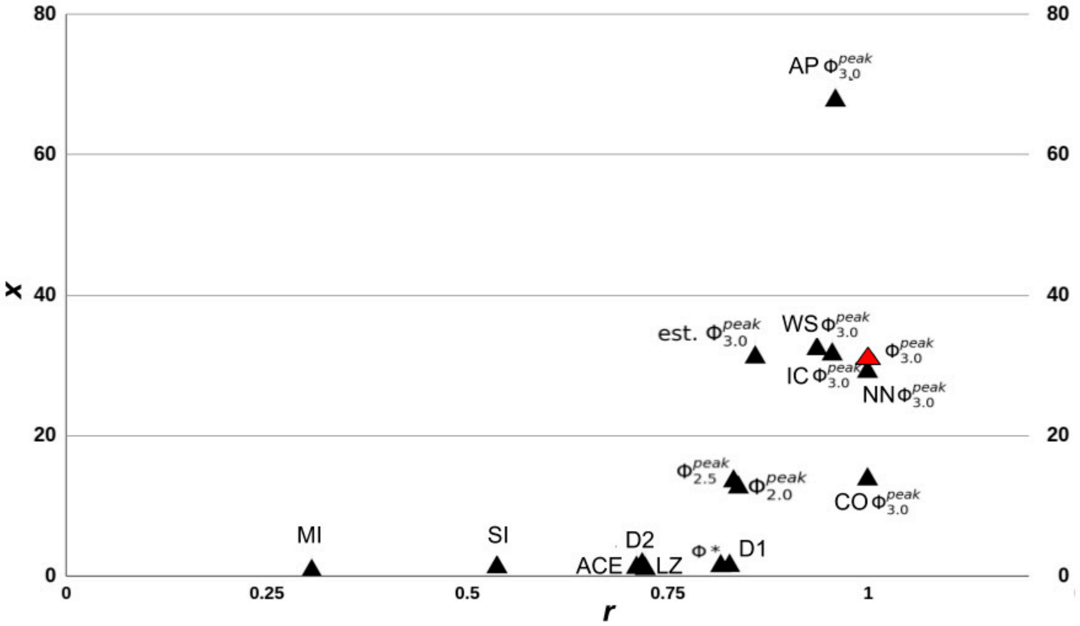


Figure A3. Overview of computational times as recorded locally, for each measure ($\Phi^{peak}_{3.0}$marked red), over correlation (r/r_s) with $\Phi^{peak}_{3.0}$. The y-axis corresponds to the exponent x of the logarithmic fit between times (in seconds) for networks of size n=3,4,5,6, in the form of y=bn^x, where b is a constant. Φ: integrated information; Φ^{peak}: maximum Φ over system states; CO: cut one approximation; NN: no new concepts approximation; AP: all partitions; WS; whole system approximation; IC: iterative cut approximation; est.$\Phi^{peak}_{3.0}$: $\Phi^{peak}_{3.0}$ estimated from 5 sample states; D1/2: state differentiation measure 1/2; *S*: state entropy; LZ: Lempel Ziv complexity; SI: stochastic interaction; MI: mutual information.

References

1. Crick, F.; Koch, C. Towards a neurobiological theory of consciousness. *Semin. Neurosci.* **1990**, *2*, 263–275.
2. Chalmers, D.J. Facing up to the problem of consciousness. *Journal of consciousness studies* **1995**, *2*, 200–219.
3. Balduzzi, D.; Tononi, G. Integrated information in discrete dynamical systems: motivation and theoretical framework. *PLoS Comput. Biol.* **2008**, *4*, e1000091.
4. Tononi, G. An information integration theory of consciousness. *BMC Neurosci.* **2004**, *5*, 42.
5. Oizumi, M.; Albantakis, L.; Tononi, G. From the phenomenology to the mechanisms of consciousness: integrated information theory 3.0. *PLoS Comput. Biol.* **2014**, *10*, e1003588.
6. Mayner, W.G.P.; Marshall, W.; Albantakis, L.; Findlay, G.; Marchman, R.; Tononi, G. PyPhi: A toolbox for integrated information theory. *PLoS Comput. Biol.* **2018**, *14*, e1006343.
7. Marshall, W.; Albantakis, L.; Tononi, G. Black-boxing and cause-effect power. *PLoS Comput. Biol.* **2018**, *14*, e1006114.

8. Marshall, W.; Kim, H.; Walker, S.I.; Tononi, G.; Albantakis, L. How causal analysis can reveal autonomy in models of biological systems. *Philos. Trans. A Math. Phys. Eng. Sci.* **2017**, *375*, 20160358.
9. Albantakis, L.; Tononi, G. The Intrinsic Cause-Effect Power of Discrete Dynamical Systems—From Elementary Cellular Automata to Adapting Animats. *Entropy* **2015**, *17*, 5472–5502.
10. Oizumi, M.; Amari, S.-I.; Yanagawa, T.; Fujii, N.; Tsuchiya, N. Measuring Integrated Information from the Decoding Perspective. *PLoS Comput. Biol.* **2016**, *12*, e1004654.
11. Barrett, A.B.; Seth, A.K. Practical measures of integrated information for time-series data. *PLoS Comput. Biol.* **2011**, *7*, e1001052.
12. Tegmark, M. Improved Measures of Integrated Information. *PLoS Comput. Biol.* **2016**, *12*, e1005123.
13. Schartner, M.; Seth, A.; Noirhomme, Q.; Boly, M.; Bruno, M.-A.; Laureys, S.; Barrett, A. Complexity of Multi-Dimensional Spontaneous EEG Decreases during Propofol Induced General Anaesthesia. *PLoS One* **2015**, *10*, e0133532.
14. Casali, A.G.; Gosseries, O.; Rosanova, M.; Boly, M.; Sarasso, S.; Casali, K.R.; Casarotto, S.; Bruno, M.-A.; Laureys, S.; Tononi, G.; et al. A theoretically based index of consciousness independent of sensory processing and behavior. *Sci. Transl. Med.* **2013**, *5*, 198ra105.
15. Marshall, W.; Gomez-Ramirez, J.; Tononi, G. Integrated Information and State Differentiation. *Front. Psychol.* **2016**, *7*, 926.
16. Haun, A.M.; Oizumi, M.; Kovach, C.K.; Kawasaki, H.; Oya, H.; Howard, M.A.; Adolphs, R.; Tsuchiya, N. Conscious Perception as Integrated Information Patterns in Human Electrocorticography. *eNeuro* **2017**, *4*.
17. Kim, H.; Hudetz, A.G.; Lee, J.; Mashour, G.A.; Lee, U.; ReCCognition Study Group Estimating the Integrated Information Measure Phi from High-Density Electroencephalography during States of Consciousness in Humans. *Front. Hum. Neurosci.* **2018**, *12*, 42.
18. Hudetz, A.G.; Liu, X.; Pillay, S. Dynamic repertoire of intrinsic brain states is reduced in propofol-induced unconsciousness. *Brain Connect.* **2015**, *5*, 10–22.
19. Mediano, P.A.M.; Seth, A.K.; Barrett, A.B. Measuring Integrated Information: Comparison of Candidate Measures in Theory and Simulation. *Entropy* **2018**, *21*, 17.
20. Kanwal, M.; Grochow, J.; Ay, N. Comparing information-theoretic measures of complexity in Boltzmann machines. *Entropy* **2017**.
21. Oizumi, M.; Tsuchiya, N.; Amari, S.-I. Unified framework for information integration based on information geometry. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 14817–14822.
22. Ferenets, R.; Lipping, T.; Anier, A.; Jäntti, V.; Melto, S.; Hovilehto, S. Comparison of entropy and complexity measures for the assessment of depth of sedation. *IEEE Trans. Biomed. Eng.* **2006**, *53*, 1067–1077.
23. Gosseries, O.; Schnakers, C.; Ledoux, D.; Vanhaudenhuyse, A.; Bruno, M.-A.; Demertzi, A.; Noirhomme, Q.; Lehembre, R.; Damas, P.; Goldman, S.; et al. Automated EEG entropy measurements in coma, vegetative state/unresponsive wakefulness syndrome and minimally conscious state. *Funct. Neurol.* **2011**, *26*, 25–30.
24. Schartner, M.M.; Carhart-Harris, R.L.; Barrett, A.B.; Seth, A.K.; Muthukumaraswamy, S.D. Increased spontaneous MEG signal diversity for psychoactive doses of ketamine, LSD and psilocybin. *Sci. Rep.* **2017**, *7*, 46421.
25. Amari, S.; Tsuchiya, N.; Oizumi, M. Geometry of information integration. *arXiv preprint arXiv:1709.02050* **2017**.

26. Kitazono, J.; Oizumi, M. *Practical PHI toolbox for integrated information analysis*; 2018;.
27. Hudetz, A.G.; Mashour, G.A. Disconnecting Consciousness: Is There a Common Anesthetic End Point? *Anesth. Analg.* **2016**, *123*, 1228–1240.
28. Lempel, A.; Ziv, J. On the Complexity of Finite Sequences. *IEEE Trans. Inf. Theory* **1976**, *22*, 75–81.
29. Schartner, M.M.; Pigorini, A.; Gibbs, S.A.; Arnulfo, G.; Sarasso, S.; Barnett, L.; Nobili, L.; Massimini, M.; Seth, A.K.; Barrett, A.B. Global and local complexity of intracranial EEG decreases during NREM sleep. *Neurosci Conscious* **2017**, *2017*, niw022.
30. Kitazono, J.; Kanai, R.; Oizumi, M. Efficient Algorithms for Searching the Minimum Information Partition in Integrated Information Theory. *Entropy* **2018**.
31. Hidaka, S.; Oizumi, M. Fast and exact search for the partition with minimal information loss. *PLoS One* **2018**, *13*, e0201126.
32. Arsiwalla, X.D.; Verschure, P.F.M.J. Integrated information for large complex networks. In Proceedings of the The 2013 International Joint Conference on Neural Networks (IJCNN); 2013; pp. 1–7.
33. Boly, M.; Massimini, M.; Tsuchiya, N.; Postle, B.R.; Koch, C.; Tononi, G. Are the Neural Correlates of Consciousness in the Front or in the Back of the Cerebral Cortex? Clinical and Neuroimaging Evidence. *J. Neurosci.* **2017**, *37*, 9603–9613.
34. Albantakis, L.; Hintze, A.; Koch, C.; Adami, C.; Tononi, G. Evolution of integrated causal structures in animats exposed to environments of increasing complexity. *PLoS Comput. Biol.* **2014**, *10*, e1003966.
35. Toker, D.; Sommer, F. Moving Past the Minimum Information Partition: How To Quickly and Accurately Calculate Integrated Information. *arXiv [q-bio.NC]* 2016.
36. Khajehabdollahi, S.; Abeyasinghe, P.; Owen, A.; Soddu, A. The emergence of integrated information, complexity, and consciousness at criticality. *bioRxiv* 2019, 521567.
37. Esteban, F.J.; Galadí, J.; Langa, J.A.; Portillo, J.R.; Soler-Toscano, F. Informational structures: A dynamical system approach for integrated information. *PLoS Comput. Biol.* **2018**, *14*, e1006154.
38. Virmani, M.; Nagaraj, N. A novel perturbation based compression complexity measure for networks. *Heliyon* **2019**, *5*, e01181.
39. Toker, D.; Sommer, F.T. Information integration in large brain networks. *PLoS Comput. Biol.* **2019**, *15*, e1006807.
40. Mori, H.; Oizumi, M. Information integration in a globally coupled chaotic system. *The 2018 Conference on Artificial Life: A Hybrid of the European Conference on Artificial Life (ECAL) and the International Conference on the Synthesis and Simulation of Living Systems (ALIFE)* **2018**, 384–385.
41. Isler, J.R.; Stark, R.I.; Grieve, P.G.; Welch, M.G.; Myers, M.M. Integrated information in the EEG of preterm infants increases with family nurture intervention, age, and conscious state. *PLoS One* **2018**, *13*, e0206237.