A framework to study dynamic dependencies in networks of interacting processes

Daniel Chicharro[1,*] and Anders Ledberg[2,3]

[1]*Center for Neuroscience and Cognitive Systems@UniTn, Istituto Italiano di Tecnologia, Via Bettini 31, 38068 Rovereto (TN)*

[2]*Center of Brain and Cognition, Universitat Pompeu Fabra, Barcelona, Spain*

[3]*SoRAD, Stockholms Universitet, Stockholm, Sweden*

(Dated: September 14, 2012)

The analysis of dynamic dependencies in complex systems such as the brain helps to understand how emerging properties arise from interactions. Here we propose an information theoretic framework to analyze the dynamic dependencies in multivariate time-evolving systems. This framework constitutes a fully multivariate extension and unification of previous approaches based on bivariate or conditional mutual information and Granger causality or transfer entropy. We define multi-information measures that allow us to study the global statistical structure of the system as a whole, the total dependence between subsystems, and the temporal statistical structure of each subsystem. We develop a stationary and a non-stationary formulation of the framework.

We then examine different decompositions of these multi-information measures. The transfer entropy naturally appears as a term in some of these decompositions. This allows us to examine its properties not as an isolated measure of interdependence but in the context of the complete framework. More generally we use causal graphs to study the specificity and sensitivity of all the measures appearing in these decompositions to different sources of statistical dependence arising from the causal connections between the subsystems. We illustrate that there is no straightforward relation between the strength of specific connections and specific terms in the decompositions. Furthermore, causal and noncausal statistical dependencies are not separable. In particular, the transfer entropy can be nonmonotonic in dependence on the connectivity strength between subsystems and is also sensitive to internal changes of the subsystems, so that it should not be interpreted as a measure of connectivity strength. Altogether, in comparison to an analysis based on single isolated measures of interdependence, this framework is more powerful to analyze emergent properties in multivariate systems, and to characterize functionally relevant changes in the dynamics.

PACS numbers: 87.19.lo, 89.70.Cf, 89.75-k, 87.19.lj, 05.45.Tp, 87.10.Mn, 87.18-h

I. INTRODUCTION

The dynamics of complex systems are often modeled as resulting from interactions between less complex subsystems, or parts. The aim is of course to describe how complex “emergent” properties arise from interactions between simpler and perhaps experimentally and conceptually more accessible parts. The success of this approach is epitomized by the kinetic theory of gases, but many, if not most, complex phenomena are investigated using the same reductionist methodology.

If enough is known about the subsystems and the way in which they interact, a fruitful approach is to construct mechanistic 'generative' models and compare the dynamics of such models to experimental data. If less is known, a data-driven approach is often needed where properties of the subsystems and their interactions are estimated from data. The latter is the situation we will be concerned with here.

Measurements of real world systems are often noisy, perhaps due to intrinsic randomness of the systems or perturbations introduced in the measurement process. Consequently, all information about the system that can be obtained from observational studies is a (potentially time-evolving) probability distribution over the observables. The all-important problem then becomes how to represent (parametrize, characterize) this, potentially high-dimensional, probability distribution so that useful information about the subsystems and their interactions are obtained.

Here we introduce a general framework to quantify different types of statistical dynamic dependencies within a system. This framework unifies previous approaches and helps to put some measures of interdependence into proper context. We will be concerned with systems that change in time and we will discuss in some detail how the predictability of the future can be decomposed into different terms. These decompositions will be related to measures of Granger causality [1, 2] and transfer entropy [3]. We will see that these measures should not be considered stand-alone measures of interdependence, but rather as components in specific decompositions of the total dependence between parts of a multivariate temporally-evolving system.

Granger causality/transfer entropy measures have acquired a preeminent status in the study of interactions between time-evolving processes. The linear measure of Granger causality [1, 2, 4] implements a criterion for uncovering causality from improved predictability [5]. The transfer entropy [3] implements in the information-theoretic framework a generalization of this criterion from a formulation in terms of predictability to a formulation in terms of conditional independence. This generalization was already described by Granger in [2], and different formulations and measures equivalent to transfer entropy have been previously introduced in different field [e. g. 6, 7], and see [8] for a review. We here will refer by Granger causality to the general criterion proposed in [2], so that Granger causality measures include both the transfer en-

*Electronic address: daniel.chicharro@iit.it

tropy and the linear Granger causality measures (see Methods for a more detail discussion of their relation).

Granger causality measures have been applied in many areas, comprising financial time series [9], mechanical systems [10], weather prediction [11], and chemical processes [12]. In particular they have especially been used to study electrophysiological and neural signals [see 8, 13–15, for review]. Studies comprise different recording modalities such as single neuron activity [16], electroencephalogram (EEG) signals [17], magnetoencephalogram signals [18], local field potentials (LFP) [19], or functional magnetic resonance imaging (fMRI) [20]. Given this profusion of applications to neural data we will focus on the field of neuroscience to specifically discuss the significance of our results and the potential usefulness of the framework we propose. Nonetheless our framework is general in the sense that it does not rely on the nature of the data analyzed, and is thus valid for any system consisting of interacting parts.

Despite their wide use, the interpretation of Granger measures remains controversial. Granger [4] referred to the linear Granger causality measure as a measure of *causality strength*. Originally, transfer entropy was considered a measure of *information flow* [6], *causality* [7, 21], or *information transfer* [3]. In experimental studies the emphasis on these different notions varies, and it is common that different terms as *information flow*, *causal influence*, or *directed dependence* are used in the same study [e. g. 22], considering the main characteristic of the Granger causality measures being directed measures of dynamic interdependence [15]. However, how the measures are interpreted has significant impact on the potential information provided about how the system actually works. In the specific context of brain connectivity analysis, it has been debated how useful they are for causal inference and if they are measures of 'effective' or 'functional connectivity' [8, 23, 24].

In a recent contribution [25] we separated different questions about causality and used this distinction to better interpret the measures mentioned above. In particular, we showed that to infer the existence of causal connections and to quantify the effects of these connections are tasks that need different methods with different requirements. In general, causal inference can only be attained by intervening, externally perturbing, the system [26]. Only under strongly restrictive assumptions (that we will refer below as *complete observability*) Granger causality measures can be used for causal inference. Since they rely on statistical dependencies, ultimately they are limited by the existence of observationally equivalent causal structures [26].

Regarding the quantification of causal effects, we demonstrated in [25] that it is often not meaningful to characterize interactions between different subsystems in terms of causes and effects. This is because in many cases the main interest is not in the effect of external perturbations, but in how the causal connections participate in the generation of the unperturbed dynamics of the system. We introduced the notion of *natural causal effects between dynamics* [25] and provided conditions for their existence. We concluded that the effect of causal connections can typically not be described in terms of the effect of one subsystem over another. Consequently, we argued that Granger causality measures, and in particular transfer entropy, cannot be used in general as measures of the strength of causal effects. We advocated a change of focus: instead of examining the causal effects resulting from the causal connections one should analyze how causal connections participate in the generation of functionally relevant dynamics [cf. 27]. In other words, quantifying causal effects between subsystems is often not meaningful, however, the dynamic dependencies between the subsystems are both meaningful and characterizable. Accordingly, in the present contribution we focus on the study of statistical dependencies. We will consider how different types of statistical dependencies arise from the causal connections in the system, but we are not interested in the quantification of causal effects as we were in [25]. We provide a unifying framework to characterize dynamic dependencies in multivariate time-evolving systems. In contrast to the majority of studies in which Granger measures are applied to study interactions between processes, we here do not consider these measures as isolated measures.

The role of Granger causality measures in a decomposition of different contributions to the mutual information between bivariate processes has been already examined [21, 28–31], mainly theoretically, both for stationary and non-stationary processes. This decomposition has also been extended to the conditional mutual information rate [32] in multivariate processes, but without addressing its applicability for non-stationary processes. Bivariate and conditional Granger causality measures were further developed in [33] to address the analysis of multivariate variables. Following this line, we here address the connection of transfer entropy with a framework more generally quantifying the statistical structure of the subsystems forming a time-evolving multivariate system. This statistical structure can be relevant to understand how the system work. For example, in neuroscience, the coexistence of some degree of *segregation* and *integration* between the dynamics of different brain regions has been suggested as an explanation of the specialization of different regions in some modalities of sensory processing and the emergence of multisensory representations and higher cognitive processes [34].

Studying the statistical structure of the system implies considering multivariate measures of dependence, such as the multi-information [34–36]. Furthermore, in a framework to study this statistical structure one needs to be able to consider different assumptions about the parts in which the system is spatially and temporally decomposed. We will see that using multi-information measures it is natural to define measures quantifying the *global* statistical structure of the system, which considers the multivariate process as a unique system, the *subsystems* statistical structure, which considers the dependence between different subsystems which are *a priori* distinguished, and the *temporal* statistical structure of each separate process. Examining the role of transfer entropy in such a framework can help to better evaluate its common application as a measure of interdependence. More generally, the whole framework is potentially more powerful to analyze emergent properties of the system, and to characterize changes in the dynamics that are functionally relevant.

We start by revising the basic information theoretic measures of dependence needed for our framework (Sec. II A). We also revise the causal graph representation of causal connections between processes and how the causal structure is related to the different sources of statistical dependence between these processes (Sec. II B). We then propose the quantities needed to build a framework which allows us to analyze the statistical structure of a multivariate system (Sec. III A). We do this first for the specific case of stationary processes, which simplifies the formulation and allows us to provide a complete perspective of the framework without addressing its whole derivation. For example, we show that the extension to multivariate systems indicates that the decomposition for the bivariate case of the mutual information into transfer entropies can neither be understood as a separation of different sources of statistical dependence nor as a preeminent decomposition different than other alternative ones.

We then focus on a particular measure appearing in the framework, namely the mutual information between the future of a single subsystem and the past of the system (Sec. III B). Examining alternative decompositions of this measure allows us to illustrate the sensitivity and specificity of transfer entropy and other measures appearing in the framework to different types of statistical dependencies. We use causal graphs [26] to illustrate the sensitivity to, on one side both causal and noncausal sources of statistical dependence, and on the other side to the internal properties of the processes apart from to the interactions between the processes. Notice that, although we illustrate that these measures are inappropriate to quantify the strength of causal connections, as it is well-known [e. g. 37, 38], we here do not address the quantification of causal effects [25].

A simple graphical analysis of the dependencies [26] suffices to examine most of the issues and to argue in favor of using the whole framework instead of just the transfer entropy to characterize the dynamic dependencies. The transfer entropy naturally appears in some of the possible decompositions of measures quantifying the statistical structure of the system and, as the other measures appearing in these decompositions, it is difficult to interpret in isolation because changes in its value may be originated by internal changes of each subsystem as well as by global changes in a way which is not separable. We use simple analytically solvable processes to illustrate this III C.

In the second part of our results (Sec. III D) we develop in detail the framework for the analysis of dynamic dependencies in a system of interacting processes. We show how the measures characterizing the statistical structure of the system can be expressed in terms of transfer entropy terms, providing a temporally causal decomposition alternative to a purely hierarchical decomposition of the dependencies [34]. We distinguish between formulation of the measures which is local in time and another that is cumulative on the whole time series, and we consider the correspondence between the nonstationary and stationary formulation. We also address how the bivariate and conditional measures are related.

II. METHODS

A. Information-theoretic measures of dependence

The basic information theory measure to characterize the uncertainty of a (possibly multivariate) random variable is the entropy [39]

$$H(X) = -\sum_x p(x) \log p(x), \tag{1}$$

where $p(x)$ is the probability distribution of the variable. Furthermore, the basic measure in information theory to quantify the difference between two probability distributions of a (possibly multivariate) random variable X is the Kullback-Leibler divergence (KL divergence) [40]. The KL divergence is defined as [39]

$$\mathrm{KL}(p^*(x), p(x)) = \sum_x p^*(x) \log \frac{p^*(x)}{p(x)}. \tag{2}$$

An important characteristic of the KL divergence is that it is a non-negative number and is zero if and only if the two distributions are identical. For a multivariate variable X, since it quantifies the divergence of the distribution $p(x)$ from $p^*(x)$ [see also 39, for a more specific interpretations in terms of code length], one can construct $p(x)$ to reflect a specific null-hypothesis about the dependence between the components of X. As particular applications of the KL divergence to quantify the interdependence between random variables one has the mutual information

$$I(X;Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}, \tag{3}$$

the conditional mutual information

$$I(X;Y|Z) = \sum_{x,y,z} p(x,y,z) \log \frac{p(x,y|z)}{p(x|z)p(y|z)}, \tag{4}$$

and the multi-information [34–36]

$$M(X;Y;Z) = \sum_{x,y,z} p(x,y,z) \log \frac{p(x,y,z)}{p(x)p(y)p(z)}. \tag{5}$$

The mutual information quantifies the dependence between two sets of variables, and the conditional mutual information the conditional dependence given a third set of variables. The multi-information quantifies the total dependence between several sets of variables, not necessarily a dependence that is common to all the sets but any dependence between any of the sets.

Here we will use these measures to study dynamic dependencies between random variables corresponding to the state of stochastic processes at a given sampling time. Consider a set of n stochastic processes $\mathcal{V}_j$, $j = 1, ..., n$. Let $V_j^N = \{V_{j1}, V_{j2}, \ldots, V_{jN}\}$ be the random variables corresponding to sampling times $1, 2, \ldots, N$ of process j. For a particular sampling time i, $V_j^i = \{V_{j1}, V_{j2}, \ldots, V_{ji}\}$. We want to characterize both the internal properties of the processes and the interdependence between these processes.

For a process $\mathcal{X}$ the entropy at sampling time i is

$$H(X_i) = -\sum_x p_{X_i}(x) \log p_{X_i}(x), \tag{6}$$

and characterizes the uncertainty of the random variable X_i. Here p_{X_i} denotes the probability distribution of the variable that appears as a sub-index. In the following we will suppress these sub-indices and alternatively we will use $p(x_i)$ to refer to the probability distribution of X_i. Since we are dealing with stochastic processes it will be relevant to consider conditional entropies such as

$$H(X_{i+1}|X^i) = -\sum_{x_{i+1},x^i} p(x_{i+1}, x^i) \log p(x_{i+1}|x^i). \tag{7}$$

This can be thought of as the average uncertainty of X_{i+1} given that we know the values in the past of $\mathcal{X}$. In general, for non-stationary processes, expressions such as Eq. 7 will depend on the temporal index i. This type of conditional entropies conditioning on the past for a given time i naturally appear if one considers the entropy of the whole time series

$$H(X^N) = -\sum_{x^N} p(x^N) \log p(x^N) = \sum_{k=0}^{N-1} H(X_{k+1}|X^k), \tag{8}$$

where the decomposition into the summands is based on the chain rule for the entropy [39]. This decomposition links local conditional entropies at each time k to a cumulative joint entropy for the whole time series. This will allow us in Sec. III D to differentiate between two definitions of the measures in the framework we propose. One is local for a particular time and the other is a cumulative definition that considers the whole time series.

For stochastic processes, instead of the cumulative joint entropy $H(X^N)$, which depends on N, one can alternatively consider the entropy rate

$$\mathcal{H}(\mathcal{X}) = \lim_{N\to\infty} \frac{H(X^N)}{N} \tag{9}$$

that characterizes the average entropy per sampling time, when the limit exists. The entropy rate can alternatively be defined as

$$\mathcal{H}(X_{i+1}|X^i) = \lim_{k\to\infty} H(X_{k+1}|X^k) \tag{10}$$

that characterizes the rate of increase of the entropy of the time series. Notice that while in $H(X_{k+1}|X^k)$ the index k indicates a particular time which limit to infinite is taken, in $\mathcal{H}(X_{i+1}|X^i)$ the index i is used as a dummy index just to indicate the relative temporal relation between the conditioning and conditioned variables. It can be proven [39] that these two entropy rates exist and are equivalent for discrete valued stationary processes, although for more general processes the existence of the rates is not assured. This equivalence will allow us to show in Sec. III D that the local and cumulative formulation of the framework are equivalent for these stationary processes. Accordingly, for simplification, we will for the moment consider stationary processes so that the definitions are independent of the index i.

Next we will describe transfer entropy and show how it relates to mutual information in the stationary case. Like in Eqs. 9 and 10 we will use script typeface to differentiate measures that are defined as rates. The mutual information rate is defined as

$$\mathcal{I}(\mathcal{X};\mathcal{Y}) = \lim_{N\to\infty} \frac{I(X^N;Y^N)}{N}. \tag{11}$$

The KL divergence also provides a measure to generally test the Granger causality criterion [1, 2]. In the bivariate case, this criterion says that $\mathcal{Y}$ is causing $\mathcal{X}$ if

$$p(\text{future of } \mathcal{X}|\text{past of } \mathcal{X} \text{ and } \mathcal{Y}) \neq p(\text{future of } \mathcal{X}|\text{past of } \mathcal{X}). \tag{12}$$

In words, $\mathcal{Y}$ is causing $\mathcal{X}$ if the future of $\mathcal{X}$ given the past of $\mathcal{X}$ is not independent of the past of $\mathcal{Y}$. Strictly speaking, this criterion depends on that we know that there are no other causes of $\mathcal{X}$ than $\mathcal{X}$ and (potentially) $\mathcal{Y}$. We will discuss its limitations in the discussion. In the original formulation, Granger used the mean of the distributions to test the above inequality, leading to the linear Granger causality measures. However, it was clear to him that other measures could be used as well [see 2]. Other nonlinear measures have been proposed based on the same criterion [e. g. 21, 41–43]. In the context of information theory, a more general approach for testing Eq. 12 was independently developed at about the same time [6]. This approach consists in directly testing the equality of the two probability distributions in Eq. 12 using the KL divergence. For stationary processes, the transfer entropy rate is

$$\mathcal{T}_{\mathcal{Y}\to\mathcal{X}} = \lim_{k\to\infty} \sum_{x^k y^k} \sum_{x_{k+1}} p(x_{k+1}, x^k y^k) \log \frac{p(x_{k+1}|x^k, y^k)}{p(x_{k+1}|x^k)}. \tag{13}$$

In the original work [6] the KL divergence was suggested as a measure of information gain for systems with feedback. In the same line it has been proved to have a rigorous interpretation as the error-free information rate in information channels with feedback [44], and has also applications in other areas like gambling theory [45]. However, the same measure was reintroduced as measure of causal dependence [7, 21, 28]. Most recently, the same measure has been re-re-introduced under the name 'transfer entropy' [3], as a measure of 'information transfer'. Due to its recent popularity, we will use this name in the sequel. See [e. g. 8, 31] for a more detailed discussion of how different formulations of Granger causality appeared in different fields and for different types of processes. The transfer entropy coincides with the linear Granger causality measure of Geweke [28] for linear Gaussian stationary processes [46], but in general they differ since the first one test for the general criterion in Eq. 12 and the second only for the equality of the means using a minimum squared error predictor.

It is straightforward to show that

$$\mathcal{T}_{\mathcal{Y}\to\mathcal{X}} = \mathcal{H}(X_{i+1}|X^i) - \mathcal{H}(X_{i+1}|X^i, Y^i) = \mathcal{I}(X_{i+1}; Y^i|X^i). \tag{14}$$

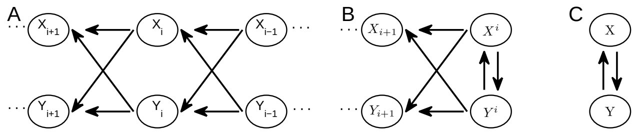

Figure 1: Graphical representation of causal connections. Causal graphs representing a generic bivariate process of order 1 with bidirectional causal connections between $\mathcal{Y}$ and $\mathcal{X}$ (A-C). From left to right the scale of the graphs changes from a microscopic level representing the dynamic interactions to a macroscopic one in which each process is represented by a single node.

That is, the transfer entropy rate is a conditional mutual information rate . Accordingly, it is zero if and only if

$$p(X_{i+1}|X^i) = p(X_{i+1}|X^i, Y^i), \tag{15}$$

which formalizes the criterion of Eq. 12. This means that the transfer entropy rate is a general measure of the criterion of Granger causality under the stationary assumption.

For stationary processes Marko [6] showed that the mutual information and transfer entropy rates are related according to

$$\mathcal{I}(\mathcal{X};\mathcal{Y}) = \mathcal{T}_{\mathcal{Y}\to\mathcal{X}} + \mathcal{T}_{\mathcal{X}\to\mathcal{Y}} + \mathcal{T}_{\mathcal{X}\cdot\mathcal{Y}}, \tag{16}$$

where $\mathcal{T}_{\mathcal{Y}\cdot\mathcal{X}}$ is a rate of instantaneous mutual information defined as

$$\mathcal{T}_{\mathcal{X}\cdot\mathcal{Y}} = \lim_{k\to\infty} \sum_{x^{k+1}y^{k+1}} p(x^{k+1}, y^{k+1}) \log \frac{p(x_{k+1}, y_{k+1}|x^k, y^k)}{p(x_{k+1}|x^k, y^k)p(y_{k+1}|x^k, y^k)}. \tag{17}$$

Eq. 16 is a general nonparametric formulation of the relation presented by Geweke in [28] for linear stationary processes. This relation for linear stationary processes was extended to the conditional case [32] and is also generalized in an information-theoretic formulation for conditional mutual information and entropy rates as

$$\mathcal{I}(\mathcal{X};\mathcal{Y}|\mathcal{Z}) = \mathcal{T}_{\mathcal{Y}\to\mathcal{X}|\mathcal{Z}} + \mathcal{T}_{\mathcal{X}\to\mathcal{Y}|\mathcal{Z}} + \mathcal{T}_{\mathcal{X}\cdot\mathcal{Y}|\mathcal{Z}}. \tag{18}$$

The definition of these conditional rates will be discussed in Sec. III D.

B. Causal graphs and the different sources of statistical dependence

Causal graphs provide an accessible framework for studying how causal connections introduce statistical dependencies [26]. The idea is simple, variables or sets of variables are nodes in the graph and causal connections between the variables are represented by directed edges (arrows). Statistical dependencies resulting from these causal connections can then be deduced by simple graph-theoretical methods. We now review the representation of stochastic processes in a causal graph at different scales, and how the causal connections produce causal and noncausal statistical dependencies between the nodes.

Consider a bivariate Markov process of order 1. In Fig. 1A-C we show three different causal graphs representing the process at different scales when bidirectional causal connections between $\mathcal{Y}$ and $\mathcal{X}$ exist. In Fig. 1A there is a complete explicit representation of the causal structure of the process. The three dots at left and right indicate that only a particular time window of the process is represented. In such causal graph each node is associated with a stochastic variable V_k, where V stands for X or Y and k for a particular time sampled. The arrows represent direct causal connections. We will call this type of causal graph the *microscopic* causal graph of the process.

By contrast, in Fig. 1B not all the nodes of the graph correspond to a unique time sampling. The rightmost nodes represent multivariate variables associated with the whole past of the processes, i.e. with X^i and Y^i. The temporal dynamics are only partially represented, by distinguishing time $i+1$ from its past. We will call this type of causal graph the *mesoscopic* causal graph of the process. Finally in Fig. 1C each node corresponds to a whole process, so that the temporal dynamics are completely implicit. This is the *macroscopic* causal graph of the bivariate process.

Macroscopic causal graphs are the most usually used [e. g. 47–49] because they have the advantage that the causal connections between two processes are represented by only two potential single directed edges. Since the macroscopic representation has fewer nodes it is easier to visualize. On the other hand it does not show the details of the causal connections, since it suffices that a directed edge $Y_k \to X_{k'}$ exists for some particular $k < k'$ in the microscopic graph, for a directed edge $Y \to X$ to appears in the macroscopic graph. Furthermore the causal connections like $X_k \to X_{k'}$ are not represented. This also occurs at the mesoscopic level (Fig. 1B) for the nodes associated with the multivariate variables X^i and Y^i. Not considering all the temporal dynamics also can lead to the existence of loops in the graph. For example, in Fig. 1B and C, the bidirectional causal connections result in directed edges pointing in both directions between X^i, Y^i and X, Y, respectively. Oppositely at the microscopic level the causal graph is always a directed acyclic graph (DAG), as long as the causal connections are assumed to be noninstantaneous.

Notice that a causal graph is only informative about the existence of causal connections, not about the causal interactions resulting from them. These causal interactions are determined by the particular physical implementation of the connections, and are reflected in a particular generative model of the processes. On the microscopic scale, where an arrow represents a single causal connection, a connectivity strength can be associated with each arrow in a given model. However, this connectivity strength should be distinguished from a quantification of the causal effect of the causal connections from one subsystem to another [see 25, for a detailed discussion].

Here we are not interested in the inference of the causal structure or the quantification of causal effects, and we use the causal graphs because they allow us to examine the existence

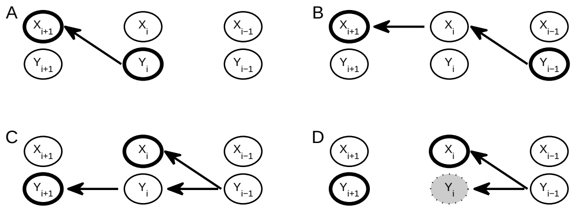

Figure 2: The relation between causal connections and the sources of statistical dependence. A-C: Different types of paths leading to different sources of statistical dependence. In bold we indicate the nodes whose statistical dependence is associated with a path. A: Directed edge that leads to a direct causal statistical dependence. B: Directed path that leads to an indirect causal statistical dependence. C: Nondirected path that leads to a noncausal statistical dependence. D: Graphical representation of the effect of conditioning on the statistical dependencies. The dotted node indicates the conditioned variable. The arrows leaving from it are removed.

of statistical dependencies between nodes (i.e. variables or sets of variables) by graph criteria, as outlined below. A *path* in the graph between two nodes V_k, $V_{k'}$, is a sequence of pairs of nodes linked by an arrow such that the first pair contains V_k, the last $V_{k'}$, and consecutive pairs have a node in common. In a path, it is allowed to go across an arrow in any direction, not only in the direction in which the arrow is pointing. A path where the directions of the edges are respected is called a *directed path*, while a path that is not a directed path is called a *nondirected* path.

From the graph criterion *d-separation* [26] the existence of statistical dependencies between two nodes can easily be determined: Two nodes are statistically dependent if there exists a path between them such that it does not contain any collider. (A collider is a combination of three nodes connected by arrows forming an inverted fork $V_k \rightarrow V_{k'} \leftarrow V_{k''}$.)

We will distinguish between three types of statistical dependencies. First, if two nodes are linked by a directed edge we say that there is a direct causal statistical dependence between the nodes (Fig. 2A). Second, if a directed path containing more than one directed edge exists between the nodes, we say that there is an indirect causal statistical dependence between the nodes (Fig. 2B). To relate these types of dependencies to direct or indirect causal connections one has to examine the microscopic graph. Finally, if a nondirected path exists between two nodes, we say that there is a non-causal statistical dependence between the nodes (Fig. 2C). Note that this statistical dependence still results from causal connections present in the system, but does not correspond to a causal flow between these particular variables.

Mutual information is sensitive to both causal and noncausal statistical dependence. The conditional mutual information, of which the transfer entropy is a particular example, is also unspecific for causal or noncausal statistical dependencies, although the conditioning removes part of the statistical dependencies. Graphically, conditioning can be understood as the blocking of all the paths going through the conditioned variable. Consider for example the statistical dependence between X_i and Y_{i+1} and the path shown in Fig. 2C. Conditioning on Y_i implies removing all the edges leaving from Y_i, and thus this path is destroyed (Fig. 2D), showing that (in this example) X_i and Y_{i+1} are conditionally independent given Y_i.

III. RESULTS

A. Motivation and framework for stationary processes

In Sec. II A we revised the definition of transfer entropy rate and its relation to mutual information rate. Given Eq. 16 the mutual information rate of the processes $\mathcal{X}$ and $\mathcal{Y}$ can be expanded as the contribution of three sources of dependence: the transfer entropy rate in the two directions, and a source of instantaneous dependence. An appealing and usual interpretation of this decomposition [7, 29] is that the dependence between the processes results from the causal influence $\mathcal{X}$ has on $\mathcal{Y}$, the causal influence $\mathcal{Y}$ has on $\mathcal{X}$, and some instantaneous dependence which cannot be explained by causal interactions between them. This interpretation is appealing because it seems to provide a separation of the different sources that result in the dependence between the processes. Here we want to examine how fundamental this decomposition is, that is, if this decomposition is preeminent among other possible decompositions of the mutual information rate. For that purpose it is important to examine (Sec. III B) the sensitivity and specificity of the measures to the different sources of dependence we distinguished in Sec. II B.

Consider that the quantities in Eq. 16 are calculated for a truly bivariate system, and assume we have access to the dynamic processes between which the interactions really occur. In this case we will see that the three terms on the rhs can be related to the existence of particular arrows in a graph analogous to the ones of Fig. 1. A nonzero transfer entropy $T_{\mathcal{Y}\rightarrow\mathcal{X}}$ indicates the existence of causal connections from $\mathcal{Y}$ to $\mathcal{X}$, and analogously for $T_{\mathcal{X}\rightarrow\mathcal{Y}}$. A nonzero term of instantaneous causality $T_{\mathcal{X}\cdot\mathcal{Y}}$ would involve the existence of vertical arrows $X_i \rightarrow Y_i$ or $Y_i \rightarrow X_i$ in the causal graph, and this would be against the principle of causality. Accordingly, the instantaneous dependence is zero when calculated from the random variables of the truly bivariate system.

However, there are several reasons why the instantaneous dependence can be nonzero when estimated from data. A positive value of $T_{\mathcal{X}\cdot\mathcal{Y}}$ can result from the violations of the assumptions that allow to link the causal structure to the measures of dynamic dependence. For example, if the measures are calculated not from the processes between which the interactions occur, but from some vicarious signals derived from them [50, 51], or if the discrete approximation is not enough to avoid time aggregation [1, 8], or if there are hidden common drivers, then the instantaneous dependencies appear. In this case, when the bivariate decomposition is calculated not for an isolated bivariate system but for two subsystems of a larger system, it is important to understand how the sources of dependence that involve the rest of the system are reflected in each of the measures in the bivariate decomposition. That a nonzero instantaneous dependence appears is not the only effect from the rest of the system and to understand how the mutual information is then divided into the three terms of Eq. 16 it is necessary to examine the link between bivariate measures and conditional measures as the ones of Eq. 18. At first sight, in the stationary formulation in which they are expressed, the decomposition of the conditional mutual information rate in Eq. 18 seems to be the natural analogous extension of the decomposition of the mutual information rate of Eq. 16. However, we will see (Sec. III D) that for non-stationary processes these two decompositions have different degrees of generality.

More generally, decompositions such as the ones of Eqs. 16 and 18 do not account for all the dynamic dependencies in a system, but only allow us to study the dependencies between two subsystems. Here we propose a truly multivariate framework to analyze the whole system of interacting processes. We consider multi-information measures that quantify the dynamic dependencies in the system and we show different ways in which they can be decomposed into different contributions (Sec. III D).

Consider a multivariate system formed by interacting subelements resulting in some temporal joint dynamics. If one wants to study the dynamic dependencies present in the dynamics of the systems there are some assumptions that one can only do based on *a priori* information about the system. In particular, one needs to decide which are the relevant subsystems to be distinguished. In the brain, this division can be based on structural information, or on the function each part is expected to perform [see 52, and references therein]. If n subsystems are distinguished and given time series recorded from them (V_j^N, where $j = 1, ..., n$), one would like to characterize the statistical structure of the system in terms of the total dynamic dependence between these subsystems. The measure we propose can be seen as an extension of the measure of "integration" introduced by Tononi et al. [34], but in which the temporal dynamics are considered explicitly. The *subsystems multi-information* is defined as

$$M(V_1^N; V_2^N; ...; V_n^N) = \sum_{v_1^N, v_2^N, ..., v_n^N} p(v_1^N, v_2^N, ..., v_n^N) \log \frac{p(v_1^N, v_2^N, ..., v_n^N)}{\prod_{j=1}^n p(v_j^N)}. \quad (19)$$

Notice that here the assumption about the distinguishability of the different subsystems is reflected in the denominator of the logarithm, where there is a factorization of the joint probability distribution of the time series. Furthermore, remember that each variable v_j can be multivariate, so that several recordings from the same subsystem are included.

As the reverse of the selection of the subsystems which are distinguished, one decides at the same time which units are considered as a unique entity regarding the purpose of the analysis. For example, in the brain, given the ubiquity of reciprocal connections between areas, it is not immediately clear which parts should be considered as integrated or segregated during a particular brain computation. To understand this balance between functional integration and functional segregation [34], it can be useful to contrast an analysis in which different regions are assumed to be distinguishable with another in which the focus is on the joint dynamics arising from the interactions. In the latter case we propose to analyze the statistical structure of the system in terms of the dynamic dependencies quantified by the *global multi-information*

$$M(V_{11}; V_{12}; ...; V_{1N}; V_{21}; V_{22}; ...; V_{2N}; V_{n1}; V_{n2}; ...; V_{nN}) = \sum_{v_1^N, v_2^N, ..., v_n^N} p(v_1^N, v_2^N, ..., v_n^N) \log \frac{p(v_1^N, v_2^N, ..., v_n^N)}{\prod_{j=1}^n \prod_{i=1}^N p(v_{ji})}, \quad (20)$$

where V_{ji} refers to time i of the time series corresponding to process j. In the global multi-information the subsystems are not considered as separate entities, and this is reflected in the symmetry of the factorization in the denominator. An alternative way to divide the variables in the dynamics system could be to distinguish different times but not different subsystems. This would result in a measure of the temporal statistical structure of the whole system considered as a multivariate process. In the case of the brain, explicitly considering time adds to the analysis of functional integration and segregation the ingredient of temporal integration and segregation, which may reflect how the same brain area is involved in different functionalities during a sequential task. Here we will not use such a measure but we will consider a measure of the temporal internal statistical structure of each subsystem. For a subsystem $\mathcal{V}_j$, the *temporal internal multi-information* is

$$M(V_{j1}; V_{j2}; ...; V_{jN}) = \sum_{v_j^N} p(v_j^N) \log \frac{p(v_j^N)}{\prod_{k=1}^N p(v_{jk})}, \quad (21)$$

which can be seen as a particular global multi-information when calculated for a single process. We consider these measures as the basic ingredients of a framework to analyze dynamic dependencies in a system formed by interacting processes. In Sec. III D we will show how the global and subsystems multi-information can be decomposed into different components comprising transfer entropy and instantaneous dependence terms. Here we will summarize these results for the stationary case. Furthermore, we will motivate why we focus on the mutual information between the future of one subsystem and the past of the whole system to examine the sensitivity and specificity of the measures, and in particular the transfer entropy, to different sources of dependence.

We will show that the subsystems multi-information rate can be decomposed as

$$\mathcal{M}(\mathcal{V}_1; \mathcal{V}_2; ...; \mathcal{V}_n) = \mathcal{T}_{\mathcal{V}_1 \cdot \mathcal{V}_2 \cdot ... \cdot \mathcal{V}_n} + \sum_{j=1}^n \mathcal{T}_{\{\mathcal{V}\} \setminus \mathcal{V}_j \to \mathcal{V}_j}, \quad (22)$$

where $\{\mathcal{V}\} \setminus \mathcal{V}_j$ refers to all the processes except $\mathcal{V}_j$. The quantity $\mathcal{T}_{\mathcal{V}_1 \cdot \mathcal{V}_2 \cdot ... \cdot \mathcal{V}_n}$ is a measure of the instantaneous dependence between all the processes while $\mathcal{T}_{\{\mathcal{V}\} \setminus \mathcal{V}_j \to \mathcal{V}_j}$ is the transfer entropy to one process from all the rest of processes together. Since it is clear from Eq. 19 that the subsystems multi-information of two systems is the mutual information between them, this decomposition can be seen as the natural extension of the one in Eq. 16. However, this extension is contradictory to the appealing interpretation of the decomposition of the mutual information rate in terms of reciprocal causal influences. The decomposition of the subsystems multi-information is less intuitive because it is not symmetric, in the sense that it only considers contributions which are from all the rest to a particular process.

We will also show that the global multi-information rate can be decomposed as

$$\begin{aligned}\mathcal{M}(\mathcal{V}) &= \sum_{j=1}^{n}\mathcal{M}(\mathcal{V}_j) + \mathcal{M}(\mathcal{V}_1;\mathcal{V}_2;...;\mathcal{V}_n)\\ &= \sum_{j=1}^{n}\mathcal{I}(\mathcal{V}_{ji+1};\mathcal{V}_1^i,...,\mathcal{V}_n^i) + \mathcal{T}_{\mathcal{V}_1\cdot\mathcal{V}_2\cdot...\cdot\mathcal{V}_n}.\end{aligned} \tag{23}$$

Here $\mathcal{M}(\mathcal{V})$ and $\mathcal{M}(\mathcal{V}_j)$ refer to the global and temporal internal multi-information rates respectively. Therefore the first equality shows that, in an intuitive way, when analyzed with multi-information measures, the global statistical structure can be expanded in terms of the temporal internal statistical structure of each subsystems and the statistical structure of the subsystems (although this should not be understood as a separation of the sources of dependence, which is clear considering for example that the temporal statistical structure of a process depends also on the interactions with the others). In the second equality the mutual information rate $\mathcal{I}(\mathcal{V}_{ji+1};\mathcal{V}_1^i,...,\mathcal{V}_n^i)$ quantifies the dependence between process $\mathcal{V}_j$ at time $i+1$ and the past of the whole system. We will focus on this quantity to consider how statistical dependencies result from the underlying causal structure. This second decomposition is intuitive because it shows that the global multi-information, apart from a term of instantaneous dependence that should be zero under the assumptions discussed above, is just the sum of how the uncertainty of the evolution of each of the processes is reduced by the past of the whole system.

Although transfer entropies do not explicitly appear in the decompositions of Eq. 23, the terms $\mathcal{I}(\mathcal{V}_{ji+1};\mathcal{V}_1^i,...,\mathcal{V}_n^i)$ can themselves be decomposed as a sum of transfer entropies. For example, for a trivariate system

$$\begin{aligned}&\mathcal{I}(\mathcal{X}_{i+1};\mathcal{X}^i,\mathcal{Y}^i,\mathcal{Z}^i)\\ &= \mathcal{I}(\mathcal{X}_{i+1};\mathcal{X}^i) + \mathcal{T}_{\mathcal{Y}\to\mathcal{X}} + \mathcal{T}_{\mathcal{Z}\to\mathcal{X}|\mathcal{Y}}\\ &= \mathcal{I}(\mathcal{X}_{i+1};\mathcal{X}^i) + \mathcal{T}_{\mathcal{Z}\to\mathcal{X}} + \mathcal{T}_{\mathcal{Y}\to\mathcal{X}|\mathcal{Z}}\\ &= \mathcal{I}(\mathcal{X}_{i+1};\mathcal{Y}^i) + \mathcal{I}(\mathcal{X}_{i+1};\mathcal{X}^i|\mathcal{Y}^i) + \mathcal{T}_{\mathcal{Z}\to\mathcal{X}|\mathcal{Y}}\\ &= \mathcal{I}(\mathcal{X}_{i+1};\mathcal{Z}^i) + \mathcal{I}(\mathcal{X}_{i+1};\mathcal{Y}^i|\mathcal{Z}^i) + \mathcal{I}(\mathcal{X}_{i+1};\mathcal{X}^i|\mathcal{Y}^i,\mathcal{Z}^i).\end{aligned} \tag{24}$$

The different decompositions reflect simply different orders in which the dependence on the past of the subsystems is conditioned. The selection of one particular decomposition of the dependencies can only be based on side information about the system. Otherwise, one particular decomposition can only be supported *a posteriori*, that is, given how useful it results to understand how the systems work and to characterize changes in the system when examined under different conditions. None of the decompositions can be considered as preeminent, for example because it contains transfer entropy terms. For example, the first and last equality in Eq. 24 do not have a different status *a priori*. We will show that zero transfer entropies can be related to the lack of particular causal connections, but that nonzero values reflect in a complicated way both causal and noncausal sources of dependence. Considering the different measures together, as well as the entropies and conditional entropies involved in their calculation, one can better understand the dynamic dependencies in the system.

B. From causal structure to statistical dependencies quantification

We now focus on decompositions of the mutual information between the future of a subsystem and the whole past of the system like the ones in Eq. 24. For conciseness, we will refer to this mutual information as the system predictive information about a subsystem. We will examine how statistical dependencies arise from the causal structure and the sensitivity and specificity of the measures in these decompositions to different sources of dependence. We focus on a bivariate process. For simplicity we will here omit the italics in the notation of the rates. In fact, as we will see in Sec. III D all the expressions considered in this section can be interpreted as stationary rates or as defined for a particular local time i. For a bivariate process formed by $\mathcal{X}$ and $\mathcal{Y}$ the system predictive information about $\mathcal{X}$ is $I(X_{i+1};X^i,Y^i) = H(X_{i+1}) - H(X_{i+1}|X^i,Y^i)$, and analogously for Y_{i+1}. This mutual information can be decomposed in several ways. In particular:

$$I(X_{i+1};X^i,Y^i) = I(X_{i+1};X^i) + I(X_{i+1};Y^i|X^i) \tag{25a}$$
$$= I(X_{i+1};Y^i) + I(X_{i+1};X^i|Y^i) \tag{25b}$$

can be derived by expressing the mutual information and conditional mutual information as differences between two entropies:

$$\begin{aligned}I(X_{i+1};X^i) &= H(X_{i+1}) - H(X_{i+1}|X^i)\\ I(X_{i+1};Y^i|X^i) &= H(X_{i+1}|X^i) - H(X_{i+1}|X^i,Y^i)\\ I(X_{i+1};Y^i) &= H(X_{i+1}) - H(X_{i+1}|Y^i)\\ I(X_{i+1};X^i|Y^i) &= H(X_{i+1}|Y^i) - H(X_{i+1}|Y^i,X^i).\end{aligned} \tag{26}$$

In fact, Eqs. 25a and 25b correspond to only two particular ways to decompose $I(X_{i+1};X^i,Y^i)$ given the chain rule for the mutual information [39]. In particular, in these two decompositions, the whole past X^i and Y^i are kept together. Other alternative decompositions do not consider X^i, Y^i as a whole, and thus have more than two summands. However, here we are interested in these two concrete decompositions given that Eq. 25a contains as the second summand the transfer entropy $T_{Y\to X}$ (Eq. 13). Analogously, the transfer entropy $T_{X\to Y}$ appears in one of the decompositions of the system predictive information about $\mathcal{Y}$, that is,

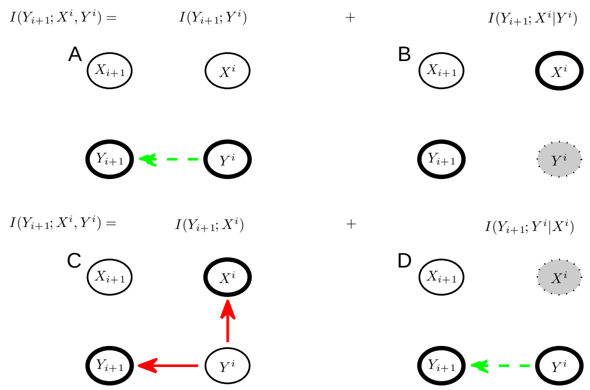

Figure 3: (Color online) The relation between causal connections and the sources of statistical dependence quantified by the terms appearing in the decomposition of the system predictive information about subsystem $\mathcal{Y}$, $I(Y_{i+1}; X^i, Y^i)$ (Eq. 27). We assume a process with unidirectional causality from $\mathcal{Y}$ to $\mathcal{X}$. We indicate in dashed green the arrows that form directed paths which are associated with a causal source of statistical dependence. We indicate in solid red the arrows that form nondirected paths which are associated to noncausal sources of statistical dependence. In each subplot we indicate in bold which are the nodes whose statistical dependence we focus on. Dotted nodes indicate conditioning, and the arrows leaving conditioned nodes are removed.

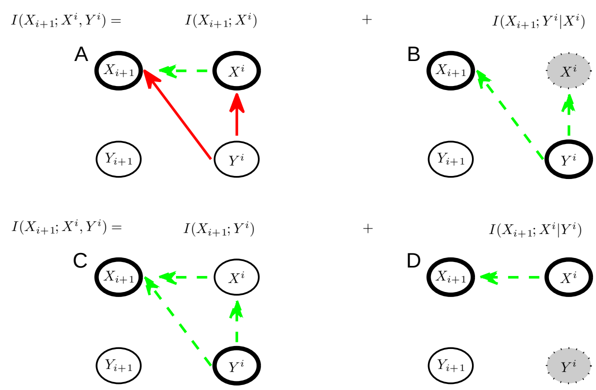

Figure 4: (Color online) The relation between causal connections and the sources of statistical dependence quantified by the terms appearing in the decomposition of the system predictive information about subsystem $\mathcal{X}$, $I(X_{i+1}; X^i, Y^i)$ (Eq. 25). The Figure is analogous to Fig. 3.

of $I(Y_{i+1}; X^i, Y^i) = H(Y_{i+1}) - H(Y_{i+1}|Y^i, X^i)$:

$$I(Y_{i+1}; X^i, Y^i) = I(Y_{i+1}; Y^i) + I(Y_{i+1}; X^i|Y^i) \quad (27a)$$
$$= I(Y_{i+1}; X^i) + I(Y_{i+1}; Y^i|X^i) \quad (27b)$$

where the mutual information and conditional mutual information appearing can be expressed in terms of a difference of entropies analogously to Eqs. 26. Given Eqs. 25 and 27, we see that the transfer entropy only appears in the decomposition of the system predictive information about a subsystem if a particular priority is chosen when quantifying the statistical dependencies.

In Figs. 3 and 4 we show how the different types of statistical dependence are quantified by the different terms of Eqs. 25 and 27, respectively. We use the mesoscopic causal graph representation because it reflects the grouping of the variables used in the transfer entropy and the measures appearing in Eqs. 25 and 27. We consider the case where there is only causality from $\mathcal{Y}$ to $\mathcal{X}$ in order to be able to discuss both the consequences of the existence and lack of causal connections across the subsystems.

We start examining the decompositions of $I(Y_{i+1}; X^i, Y^i)$ (Eq. 27) in Fig. 3. In Fig. 3A we see that the only path from Y^i to Y_{i+1} is a directed path, thus associated with a causal statistical dependence. The conditioning on X^i (Fig. 3D) does not affect this path since graphically conditioning only implies the removal of the arrows leaving the conditioned node. However, that Figs. 3A and 3D contain the same unique path does not imply that $I(Y_{i+1}; Y^i|X^i)$ is equal to $I(Y_{i+1}; Y^i)$. This is because the graphical criterion only indicates the existence of a statistical dependence, but not its magnitude.

This difference can be better appreciated considering the other terms appearing in the alternative decompositions of $I(Y_{i+1}; X^i, Y^i)$. In Fig. 3C we see that, although there are no causal connections from $\mathcal{X}$ to $\mathcal{Y}$, there is a nondirected path from X^i to Y_{i+1} that leads to $I(Y_{i+1}; X^i) > 0$. In this case conditioning on Y^i (Fig. 3B) blocks the nondirected path, as reflected in the removal of the edge from Y^i to Y_{i+1} and the lack of a path from X^i to Y_{i+1}. In consequence $I(Y_{i+1}; X^i|Y^i) = 0$. That is, the conditioning renders $I(Y_{i+1}; X^i|Y^i)$ specific for causal sources of statistical dependence from $\mathcal{X}$ to $\mathcal{Y}$ because it blocks the nondirected paths like the one leading to $I(Y_{i+1}; X^i) > 0$ in Fig. 3C.

We can appreciate the implications of choosing the alternative decompositions of $I(Y_{i+1}; X^i, Y^i)$. For the decomposition in Eq. 27a the first term $I(Y_{i+1}; Y^i)$ quantifies the underlying causal source of statistical dependence of Y_{i+1} with the past of $\mathcal{Y}$. In consequence the second term $I(Y_{i+1}; X^i|Y^i) = 0$ shows that the entropy cannot be further reduced. However, if the decomposition in Eq. 27b is selected, the first term $I(Y_{i+1}; X^i) > 0$ does not quantify a causal source of statistical dependence. Accordingly, the second term $I(Y_{i+1}; Y^i|X^i)$, despite the presence of the same directed path related to a causal statistical dependence in Figs. 3A and 3D, only quantifies the remaining entropy reduction. Therefore unidirectional causality from $\mathcal{Y}$ to $\mathcal{X}$ leads to $I(Y_{i+1}; Y^i) = I(Y_{i+1}; X^i) + I(Y_{i+1}; Y^i|X^i)$.

We now consider the alternative decompositions of $I(X_{i+1}; X^i, Y^i)$. Fig. 4A shows the paths contributing to the statistical dependence between X^i and X_{i+1}. This dependence is quantified by $I(X_{i+1}; X^i)$, the first term in the decomposition 25a. X_{i+1} depends on X^i through a directed path $X^i \to X_{i+1}$, but also due to a nondirected path. The mutual information $I(X_{i+1}; X^i)$ quantifies both the causal and the noncausal statistical dependence. The latter is present due to the existence of causal connections from $\mathcal{Y}$ to $\mathcal{X}$. Accordingly, the arrow $X^i \to X_{i+1}$ could be removed, so that no causal interactions would exist between X^i and X_{i+1}, but still we would have $I(X_{i+1}; X^i) > 0$.

In Fig. 4C we study the statistical dependence between Y^i and X_{i+1}, quantified by the first term in the decomposition of Eq. 25b, $I(X_{i+1}; Y^i)$. There are two directed paths from Y^i to X_{i+1}, associated with a direct causal statistical dependence and an indirect causal statistical dependence, respectively. Notice that these paths contain the same arrows involved in the dependence between X^i and X_{i+1} (Fig. 4A), but now there is no nondirected path contributing to the dependence.

The transfer entropy from $\mathcal{Y}$ to $\mathcal{X}$, which is the second term of Eq. 25a, results from the blocking of the directed path from Y^i to X_{i+1} mediated by X^i (Fig. 4B). Graphically, conditioning implies removing any edge leaving from X^i. Accordingly only the path $Y^i \rightarrow X_{i+1}$ remains. But notice that this arrow also is part of the nondirected path contributing as a noncausal source to the statistical dependence between X^i and X_{i+1}. Since the reduction of entropy produced by X^i is considered first in the decomposition of Eq. 25a, the transfer entropy $I(X_{i+1}; Y^i|X^i)$ only quantifies the extra reduction associated to $Y^i \rightarrow X_{i+1}$ that has not been quantified in $I(X_{i+1}; X^i)$, as evident from Eq. 26. This further illustrates, like the comparison of $I(Y_{i+1}; Y^i)$ and $I(Y_{i+1}; Y^i|X^i)$ and Figs. 3A and 3D, that the connection to a particular path does not determine the magnitude of the statistical dependence quantified.

The fact that graphically we conclude that $I(X_{i+1}; Y^i|X^i) > 0$ and $I(Y_{i+1}; X^i|Y^i) = 0$, consistently with the existence of unidirectional causal connections from $\mathcal{Y}$ to $\mathcal{X}$, indicates the validity of using the transfer entropy to test for causality. This is because here *complete observability* is fulfilled, since we have access to all the relevant variables. However, the existence of the nondirected path in Fig. 4A related to the existence of causality from $\mathcal{Y}$ to $\mathcal{X}$ also shows that the transfer entropy cannot be used as a measure of the strength of the connection between the subsystems. This becomes more evident in the case where $\mathcal{X}$ strongly depends on $\mathcal{Y}$. In the limit of $\mathcal{X}$ and $\mathcal{Y}$ being strongly synchronized conditioning on Y^i can almost not further reduce the entropy. In general, the transfer entropy is nonmonotonic with respect to the strength of the causal connection. More generally, given the overlapping of the different paths that share common edges, it is not possible to associate the magnitude of the measures with the strength of a specific connection. Only zero values indicate the lack of some arrows, as $I(Y_{i+1}; X^i|Y^i) = 0$ results from the lack of causal connections from $\mathcal{X}$ to $\mathcal{Y}$. The magnitudes of the measures are intertwined as it is clear from their expression as a difference of entropies in Eqs. 26.

C. Dynamic dependencies in exemplary processes

We will now examine some simple simulated examples. We analyze simple examples of stationary Markov binary processes and linear Gaussian stationary stochastic processes, respectively. We start by discussing a binary Markov process for which the nonmonotonicity of the transfer entropy with the strength of the causal connection is evident. We then focus on the case of the linear Gaussian stationary processes. In all these cases the information theoretic measures used are calculated analytically to isolate the fundamental properties of the measures from any estimation bias.

1. Causality in bivariate Markov binary processes

As a first example we study a stationary bivariate Markov binary process of order 1. Both $\mathcal{X}$ and $\mathcal{Y}$ take only values 0 and 1. The process is completely determined by the transition probabilities and by the condition of stationarity:

$$p(x_{i+1}, y_{i+1}) = \sum_{x_i, y_i} p(x_{i+1}, y_{i+1}|x_i, y_i) p(x_i, y_i) = p(x_i, y_i). \tag{28}$$

Furthermore, we consider that only causal connections from $\mathcal{Y}$ to $\mathcal{X}$ exists. In particular, we determine the transition probabilities for $\mathcal{Y}$ by $p(y_{i+1} = y|y_i = y) = d$, that is, d is the probability that the same value is taken at subsequent steps. The transition probabilities for $\mathcal{X}$ are such that $p(x_{i+1} = y|y_i = y) = \frac{1+g}{2}$, independently of the value of X_i. Therefore g determines the strength of the causal connection from $\mathcal{Y}$ to $\mathcal{X}$. Accordingly, the transition probabilities can be separated as the product $p(x_{i+1}, y_{i+1}|x_i, y_i) = p(x_{i+1}|y_i)p(y_{i+1}|y_i)$. For $g = 0$, X_{i+1} takes value 0 or 1 with equal probability and independently of X_i and Y_i. In the case $d = 0$, when $\mathcal{Y}$ deterministically alternates between 0 and 1, this example corresponds to one already discussed in Kaiser and Schreiber (2002) [53]. However, we here examine the dependence on d when $\mathcal{Y}$ has different degrees of stochasticity. In Fig. 5 we examine, in dependence on g and for three different values of d, some of the terms in the alternative decompositions of $I(X_{i+1}; X^i, Y^i)$ (Eq. 25). We calculate the measures using 3 time lags for X^i and Y^i, because this is enough for convergence and for more time lags the values obtained do not differ significantly.

In Fig. 5A we show $I(X_{i+1}; X^i)$, the first term in the decomposition of Eq. 25a. For $d = \frac{1}{2}$, $I(X_{i+1}; X^i) = 0$ independently of g. Apart from the lack of a causal connection from X^i to X_{i+1}, which results from the generative mechanisms of this process, also the statistical dependence produced by the nondirected path of Fig. 4A disappears for $d = \frac{1}{2}$. This is because Y_{i+1} is independent of Y^i for this d value, thus destroying the nondirected path. For $d \neq \frac{1}{2}$, Y_{i+1} statistically depends on Y^i and the presence of the nondirected path depends on the existence of a causal interaction from $\mathcal{Y}$ to $\mathcal{X}$, controlled by g. For $g \neq 0$ the nondirected path exists, leading to a noncausal statistical dependence between X^i and X_{i+1}. Accordingly, $I(X_{i+1}; X^i)$ increases with $|g|$.

In contrast to $I(X_{i+1}; X^i)$ the transfer entropy $I(X_{i+1}, Y^i|X^i)$ (Fig. 5B) decreases with d. Furthermore, we observe that it is nonmonotonic with g for low values of d. This can be understood because $I(X_{i+1}; X^i)$ already accounts for a higher part of the entropy reduction when the processes become more deterministic. Despite this nonmonotonicity, the transfer entropy fulfills the require-

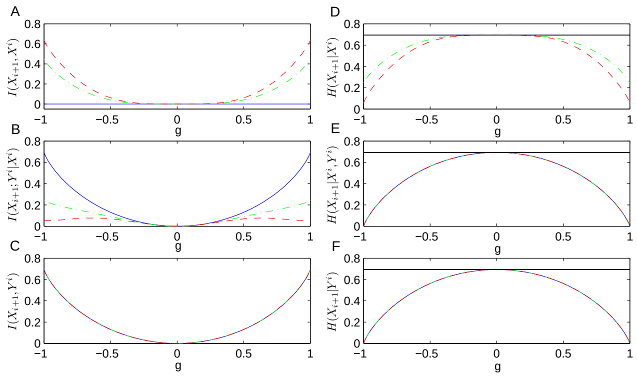

Figure 5: (Color online) Information theoretic measures involved in the Eqs. 25 for the decomposition of the system predictive information about subsystem $\mathcal{X}$ calculated for a bivariate binary Markov process of order 1. See the text for a description of the process. The measures are calculated analytically using 3 time lags to account for the past X^i and Y^i. The results are shown for $d = 1/2$ (solid blue), $d = 1/15$ (dash-dotted green), and $d = 1/100$ (dashed red), where d is the probability that $Y_{i+1} = Y_i$. In C, E, and F the three lines overlap. The black horizontal line in D-F indicates the constant value of $H(X_{i+1})$, which overlaps with the blue line in D.

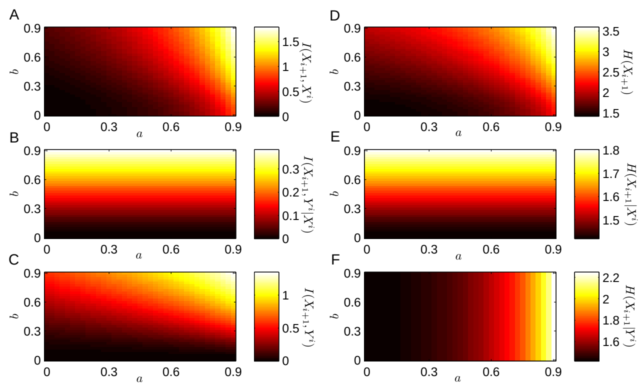

Figure 6: (Color online) Information theoretic measures involved in Eq. 25 for the decomposition of the system predictive information about subsystem $\mathcal{X}$ calculated for a bivariate linear Gaussian stationary autoregressive process of order 1 (Eq. 29). We show the dependence on the coefficients a and b for $c = 0.8$. A hot scale is used from dark (low) to light (high).

ment of a causality test, being zero if and only if $g = 0$. Furthermore, the transfer entropy in the opposite direction $I(Y_{i+1}, X^i|Y^i) = 0$ for all d, g, consistently with the lack of causality from $\mathcal{X}$ to $\mathcal{Y}$ (Results not shown).

The dependence of $I(X_{i+1}; X^i)$ and $I(X_{i+1}, Y^i|X^i)$ on d and g are related and can easily be understood considering their expression as a difference of entropies (Eqs. 26). For this particular binary Markov process $H(X_{i+1})$ is constant because $p(X_{i+1} = 0) = p(X_{i+1} = 1) = 1/2$ independently of d and g. The increase of $I(X_{i+1}; X^i)$ reflects the decay of $H(X_{i+1}|X^i)$ (Fig. 5D). Notice that this decay occurs without any causal connection from X^i to X_{i+1}, and all the statistical dependence arises from the nondirected path of Fig. 4A. Accordingly, the transfer entropy, that compares $H(X_{i+1}|X^i)$ with $H(X_{i+1}|X^i, Y^i)$, is smaller.

In Fig. 5C we show $I(X_{i+1}; Y^i)$, the first term in the alternative decomposition (Eq. 25b). The constant value $H(X_{i+1})$ is compared to $H(X_{i+1}|Y^i)$ (Fig. 5F). Notice that $H(X_{i+1}|Y^i) = H(X_{i+1}|X^i, Y^i)$, because only the directed edge $Y^i \to X_{i+1}$ exists. In contrast to $I(X_{i+1}, Y^i|X^i)$, $I(X_{i+1}; Y^i)$ quantifies all the causal statistical dependencies, as shown in Fig. 4C, and thus it monotonically increases with $|g|$. The independence of $I(X_{i+1}, Y^i)$ from d is particular for this process and reflects that, independently of d, $p(Y_i = 0) = p(Y_i = 1) = \frac{1}{2}$. Furthermore it is only zero for $g = 0$.

Although it is not our purpose in this work to address the quantification of causal effects from one process to another, notice that the fact that $I(X_{i+1}; X^i)$ quantifies a noncausal statistical dependence that results from the existence of causal connections $Y^i \to X_{i+1}$ and not from $X^i \to X_{i+1}$, and thus that the transfer entropy only quantifies the remaining dependence resulting from $Y^i \to X_{i+1}$, shows that the transfer entropy can be small in cases in which the existence of connections from $\mathcal{Y}$ to $\mathcal{X}$ has a high impact on the properties of the process $\mathcal{X}$. This indicates that, apart from not being appropriate as a measure of connectivity strength, the transfer entropy can only reflect some aspects of the existing causal effects. The fact that causal connections result in different types of causal effects was discussed in more detail in [25].

2. *Causality in bivariate linear Gaussian processes*

We now focus on linear Gaussian stationary processes. In this case the study of causality is simplified because the information theoretic measures depend only on second order moments [39]. In particular, the transfer entropy corresponds to the measure of causality proposed by Geweke [28] as shown in [46]. We consider the following linear Gaussian autoregressive process:

$$\begin{aligned} x_{i+1} &= a x_i + b y_i + \epsilon_{x,i+1} \\ y_{i+1} &= c y_i + \epsilon_{y,i+1}, \end{aligned} \tag{29}$$

where the innovations (ϵ) have zero mean and $E[\epsilon_{x,i}\epsilon_{x,j}] = \sigma^2(\epsilon_x)\delta_{ij}$, $E[\epsilon_{y,i}\epsilon_{y,j}] = \sigma^2(\epsilon_y)\delta_{ij}$, and $E[\epsilon_{x,i}\epsilon_{y,j}] = 0\ \forall i, j$. The values of a, b, and c are chosen so that the process is stationary. Unidirectional causal connections from $\mathcal{Y}$ to $\mathcal{X}$ exist for $b > 0$. We calculate the information theoretic measures analytically. Unless stated otherwise we use 10 time lags to account for the influence of the past X^i and Y^i. Closed form analytical expressions considering infinite lags can be obtained for the transfer entropy and some of the other measures using their spectral decomposition [54].

In Fig. 6 we show how terms in the decompositions of $I(X_{i+1}; X^i, Y^i)$ depend on the coefficients a and b, while keeping a constant value $c = 0.8$ and the variance of the innovations $\sigma^2(\epsilon_x) = \sigma^2(\epsilon_y) = 1$. In Fig. 6A we see $I(X_{i+1}; X^i)$, the first term of the decomposition in Eq. 25a. The influence of the noncausal source of statistical dependence (Fig. 4A) is reflected in the dependence on b, especially in the fact that for sufficiently high values of b, $I(X_{i+1}; X^i)$ is nonzero even if $a = 0$.

By contrast, the transfer entropy from $\mathcal{Y}$ to $\mathcal{X}$ (Fig. 6B) depends only on b, increasing monotonically with it. In general though, for processes of order higher than one, the transfer entropy is not specifically related to a single coupling coefficient like here it is with b, but would reflect the joint influence of the different connections $Y_{i-k} \to X_{i+1}$, for $k \geq 0$. Furthermore, as we discuss below, the independence of the transfer entropy from the coefficients associated with directed edges from the past of $\mathcal{X}$ to X_{i+1} should not be considered as resulting from the separability of the different sources of dependence even for linear processes.

These dependencies can be understood considering the expression of the measures as a difference of entropies (Eqs. 26). Oppositely to the Markov process studied above, here $H(X_{i+1})$ (Fig. 6D) increases with a and b, reflecting an increase in the variance. Given the linearity of the autoregressive process the conditional entropy $H(X_{i+1}|X^i)$ is independent from a. This independence is inherited by the transfer entropy since the other term in Eq. 26, $H(X_{i+1}|X^i, Y^i)$, is constant and determined by the variance of the innovation term, which is the only remaining source of uncertainty. Analogously to $H(X_{i+1}|X^i)$, the conditional entropy $H(X_{i+1}|Y^i)$ is independent from b (Fig. 6F). The independence from b is inherited by $I(X_{i+1}; X^i|Y^i)$ (Results not shown). This may also suggest that, given linearity, the dependence on the past of $\mathcal{X}$ and on the past of $\mathcal{Y}$ are separable.

By contrast, $I(X_{i+1}; Y^i)$, depends both on b and a, as can be seen in Fig. 6C. According to Fig. 4C, $I(X_{i+1}; Y^i)$ specifically reflects only causal statistical dependencies. Since it is sensitive also to the dependence produced by the indirect path from Y^i to X_{i+1} mediated by X^i, $I(X_{i+1}; Y^i)$ quantifies the total dependence produced by the causal connections between the processes, and resulting from paths including arrows $Y_{i-k} \to X_{i-k+1}$, $X_{i-k} \to X_{i-k+1}$ and $Y_{i-k} \to Y_{i-k+1}$, which explains the dependence on both b and a.

We continue further examining in Fig. 7 the dependence on the coefficients b and c for $a = 0.1$. For this value of a the statistical dependence between X_{i+1} and X^i can be dominated by the influence of nondirected paths (Fig. 4A). In particular we see in Fig. 7A that $I(X_{i+1}; X^i)$ substantially increases when both b and c are high, since both arrows $Y_k \to Y_{k+1}$ and $Y_i \to X_{i+1}$ are needed for the nondirected paths to be effective.

The transfer entropy $I(X_{i+1}, Y^i|X^i)$ depends not only on b but on c (Fig. 7B). This coefficient determines the entropy rate of $\mathcal{Y}$ and thus also how coherent are the causal interactions resulting from causal connections from $\mathcal{Y}$ to $\mathcal{X}$ at subsequent times. As we mentioned $H(X_{i+1}|X^i, Y^i)$ is constant for a fixed variance of the innovations, so that the transfer entropy inherits its dependencies from $H(X_{i+1}|X^i)$ (Fig. 7E).

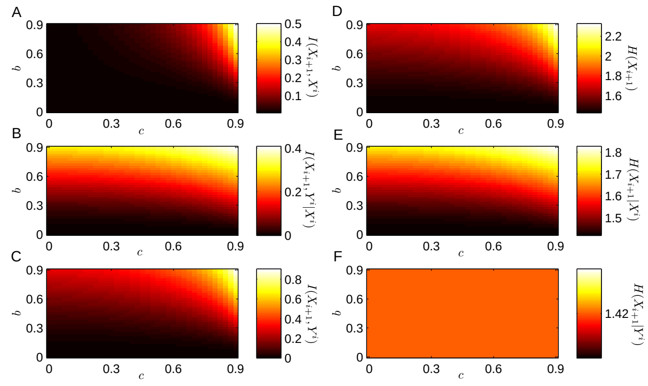

Figure 7: (Color online) Information theoretic measures involved in Eq. 25 for the decomposition of the system predictive information about subsystem $\mathcal{X}$ calculated for a bivariate linear Gaussian stationary autoregressive process of order 1 (Eq. 29). We show the dependence on the coefficients b and c for $a = 0.1$. A hot scale is used from dark (low) to light (high).

Finally, $I(X_{i+1}; Y^i)$ also depends on b and c (Fig. 7C), and inherits this dependence from $H(X_{i+1})$ (Fig. 7D). Given the linearity of the autoregressive process $H(X_{i+1}|Y^i)$ is constant for fixed a.

Altogether we can see that the dependencies found in Figs. 6 and 7 can be inferred from the arrows that are involved in the paths that produce the sources of dependence (Fig. 4). At the mesoscopic level one also need to have into account that the internal structure of Y^i is determined by c. In the microscopic scale the criterion of examining the involved arrows would be sufficient, but we preferred the mesoscopic scale to find a compromise between simplicity and informativeness of the representation.

As we mentioned, the independence of the transfer entropy from a could lead to believe that for linear Gaussian stationary processes direct and indirect effects are separable by means of conditioning. However, this is clearly contradicted by the inequality of the transfer entropy and $I(X_{i+1}; Y^i)$ for $a = 0$, as can be seen in Fig. 8A. To see this, we consider the dependence of the transfer entropy on the number of time lags included in the past of $\mathcal{X}$ and $\mathcal{Y}$. We discuss first the transfer entropy from $\mathcal{X}$ to $\mathcal{Y}$, $I(Y_{i+1}, X^i|Y^i)$. As explained in Sec. III B conditioning on Y^i is needed to avoid a false positive detection of causality due to the noncausal statistical dependence produced by the nondirected path in Fig. 3C. However, in this example, since the causal connections from the past of $\mathcal{Y}$ to Y_{i+1} are of order 1, namely $Y_i \to Y_{i+1}$, any nondirected path from X^i to Y_{i+1} included in Fig. 3C is blocked when conditioning only on Y_i. In general, if $i - k'$ is the oldest time in the past for which a directed edge to Y_{i+1} exists, it suffices to condition on $Y_i, ..., Y_{i-k'}$ to remove the influence of the noncausal sources of statistical dependence. Therefore here $I(Y_{i+1}, X^i|Y_i)$ is already zero.

We now consider the dependence of $I(X_{i+1}, Y^i|X^i)$ on the number of time lags. As said above $H(X_{i+1}|X^i, Y^i)$ re-

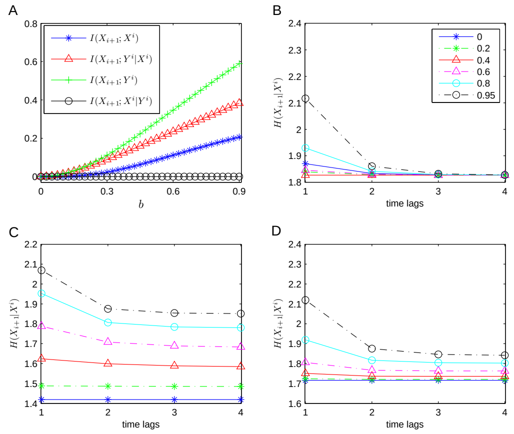

Figure 8: (Color online) Statistical dependencies quantified by the transfer entropy. A: Comparison of the transfer entropy with the other measures appearing in the decompositions in Eq. 25 for the process of Eq. 29 in dependence on b with $a = 0$ and $c = 0.8$. B-D: Dependence of the conditional entropy $H(X_{i+1}|X^i)$ on the time lags used to account for the past of $\mathcal{X}$. Different colors (symbols-line styles) correspond to values of the parameter (indicated in B) varied in each panel. The other two parameters are fixed to 0.8. B: Parameter a is changed. C: Parameter b is changed. D: Parameter c is changed.

mains constant as long as $\sigma^2(\epsilon_x)$ does not change, and thus we should focus on the dependence of $H(X_{i+1}|X^i)$ (Fig. 8B-D). Given Eq. 29 it suffices to condition on X_i to quantify all the entropy reduction that is associated with the causal source of statistical dependence between X_{i+1} and X^i in Fig. 4A. But the existence of the nondirected path in Fig. 4A implies that the rest of the past of $\mathcal{X}$ has still a dependence with X_{i+1}. This nondirected path overlaps with the directed path $Y^i \rightarrow X_{i+1}$ (Fig. 4C) associated with the transfer entropy from $\mathcal{Y}$ to $\mathcal{X}$. Accordingly further conditioning on X^i beyond X_i removes part of the statistical dependence which is introduced due to the term bY_i. Similarly, for $H(X_{i+1}|Y^i)$ it suffices to take one time lag to quantify the direct causal dependence associated with bY_i, but the rest of the past of $\mathcal{Y}$ is still connected to X_{i+1} after conditioning on Y_i through the indirect path mediated by X^i.

In particular, we see in Fig. 8B that $H(X_{i+1}|X^i)$ and thus the transfer entropy depends on a for $1, 2$ time lags. These time lags are sufficient to correctly infer the causal structure of the graph and retrieve $I(Y_{i+1}, X^i|Y^i) = 0$ but provide a different value of $I(X_{i+1}, Y^i|X^i)$. Furthermore, the dependence on a is nonmonotonic, since a increases the variance of $\mathcal{X}$ but also its predictability, which have opposite effects on $H(X_{i+1}|X^i)$. By contrast, $H(X_{i+1}|X^i)$ monotonically increases with b (Fig. 8C), and c (Fig. 8A) for any number of time lags. For higher b and c the dependence on this number is higher.

The dependence of the transfer entropy on the time lags shows that conditioning does not separate the direct and indirect components of the dependence resulting from the causal connections. The statistical dependence of X_{i+1} on X^i and on Y^i are entangled since they arise from overlapping paths. Despite the linear structure of Eq. 29 the covariance matrices and thus the measures quantifying the dependencies in the process are nonlinear in the parameters a, b, c, including terms like cb (See for example Eq. 43 in [54]).

D. An information theoretic framework of dynamic dependencies in networks of interacting processes

Above we illustrated how the different components of the system predictive information about a subsystem (Eqs. 25 and 27) are sensitive to different sources of dependence. This has been useful to understand how the causal structure produces statistical dependencies and how the resulting statistical structure of the system is captured in different decompositions. We have seen that causal and non causal sources of dependence are not separable, and that the measures are not specific to single properties of the causal structure, i. e. changes in the transfer entropy may reflect changes in the autocorrelation of a process and not changes in interdependence. Accordingly, to characterize the statistical structure of the system a unified framework of measures should be preferred to isolated measures. We now want to formulate in detail the general framework to characterize all the dynamic dependencies in a network of processes. We outlined the main components of this formalism in Sec. III A for the stationary case. Here we will consider two alternative non-stationary formulations of the framework, one local in time, and the other cumulative on the whole time series. We start reviewing in Sec. III D 1 the decomposition of the mutual information into components that comprise the transfer entropy. We then comment on the lack of an analogous relationship for the conditional measures (Sec. III D 2) in the non-stationary case, thus indicating the differences between the unconditional (Eq. 16) and conditional (Eq. 18) decompositions. We also consider how the bivariate measures are sensitive to different sources of dependence arising in a multivariate system and we link them to the conditional measures (Sec. III D 3). We finally consider the multivariate measures to characterize the statistical structure proposed in Sec. III A and how they can be decomposed into different components in Sec. III D 4. Motivated by these decompositions we discuss different types of normalizations of the transfer entropy (Sec. III D 5). We then also consider alternative decompositions comprising transfer-entropy-like terms, but which have a larger horizon in the future than $i + 1$.

1. *Mutual information and transfer entropy in bivariate systems*

We here follow [30, 31, 55] and review a non-stationary formulation of the relationship between mutual information and transfer entropy. The corresponding formulation for stationary processes was presented in [21, 28]. The cumulative mutual information between two time series $X^N = \{X_1, X_2, ..., X_N\}$, $Y^N = \{Y_1, Y_2, ..., Y_N\}$ recorded from $\mathcal{X}$

and $\mathcal{Y}$, respectively, is

$$I(X^N;Y^N) = \sum_{w_{xy}^N} p(w_{xy}^N) \log \frac{p(w_{xy}^N)}{p(x^N)p(y^N)}$$
$$= \sum_{k=0}^{N-1} \mathrm{KL}(p(x_{k+1}, y_{k+1}|w_{xy}^k); p(x_{k+1}|x^k)p(y_{k+1}|y^k))$$
$$= H(X^N) + H(Y^N) - H(X^N, Y^N)$$
$$= \sum_{k=0}^{N-1} H(X_{k+1}|X^k) + \sum_{k=0}^{N-1} H(Y_{k+1}|Y^k) -$$
$$\sum_{k=0}^{N-1} H(X_{k+1}, Y_{k+1}|X^k, Y^k), \tag{30}$$

where $w_{xy}^N = \{x^N, y^N\}$.

The cumulative transfer entropy from $\mathcal{Y}$ to $\mathcal{X}$ is

$$\begin{aligned} T_{Y^N \to X^N} &= \sum_{w_{xy}^N} p(w_{xy}^N) \log \frac{\prod_{k=0}^{N-1} p(x_{k+1}|w_{xy}^k)}{\prod_{k=0}^{N-1} p(x_{k+1}|x^k)} \\ &= \sum_{k=0}^{N-1} \mathrm{KL}(p(x_{k+1}|w_{xy}^k); p(x_{k+1}|x^k)) \\ &= \sum_{k=0}^{N-1} H(X_{k+1}|X^k) - \sum_{k=0}^{N-1} H(X_{k+1}|X^k, Y^k) \\ &= \sum_{k=0}^{N-1} I(X_{k+1}; Y^k|X^k). \end{aligned} \tag{31}$$

A measure of cumulative instantaneous dependence between the time series is given by

$$T_{Y^N \cdot X^N} = \sum_{w_{xy}^N} p(w_{xy}^N) \log \frac{\prod_{k=0}^{N-1} p(x_{k+1}, y_{k+1}|w_{xy}^k)}{\prod_{k=0}^{N-1} p(x_{k+1}|w_{xy}^k)p(y_{k+1}|w_{xy}^k)}$$
$$= \sum_{k=0}^{N-1} \mathrm{KL}(p(x_{k+1}, y_{k+1}|w_{xy}^k); p(x_{k+1}|w_{xy}^k)p(y_{k+1}|w_{xy}^k))$$
$$= \sum_{k=0}^{N-1} H(X_{k+1}|X^k, Y^k) + \sum_{k=0}^{N-1} H(Y_{k+1}|X^k, Y^k)$$
$$- \sum_{k=0}^{N-1} H(X_{k+1}, Y_{k+1}|X^k, Y^k)$$
$$= \sum_{k=0}^{N-1} I(X_{k+1}; Y_{k+1}|X^k, Y^k). \tag{32}$$

These measures are related so that the cumulative mutual information can be decomposed as

$$I(X^N;Y^N) = T_{Y^N \to X^N} + T_{X^N \to Y^N} + T_{Y^N \cdot X^N}. \tag{33}$$

We discussed the interpretation of this decomposition in Sec. III A. For a bivariate system, and assuming we have access to the dynamic processes between which the interactions really occur, the three terms in the rhs can be related to the existence of particular arrows in a graph analogous to the ones of Fig. 1 (but considering in general that the connections do not need to be stationary). A nonzero transfer entropy $T_{Y^N \to X^N}$ indicates that at least there is some causal connection from $\mathcal{Y}$ to $\mathcal{X}$, and analogously for $T_{X^N \to Y^N}$. The exact magnitude of positive values depends on different sources of dependence as we have vastly discussed above. As mentioned, the term of instantaneous causality $T_{Y^N \cdot X^N}$ should be zero for a truly bivariate process.

An analogous decomposition is obtained when the transfer entropy and instantaneous dependence are defined locally. The transfer entropy at time i is defined as

$$T_{Y \to X}(i) = I(X_{i+1}; Y^i|X^i), \tag{34}$$

and the instantaneous dependence at time i as

$$T_{Y \cdot X}(i) = I(X_{i+1}; Y_{i+1}|X^i, Y^i). \tag{35}$$

It can easily be checked [55] that

$$\begin{aligned} I(X_{i+1}, X^i; Y_{i+1}, Y^i) - I(X^i; Y^i) = T_{Y \to X}(i) \\ + T_{X \to Y}(i) + T_{Y \cdot X}(i). \end{aligned} \tag{36}$$

Furthermore, for discrete valued processes and assuming stationarity, it can be checked [44], based on the equivalence between the average and conditional entropy rates [Theorem 4.2.1 in 39] (and see Sec. II A), that

$$\begin{aligned} \lim_{i \to \infty} I(X_{i+1}, X^i; Y_{i+1}, Y^i) - I(X^i; Y^i) \\ = \lim_{N \to \infty} \frac{I(X^N; Y^N)}{N}, \end{aligned} \tag{37}$$

that is, the average mutual information is equal to the increase rate in the information when including one sample more. So both the local and cumulative formulation converge into Eq. 16 for stationary processes.

The equality of these two formulations when assuming stationarity justifies why it is common to drop the temporal argument in $T_{Y \to X}(i)$. The assumption of stationarity in practical applications depends on the way the measure is estimated, either sampling the distributions across time or across trials with a reference of time related to some event, like an stimulus presentation. We will here drop the argument in cases in which we refer to properties that hold for both the cumulative and the local measures. Furthermore, notice that here *local* means only local in time, in contrast to the local transfer entropy introduced in [56], which is local in the values taken by the variables. On the other hand, *cumulative* means for the whole time series.

2. *Conditional mutual information and conditional transfer entropy in multivariate systems*

We here develop in detail some results sketched in [55] and we will build on them to relate bivariate and conditional measures. The cumulative conditional information of $\mathcal{X}$ and $\mathcal{Y}$ given $\mathcal{Z}$ from the time series $X^N = \{X_1, X_2, ..., X_N\}$, $Y^N = \{Y_1, Y_2, ..., Y_N\}$, $Z^N = \{Z_1, Z_2, ..., Z_N\}$ is

$$\begin{aligned} I(X^N; Y^N|Z^N) &= \sum_{w_{xyz}^N} p(w_{xyz}^N) \log \frac{p(w_{xy}^N|z^N)}{p(x^N|z^N)p(y^N|z^N)} \\ &= \sum_{k=0}^{N-1} \mathrm{KL}(p(x_{k+1}, y_{k+1}|w_{xy}^k, z^N); \\ &\quad p(x_{k+1}|x^k, z^N)p(y_{k+1}|y^k, z^N)), \end{aligned} \tag{38}$$

which can be expressed in terms of sums of conditional entropies in analogy to $I(X^N; Y^N)$ (Eq. 30). The cumulative conditional transfer entropy from $\mathcal{Y}$ to $\mathcal{X}$ given $\mathcal{Z}$ is

$$\begin{aligned} T_{Y^N \to X^N|Z^N} &= \sum_{w_{xyz}^N} p(w_{xyz}^N) \log \frac{\prod_{k=0}^{N-1} p(x_{k+1}|w_{xyz}^k)}{\prod_{k=0}^{N-1} p(x_{k+1}|w_{xz}^k)} \\ &= \sum_{k=0}^{N-1} I(X_{k+1}; Y^k|X^k, Z^k). \end{aligned} \tag{39}$$

Similarly, the cumulative conditional instantaneous dependence of $\mathcal{Y}$ and $\mathcal{X}$ given $\mathcal{Z}$ is

$$\begin{aligned} &T_{Y^N \cdot X^N|Z^N} \\ &= \sum_{w_{xyz}^N} p(w_{xyz}^N) \log \frac{\prod_{k=0}^{N-1} p(x_{k+1}, y_{k+1}|w_{xyz}^k)}{\prod_{k=0}^{N-1} p(x_{k+1}|w_{xyz}^k)p(y_{k+1}|w_{xyz}^k)} \\ &= \sum_{k=0}^{N-1} I(X_{k+1}; Y_{k+1}|X^k, Y^k, Z^k). \end{aligned} \tag{40}$$

From these definitions it is easy to check that, in contrast to the bivariate case,

$$\begin{aligned} I(X^N; Y^N|Z^N) \neq T_{Y^N \to X^N|Z^N} + T_{X^N \to Y^N|Z^N} \\ + T_{Y^N \cdot X^N|Z^N}. \end{aligned} \tag{41}$$

An extension under the stationary assumption of the decompositions of the mutual information in Eqs. 33 and 36 to the conditional case was provided in [32] (further assuming linearity) and corresponds to the one shown in Eq. 18. There it was not specified in which way this extension results as a limit from the non-stationary case. In fact, it can be seen that Eq. 18 implies taken a limit from the local definitions of the measures. In particular, given the local conditional transfer entropy

$$\begin{aligned} T_{Y \to X|Z}(i) &= H(X_{i+1}|X^i, Z^i) - H(X_{i+1}|X^i, Y^i, Z^i) \\ &= I(X_{i+1}; Y^i|X^i, Z^i) \end{aligned} \tag{42}$$

and local conditional instantaneous dependence

$$T_{Y \cdot X|Z}(i) = I(X_{i+1}; Y_{i+1}|X^i, Y^i, Z^i), \tag{43}$$

it is immediate to check that

$$\begin{aligned} &I(X_{i+1}, X^i; Y_{i+1}, Y^i|Z^i) - I(X^i; Y^i|Z^i) \\ &= T_{Y \to X|Z}(i) + T_{X \to Y|Z}(i) + T_{Y \cdot X|Z}(i). \end{aligned} \tag{44}$$

However, although

$$\begin{aligned} \lim_{N \to \infty} \frac{T_{Y^N \to X^N|Z^N}}{N} + \frac{T_{X^N \to Y^N|Z^N}}{N} + \frac{T_{Y^N \cdot X^N|Z^N}}{N} \\ = \lim_{i \to \infty} T_{Y \to X|Z}(i) + T_{X \to Y|Z}(i) + T_{Y \cdot X|Z}(i), \end{aligned} \tag{45}$$

the limit for $N \to \infty$ of the lhs of Eq. 44 and the one of the average conditional mutual information in Eq. 38 are in general not equivalent definitions of a conditional mutual information rate

$$\begin{aligned} \lim_{i \to \infty} I(X_{i+1}, X^i; Y_{i+1}, Y^i|Z^i) - I(X^i; Y^i|Z^i) \\ \neq \lim_{N \to \infty} \frac{I(X^N; Y^N|Z^N)}{N}. \end{aligned} \tag{46}$$

To understand why they are not equivalent in general it is helpful to consider in which cases they are. In particular we need to see when z^N can be substituted by z^k in the conditioning part of the distributions appearing in the expression of $I(X^N; Y^N|Z^N)$ as an average of Kullback-Leibler divergences. It can be checked that this is only possible if there are no connections from $\mathcal{X}$ or $\mathcal{Y}$ to $\mathcal{Z}$. If the future of $\mathcal{Z}$ is caused by the past of the other processes, the conditioning on this future is not compatible with a temporally asymmetric decomposition as the one comprising the conditional transfer entropies.

3. *The link between bivariate and conditional measures*

The conditional measures or the rates obtained from them, are related to the bivariate measures. In particular, considering a generalization of the relation in [32]

$$\begin{aligned} T_{YZ \to X} &= T_{Z \to X} + T_{Y \to X|Z} \\ &= T_{Y \to X} + T_{Z \to X|Y}, \end{aligned} \tag{47}$$

where $T_{XY \to Z}$ is the transfer entropy of $\mathcal{X}$ and $\mathcal{Y}$ considered as a joint subsystem to $\mathcal{Z}$. This relation is just a straightforward application of the chain rule for conditional mutual information [39]. Accordingly

$$T_{Y \to X} = T_{Y \to X|Z} + T_{Z \to X} - T_{Z \to X|Y}. \tag{48}$$

Even if there is no direct causal connection from $\mathcal{Y}$ to $\mathcal{X}$ ($T_{Y \to X|Z} = 0$) the bivariate measure can still be positive. It is well-known [2] that if there exist

$$\text{indirect causal connections from } Y \text{ to } X \Rightarrow T_{Y \to X} > 0 \tag{49}$$

and that if there exist

$$\text{common drivers of } Y \text{ and } X \Rightarrow T_{Y \to X} > 0. \tag{50}$$

The process $\mathcal{Z}$ can both be a common driver or mediate indirect causal connections.

To our knowledge less attention has been paid to the relation between the instantaneous dependencies. In particular we found that

$$T_{X \cdot Y} = T_{X \cdot Y|Z} + T_{Z \to X|Y} + T_{Z \to Y|X} - T_{Z \to XY}. \quad (51)$$

It can be seen that in contrast to the transfer entropy

$$\text{indirect causal connections from } Y \text{ to } X \nRightarrow T_{Y \cdot X} > 0. \quad (52)$$

However, like for the transfer entropy

$$\text{common drivers of } Y \text{ and } X \Rightarrow T_{Y \cdot X} > 0. \quad (53)$$

That is, even if in the multivariate system including $\mathcal{Z}$ there are no instantaneous dependencies ($T_{X \cdot Y|Z} = 0$), the bivariate instantaneous dependence can be positive when $\mathcal{Z}$ is a common driver and is not considered explicitly. If the bivariate decomposition of the mutual information (Eqs. 33 and 36) is applied to part of a multivariate system, the indirect causal statistical dependencies due to indirect connections and the noncausal statistical dependencies due to common drivers, are not separable. The transfer entropy is sensitive to both and the instantaneous dependence quantifies the extra dependence due to common drivers which is not already accounted by transfer entropies. When dealing with a bivariate process as we did in the exemplary analysis presented above we already illustrated the difficulty to analyze the transfer entropy without the context of the other measures. This comparison of the bivariate and conditional measures serves to further indicate that the dependencies produced by hidden variables can be distributed in the three terms of the decomposition of the bivariate mutual information in a not straightforward way.

4. *Statistical structure and measures of dynamic dependence in multivariate systems*

We now finally address the fully multivariate characterization of the dynamic dependencies in a network of interacting processes. We start by considering the statistical structure of the system when a number of subsystems are distinguished. We will proceed considering first a multivariate system formed by subsystems $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$, and after see how the analysis is generalized. Consider the cumulative subsystems multi-information

$$M(X^N;Y^N;Z^N) = \sum_{w_{xyz}^N} p(w_{xyz}^N) \log \frac{p(w_{xyz}^N)}{p(x^N)p(y^N)p(z^N)} \quad (54)$$

that quantifies the total dependence between the time series of the three subsystems. We want to see how it can be decomposed into different components. To do this we use a general property of decomposability of the multi-information [34]. Given a bipartition of the subsystems in two groups of n' and $n-n'$ subsystems respectively

$$\begin{aligned} M(V_1^N;&V_2^N;...;V_n^N) \\ &= M(V_1^N;V_2^N;...;V_{n'}^N) \\ &+ M(V_{n'+1}^N;V_{n'+2}^N;...;V_n^N) \\ &+ I(V_1^N,V_2^N,...,V_{n'}^N;V_{n'+1}^N,V_{n'+2}^N,...,V_n^N), \end{aligned} \quad (55)$$

so that the total dependence can be divided into the total dependence internal to each group and the mutual information between these two groups. In particular, since the multi-information corresponds to the mutual information for the bivariate case,

$$M(X^N;Y^N;Z^N) = I(X^N;Y^N) + I(X^N,Y^N;Z^N). \quad (56)$$

Furthermore, let us define the cumulative multivariate instantaneous dependence in agreement with [32] as

$$\begin{aligned} T(X^N \cdot Y^N \cdot Z^N) &= \sum_{w_{xyz}^N} p(w_{xyz}^N) \\ \log &\frac{\prod_{k=0}^{N-1} p(x_{k+1},y_{k+1},z_{k+1}|w_{xyz}^k)}{\prod_{k=0}^{N-1} p(x_{k+1},|w_{xyz}^k)p(y_{k+1}|w_{xyz}^k)p(z_{k+1}|w_{xyz}^k)} \\ &= \sum_{k=0}^{N-1} M(X_{k+1};Y_{k+1};Z_{k+1}|X^k,Y^k,Z^k). \end{aligned} \quad (57)$$

This quantity can be expressed as the sum of a bivariate and a conditional instantaneous dependence given the relation provided in [32]:

$$T_{X \cdot Y \cdot Z} = T_{XY \cdot Z} + T_{X \cdot Y|Z}. \quad (58)$$

We can now use the decomposition of the two terms on the rhs of Eq. 56 in terms of transfer entropies and instantaneous dependencies to obtain

$$\begin{aligned} M(X^N;Y^N;Z^N) = &T_{Y^N Z^N \to X^N} + T_{X^N Z^N \to Y^N} \\ &+ T_{X^N Y^N \to Z^N} + T_{X^N \cdot Y^N \cdot Z^N}. \end{aligned} \quad (59)$$

Taking the local counterparts of the measures appearing in the rhs of Eq. 59 one can see that

$$\begin{aligned} M(X_{i+1},X^i;Y_{i+1},Y^i;Z_{i+1},Z^i) - M(X^i;Y^i;Z^i) \\ = T_{YZ \to X}(i) + T_{XZ \to Y}(i) \\ + T_{XY \to Z}(i) + T_{X \cdot Y \cdot Z}(i). \end{aligned} \quad (60)$$

Furthermore, the resulting rates from the limit $N \to \infty$ are equivalent for the local and average formulation

$$\begin{aligned} \lim_{i \to \infty} M(X_{i+1},X^i;Y_{i+1},Y^i;Z_{i+1},Z^i) - M(X^i;Y^i;Z^i) \\ = \lim_{N \to \infty} \frac{M(X^N;Y^N;Z^N)}{N}, \end{aligned} \quad (61)$$

as it is the case for the mutual information in bivariate processes (Sec. III D 1). In fact, as we said, this formulation can be seen as the natural multivariate generalization of the formulation for bivariate process, and the decompositions of the subsystems multi-information in Eqs. 59 and 60 subsume the ones in Eqs 33 and 36, respectively.

According to the general definition of subsystems multi-information in Eq. 19, in general we have that the subsystems multi-information of n subsystems can be decomposed as

$$\begin{aligned} M(V_1^N; V_2^N; ...; V_n^N) &= \sum_{j=1}^{n-1} I(V_j^N; V_{j+1}^N, ..., V_n^N) \\ &= T_{V_1^N \cdot V_2^N \cdot ... \cdot V_n^N} + \sum_{j=1}^{n} T_{\{V^N\} \setminus V_j^N \to V_j^N}. \end{aligned} \tag{62}$$

The first decomposition in terms of mutual informations results directly from Eq. 55. It is not symmetric because it is based on a hierarchy chosen to cluster the subsystems which is a priori arbitrary except further information about the structure of the system is used to select it. Furthermore the mutual informations do not have any temporal asymmetry. Oppositely, the second decomposition is symmetric in the decomposition of the subsystems and uses the temporal asymmetry. This decomposition corresponds to the result outlined in Eq. 22. The first term is the multivariate instantaneous dependence. If we assume that we include the whole system in the analysis, so that there are no dependencies produced by external processes, this term is zero. The second term, the summand of transfer entropies, is symmetric because it comprises all the transfer entropies from the rest of the system to a given subsystem. The transfer entropy $T_{\{V^N\} \setminus V_j^N \to V_j^N}$ accounts for any further entropy reduction that can be obtained considering the past of the rest of the system once the own past of $\mathcal{V}_j$ has been considered. Since the partition is bivariate and we consider that there are not external influences, this reduction results exclusively from causal statistical dependencies from the rest of the system to $\mathcal{V}_j$. However, as discussed in Sec. III A, this decomposition does not allow us to interpret it in terms of reciprocal causal influences between the subsystems.

We continue considering the cumulative temporal internal multi-information of a single process, which we will need later to decompose the global multi-information. According to Eq. 21, for a subsystem $\mathcal{X}$

$$\begin{aligned} M(X_1; X_2; ...; X_N) &= \sum_{x^N} p(x^N) \log \frac{p(x^N)}{\prod_{k=0}^{N-1} p(x_{k+1})} \\ &= \sum_{k=0}^{N-1} I(X_{k+1}; X^k). \end{aligned} \tag{63}$$

We see that this measure corresponds to the cumulative mutual information between each variable X_{k+1} and its past. This is valid not only for single processes but for any subset, or for the whole set $\mathcal{W}$ of processes in the system. For example, for $\mathcal{W} = \{\mathcal{X}, \mathcal{Y}, \mathcal{Z}\}$ we have that

$$\begin{aligned} &M(W_1; W_2; ...; W_N) \\ &\quad = \sum_{k=0}^{N-1} I(X_{k+1}, Y_{k+1}, Z_{k+1}; X^k, Y^k, Z^k). \end{aligned} \tag{64}$$

The multi-information in Eq. 63 is a quantification of the temporal dependencies of the subsystem. It is a natural generalization of the autocorrelation and as discussed in Sec. III B it depends also on the interaction with the other subsystems and not only on the internal interactions.

The local increase of the cumulative temporal internal multi-information of $\mathcal{X}$ is

$$M(X_1; X_2; ...; X_{i+1}) - M(X_1; X_2; ...; X_i) = I(X_{i+1}; X^i), \tag{65}$$

where the equality results from Eq. 55. Accordingly, an equivalent rate for stationary processes can also be obtained from the average or local temporal internal multi-information

$$\begin{aligned} &\lim_{i \to \infty} M(X_1; X_2; ...; X_{i+1}) - M(X_1; X_2; ...; X_i) \\ &= \lim_{N \to \infty} \frac{M(X_1; X_2; ...; X_N)}{N}. \end{aligned} \tag{66}$$

Finally, we consider the global multi-information used to characterize the global statistical structure of the system (Eq. 20) considered as a unique unit. For the case of a bivariate process the cumulative global multi-information is

$$\begin{aligned} &M(X_1; X_2; ...; X_N; Y_1; Y_2; ...; Y_N) \\ &= M(X_1; X_2; ...; X_N) + M(Y_1; Y_2; ...; Y_N) \\ &+ I(X^N; Y^N) \\ &= \sum_{k=0}^{N-1} I(X_{k+1}; X^k) + \sum_{k=0}^{N-1} I(Y_{k+1}; Y^k) + T_{Y^N \to X^N} \\ &+ T_{X^N \to Y^N} + T_{Y^N \cdot X^N}. \end{aligned} \tag{67}$$

We see that the global multi-information can be decomposed into components accounting for the temporal internal statistical structure of each individual process and the mutual information between the processes accounting for the statistical structure across processes. The second equality is obtained combining Eq. 63 and the decomposition of the mutual information into transfer entropies and instantaneous dependence (Eq. 33). Furthermore, assembling the temporal internal mutual informations with the transfer entropies we get

$$\begin{aligned} &M(X_1; X_2; ...; X_N; Y_1; Y_2; ...; Y_N) \\ &= \sum_{k=0}^{N-1} I(X_{k+1}; X^k, Y^k) + \sum_{k=0}^{N-1} I(Y_{k+1}; X^k, Y^k) \\ &+ T_{Y^N \cdot X^N}. \end{aligned} \tag{68}$$

Accordingly the cumulative global multi-information can be expressed as a sum of the cumulative instantaneous dependence and the cumulative mutual information of each subsystem with the whole past of the system, which is the cumulative system predictive information about a subsystem.

Given the general definition for n subsystems (Eq. 20) we have that

$$\begin{aligned} &M(V_{11}; V_{12}; ...; V_{1N}; V_{21}; V_{22}; ...; V_{2N}; V_{n1}; V_{n2}; ...; V_{nN}) \\ &= \sum_{j=1}^{n} M(V_{j1}; V_{j2}; ...; V_{jN}) + M(V_1^N; V_2^N; ...; V_n^N) \\ &= \sum_{j=1}^{n} \sum_{k=0}^{N-1} I(V_{jk+1}; V_j^k) + T_{V_1^N \cdot V_2^N \cdot ... \cdot V_n^N} \\ &+ \sum_{j=1}^{n} T_{\{V\}^N \setminus V_j^N \to V_j^N}. \end{aligned} \tag{69}$$

The first equality shows the relation between the global, temporal, and subsystems multi-information. The second is a temporally asymmetric decomposition of the dependencies. Also generally

$$\begin{aligned} &M(V_{11}; V_{12}; ...; V_{1N}; V_{21}; V_{22}; ...; V_{2N}; V_{n1}; V_{n2}; ...; V_{nN}) \\ &= \sum_{j=1}^{n} \sum_{k=0}^{N-1} I(V_{jk+1}; V_1^k, ..., V_n^k) + T_{V_1^N \cdot V_2^N \cdot ... \cdot V_n^N}, \end{aligned} \tag{70}$$

which corresponds to the result outlined in Eq. 23.

The decompositions are equivalently valid for the local definition of the measures. For example in the bivariate case, from Eq. 55

$$\begin{aligned} &M(X_1; X_2; ...; X_{i+1}; Y_1; Y_2; ...; Y_{i+1}) \\ &- M(X_1; X_2; ...; X_i; Y_1; Y_2; ...; Y_i) \\ &= I(X_{i+1}; X^i) + I(Y_{i+1}; Y^i) + I(X^{i+1}; Y^{i+1}) - I(X^i; Y^i), \end{aligned} \tag{71}$$

and given Eq. 36

$$\begin{aligned} &M(X_1; X_2; ...; X_{i+1}; Y_1; Y_2; ...; Y_{i+1}) \\ &- M(X_1; X_2; ...; X_i; Y_1; Y_2; ...; Y_i) \\ &= I(X_{i+1}; X^i, Y^i) + I(Y_{i+1}; X^i, Y^i) + T_{Y \cdot X}(i). \end{aligned} \tag{72}$$

The rates are equivalent also when obtained from the average or local global multi-information

$$\begin{aligned} &\lim_{i \to \infty} M(X_1; X_2; ...; X_{i+1}; Y_1; Y_2; ...; Y_{i+1}) \\ &- M(X_1; X_2; ...; X_i; Y_1; Y_2; ...; Y_i) \\ &= \lim_{N \to \infty} \frac{M(X_1; X_2; ...; X_N; Y_1; Y_2; ...; Y_N)}{N}. \end{aligned} \tag{73}$$

Of course, the decomposition of the global multi-information in which we have focused is one among many. It is one in which in particular we use time information and the identity of the subsystems to separate different terms. The same flexibility exists when decomposing the system predictive information about a subsystem, as exemplified in Eq. 24. We also mentioned this when introducing Eqs. 25 and 27, and motivated the selection of the decompositions that we focus on because of the presence of the transfer entropy. For multivariate systems the number of decompositions containing transfer entropy terms increases combinatorially.

5. *The normalization of the transfer entropy*

The necessity to consider a reference to evaluate the value of the transfer entropy has been considered previously. In [16] it was proposed to use a normalized version of the transfer entropy when comparing the strength of the influence between the regions:

$$NT_{Y \to X} = \frac{T_{Y \to X}}{H(X_{i+1}|X^i)} = 1 - \frac{H(X_{i+1}|X^i, Y^i)}{H(X_{i+1}|X^i)}. \tag{74}$$

This measure has been used when applying the transfer entropy to study the effective connectivity between brain regions [e. g. 57]. It is normalized so that a zero value is obtained when the transfer entropy is zero and a value of one is obtained if the past of $\mathcal{Y}$ can account for all the uncertainty about X_{i+1}. This type of normalized measures cannot be integrated in a framework like we did above for the unnormalized ones. In fact the necessity of a relative measure simply reflects that unless one evaluates the transfer entropy considering more in detail other properties of the processes the value obtained is not very informative. However, once one is aware of how the transfer entropy terms appear in the decomposition of the mutual information between subsystems, of the system predictive information about a subsystem, or of the subsystems and global multi-information, it is clear that the relative measure proposed in [16] is not unique and is not necessarily the more relevant depending on the aspects one is interested in. For example, it may be relevant to know the relative contribution of $T_{Y \to X}$ to the mutual information between the processes, or to the total entropy reduction of $\mathcal{X}$, or to the global multi-information.

In general, a proper normalization like the one of Eq. 74 has the form

$$\frac{T_{Y \to X} - T_{Y \to X\ min}}{T_{Y \to X\ max} - T_{Y \to X\ min}}, \tag{75}$$

where $T_{Y \to X\ min}$ and $T_{Y \to X\ max}$ refer to the minimum and maximum attainable values, respectively. Selecting this minimum and maximum implies some assumptions about which are the configurations against which one compares. In Eq. 74 $T_{Y \to X\ max} = H(X_{i+1}|X^i)$ and $T_{Y \to X\ min} = 0$, which implies assuming that the changes in $T_{Y \to X}$ are produced only by changes in $H(X_{i+1}|X^i, Y^i)$, that can have values in the range $[0, H(X_{i+1}|X^i)]$. But alternatively one could consider that $H(X_{i+1}|X^i, Y^i)$ is fixed and $H(X_{i+1}|X^i)$ varies in the range $[H(X_{i+1}|X^i, Y^i), H(X_{i+1})]$. This will result in $T_{Y \to X\ max} = I(X_{i+1}; X^i, Y^i)$ and $T_{Y \to X\ min} =$

0. Or one can also consider that both $H(X_{i+1}|X^i)$ and $H(X_{i+1}|X^i, Y^i)$ can vary and then $T_{Y\to X\ max} = H(X_{i+1})$ and $T_{Y\to X\ min} = 0$. If any of these analysis is more meaningful will depend on the particular data analyzed. All of them are different ways to select and synthesize the information about the dynamic dependencies in the system, in contrast to the view of the transfer or normalized transfer entropy as distinct isolated measures.

6. Alternative routes of decomposition, the m-horizon measures

We stressed that transfer entropies appear in the decompositions depending on the particular order of conditioning of the dependencies chosen. We here discuss some alternative routes of decomposition in which genuine transfer entropies do not appear but that may be preferable for some analysis. This also contributes to contemplate transfer entropy as one over different alternatives. Transfer entropy appears in a decomposition in which time steps are added one by one. One can alternatively consider decompositions in steps of m and the decompositions of mutual information as well as the different multi-informations still hold. For example

$$\begin{aligned} &I(X^{i+m}; Y^{i+m}) - I(X^i; Y^i) \\ &= I(X_{i+m}, ..., X_{i+1}; Y^i|X^i) \\ &+ I(Y_{i+m}, ..., Y_{i+1}; X^i|Y^i) \\ &+ I(X_{i+m}, ..., X_{i+1}; Y_{i+m}, ..., Y_{i+1}|X^i, Y^i), \end{aligned} \tag{76}$$

so that the m-horizon transfer entropy is

$$T^{(m)}_{Y\to X}(i) = I(X_{i+m}, ..., X_{i+1}; Y^i|X^i) \tag{77}$$

and the m-horizon instantaneous dependence is

$$T^{(m)}_{X\cdot Y}(i) = I(X_{i+m}, ..., X_{i+1}; Y_{i+m}, ..., Y_{i+1}|X^i, Y^i). \tag{78}$$

For bivariate systems the m-horizon transfer entropy is still zero if and only if there are no causal connections from $\mathcal{Y}$ to $\mathcal{X}$. In multivariate systems even the conditional m-horizon transfer entropy cannot be used to infer the causal structure because it is sensitive to indirect connections occurring during the m-interval. Furthermore, notice that the m-horizon transfer entropy differs from the measure proposed in [58] where only the sample from $i + m$ is used instead of all the samples in the interval. Such other measure thus does not appear in a decomposition of the mutual information. The m-horizon instantaneous dependence is not supposed to be theoretically zero like the instantaneous dependence with $m = 1$ if instantaneous interactions do not exist. This is because it quantifies the extra dependence that arises in the m-interval which cannot be accounted by the joint past. In that sense it is a measure that can be useful when examined referenced to a particular event .

Analogous m-measures can be defined in the cumulative case. Consider time series of length mN:

$$\begin{aligned} I(X^{mN}; Y^{mN}) = T^{(m)}_{Y^{mN}\to X^{mN}} + T^{(m)}_{X^{mN}\to Y^{mN}} \\ + T^{(m)}_{X^{mN}\cdot Y^{mN}}, \end{aligned} \tag{79}$$

where the cumulative m-horizon transfer entropy is

$$\begin{aligned} &T^{(m)}_{Y^{mN}\to X^{mN}} \\ &\quad = \sum_{k=0}^{N-1} H(X_{mk+m}, ..., X_{mk+1}|X^{mk}) \\ &\quad - \sum_{k=0}^{N-1} H(X_{mk+m}, ..., X_{mk+1}|X^{mk}, Y^{mk}) \end{aligned} \tag{80}$$

and the cumulative m-horizon instantaneous dependence is

$$\begin{aligned} &T^{(m)}_{Y^{mN}\cdot X^{mN}} \\ &= \sum_{k=0}^{N-1} I(X_{mk+m}, ..., X_{mk+1}; Y_{mk+m}, ..., Y_{k+1}|X^{mk}, X^{mk}). \end{aligned} \tag{81}$$

These m-horizon measures may be convenient in several ways. In first place one can expect them to be more robust to the selection of a particular sampling frequency and to the problem of time aggregation, which is a problem to take into account when applying the measures to temporally continuous signals [59]. Second, m can be selected in order to cover some intrinsically significant cycle of the dynamics, so that these measures can be more informative about the dependencies related to certain time scales or rhythms [60, 61]. Furthermore, also the different multi-informations can be expressed in terms of m-horizon measures. In particular, in analogy to Eq. 62, such a decomposition involves only bivariate m-horizon transfer entropies based on bipartitions of the system, and thus the connection of zero m-horizon transfer entropies to the causal structure holds.

IV. DISCUSSION

In this work we have examined an information theoretic framework to analyze the dynamic dependencies in networks of interacting processes. In Sec. III A we motivated our proposal asking for how fundamental it is the decomposition of the mutual information into causal influences components, and we presented the framework to analyze dynamic dependencies in multivariate systems for the case of stationary processes. The core quantities of this framework are a set of measures to characterize the statistical structure of a multivariate system. In particular, *subsystems multi-information* quantifies the dynamic dependencies of a set of subsystems which are *a priori* assumed to be distinguishable, based on the anatomical structure or functionality of the system. On the other hand *global multi-information* considers the system as some unique joint dynamics that arise from all the interactions between the parts.

In Sec. III B we used a graphical approach to understand how both causal and noncausal sources of dependence are quantified by the different components appearing in alternative decompositions of the mutual information between the future of one subsystem and the past of the whole system. One of these components is the transfer entropy [3], which has been extensively used to analyze the interactions between processes and in particular the functional or effective connectivity between brain regions. We illustrated how the effect of the different sources of dependence is not separable, and in particular that the transfer entropy cannot be used as a measure of the strength of the connections. In Sec. III C we examined some simple analytically solvable examples of bivariate processes to further illustrate the relation of the transfer entropy with the other measures appearing in the decompositions. Using analytical examples we avoided the difficulties one has to deal with when studying experimental data, namely the estimation of the probability distributions from finite data samples. In this way we could examine the sensitivity and specificity of the measures without any effect of estimation biases. This first part of our results indicates how transfer entropy can reflect changes in the internal properties of the subsystems and not only in the interactions among them, and the necessity to consider it into a wider framework that characterizes the dynamic dependencies in the system. The advantage of considering the measures as parts of a framework is not particular for the transfer entropy, but holds for any other of the measures that are part of it.

In the second part of our results (Sec. III D) we developed the complete framework to characterize the statistical structure of a multivariate system. We provided a local and a cumulative non-stationary formulation of the framework and showed how they converge for stationary processes. We derived decompositions of the multi-information measures into components that comprise transfer entropy terms, and we indicated how they generalize the decomposition of the mutual information previously considered in [6, 7, 28, 30, 31]. In Sec. III D 5 we discussed how the fact that the transfer entropy appears in different decompositions leads to different possible normalizations. We argued that these normalized measures are just synthetic ways to compare some different components of the framework, but again they should not be used in isolation. Finally in Sec. III D 6 we show that the decomposition of the mutual information into transfer entropies is just one of a more generic type of decompositions which consider the dynamic dependencies in intervals of m sampling times instead of for each sampling time.

Granger causality measures and in particular transfer entropy have been often used for inferring causal connections. However, there are strong limitations for their applicability to infer causality [see 8, 25, and references therein]. Briefly, one has to assume that it is possible to identify a set of processes which is complete in the sense that there are not hidden interacting processes. This assumption is certainly difficult to be fulfilled in application to experimental data. In neural applications, one cannot expect to record from a complete set and furthermore there is not a clear separation between different processes, for example when considering the signals recorded at close electrodes or voxels. Several refinements or extensions of the Granger causality measures have been proposed to deal with these problems [33, 62, 63], but the problem can at best be attenuated. Furthermore even if the assumptions were fulfilled, transfer entropy cannot be used to infer the causal connections in the case of deterministic or synchronized dynamics. This limitation is well-known [e. g. 37, 38, 64] and it is just the extreme manifestation of the nonmonotonicity with the strength of the causal connections that results from the inappropriateness of transfer entropy as a measure of connectivity strength (Section III C). This limitation for causal inference is due to the specific formulation of the Granger causality criterion using the conditioning on the own past, and not to the difficulty common to all the information theoretic measures (based on probability distributions) to be applied to deterministic systems (for example, for deterministic chaotical dynamical systems, the quantities can be calculated from the density of the trajectories in the manifolds [e. g. 65]).

Apart from its use for causal inference, Granger causality measures have been used to study asymmetric dynamic dependencies between brain dynamics [e. g. 18–20, 57, 66, 67]. Using the term *influence* here in a general way that may vary depending on the study (e. g. information transfer, information flow, causal influence), it is common that one is interested in comparing the strength of different influences. For example, the influences in opposite directions between two regions, the influences between different regions [e. g. 20, 66], or the influences across different frequency bands [e. g. 19, 29]. One can also compare the influences across different setups, like during the performance of different tasks [e. g. 20]. The interpretation of these comparisons relies on the sensitivity and specificity of the measures to the different interactions between the processes. We have shown that this interpretation is not easy since the transfer entropy is sensitive to causal and noncausal dependencies, and reflects internal properties of the processes apart from the interactions between them.

Considering that the transfer entropy is not specifically quantifying single interactions or sources of dependence it is important to determine what it quantifies. What does 'information transfer' (or information flow) mean? We discussed that the criterion of Granger causality defines *causality* operationally in terms of conditional independence. In the same sense *information transfer* is defined operationally from the definition of the measure transfer entropy. In that sense the information transfer is just what the transfer entropy quantifies, a reduction of entropy given the past of the other process. Only in particular settings, like for example an information channel with feedback, the transfer entropy can be understood as the error-free information rate attainable in the channel [44, 45], but to our knowledge the extension of this interpretation in terms of an attainable information rate remains an open question for networks of interacting processes.

In the case of the brain, as emphasized in [68], the presence of statistical dependencies *per se* is not very surprising given the ubiquitous existence of reciprocal connections in the brain. A measure of functional connectivity is only useful if its changes from one condition or moment to another allow us to understand how the system works. The question is then: When changes in a map of functional connectivity based on the transfer entropy (or another Granger causality measure) are analyzed, instead of the ones on a map based for example on mutual information, what extra information can we obtain?

If we accept that the notion of information transfer (information flow) is operationally defined, then expressions often used such as 'transfer entropy quantifies the information flow...' are clearly tautological. Furthermore, the sensitivity to dynamic dependencies is not specific of the transfer entropy and can be also obtained by the mutual information if not restricted to univariate variables [e. g. 22], that is, when it is calculated in a cumulative way for the whole processes as considered in Sec. III D. Directionality or asymmetry, which is considered as another useful property, is only an advantage if one is capable to give a physical meaning to difference of the values in opposite directions. Therefore, when evaluating what new insights about the system are provided by an analysis based on transfer entropy (or any Granger causality measure) with respect to mutual information, one cannot interpret the results based on *causal* or *information transfer* taken as something that transcends the definition of the measures.

The applicability of transfer entropy as a measure of functional connectivity can be especially delicate in the case when it is used to study the network properties based on graphs theory measures [52, 69]. Since the selection of the connections usually implies using a threshold to go from a weighted to an unweighted graph [22, 70], the nonmonotonicity of the transfer entropy with the strength of the connections may obstruct the comparison of the properties of the anatomical and functional networks [71, 72]. Despite these concerns, some success has been made in reconstructing known even complex biologically plausible structural networks from the networks based on transfer entropy analysis [e. g. 22, 73, 74]. This may be explained because of the presence of only weak degrees of coupling in the dynamics studied, and future work is needed to find general criteria for the applicability of Granger causality measures for structural network reconstruction.

The usefulness of the transfer entropy is more clear when considered as part of the whole framework to study dynamical dependencies. The changes of the components of one decomposition are better understood examining the changes in the other components, as illustrated in Secs. III B and III C. We have seen that the multi-information measures that characterize the statistical structure of the system can be decomposed either in a temporally asymmetric decomposition which includes transfer entropy terms or in an alternative decomposition based on a hierarchical clustering of the subsystems, which only includes mutual information terms.

The framework here presented has a great flexibility in the sense that it integrates measures to analyze the global, temporal, and subsystems statistical structure of the system. It extends previous measures of integration [34] by explicitly considering the temporal dynamics. It can be easily extended to consider measures of complexity and degeneracy [34, 75] to study the interplay of functional segregation and integration. The decomposition of the subsystems multi-information as a sum of transfer entropies also helps to further understand the relation examined in [76] between the measure of integrated information introduced in [34], and *causal density* [77], which have been both used to characterize consciousness levels. Furthermore, the framework can also be applied to consider the integration of functionally relevant modules [78] instead of elementary subsystems. Following [79], the multi-information measures explicitly considering the dynamics could be used to analyze the relevant motifs in the microscopic causal structure (Section II B) leading to dynamic complexity in a system.

Some of the measures that appear in the alternative decompositions of the mutual information between the future of a subsystem and the past of the whole system (Section III B) have been previously studied. In [80] it was defined a measure of G-autonomy which generalization as an information-theoretic measure is one of the measures in these decompositions. In that sense, the discussion about the specificity of the transfer entropy is also relevant to understand what a measure of conditional dependence can tell us or not about the mechanistic autonomy of a system. Furthermore, some of these decompositions were studied in [81] considering together what is called the *self storage* and the transfer entropy in boolean networks. In a series of studies [56, 81–83] it has been shown how these measures can be informative about the computations in different networks and how the joint examination of the measures and the entropies that compound them helps to better analyze their relation to particular spatiotemporal structures, or to structural properties in the network. Apart from more widely considering measures of statistical structure of multivariate systems, our analysis adds to this previous work considering also decompositions in which the order of conditioning of the dependencies does not give priority to the own past. Removing this restriction, and comparing the decompositions using causal graphs, helps to better appreciate the entanglement of the measures and their specificity.

Despite the potential value of this framework, the problems of estimation are a serious impediment for its application to experimental data, as they are already for the application of conditional transfer entropy. Since the measures in the framework are information-theoretic quantities, the standard techniques for estimation [84] and bias reduction [85] can be applied. Especially in multivariate systems it is necessary to sample a high dimensional space that comprises the past of all the recorded signals. In practice a good sampling of the probability distributions cannot be achieved. In the majority of the applications of Granger causality measures the bivariate measures are used [e. g. 18–20, 57, 66, 67, 86, 87]. Even for the bivariate measures it is still difficult to estimate the probability distributions needed to account for the whole past of the systems. The past is often reduced to one or a few samples at particular time lags [e. g. 57]. However, in general it is more usual to apply linear Gaussian Granger causality, and model the data with bivariate or sometimes multivariate autoregressive models [e. g. 19, 20, 67, 88]. As part of the estimation one has to assess the significance of nonzero values. For linear Granger causality measures a significance level can be assigned analytically for Gaussian processes [28, 89]. This analytical approach can be extended to the other measures in the framework here proposed taking into account that the information-theoretic formulation subsumes the formulation of Geweke for Gaussian processes [46, 55]. However, in general, some surrogate data are necessary to obtain a significance level [e. g. 20, 57].

In practice, another issue concerns the selection of which decompositions can be more relevant for the analysis of some particular system, since the number of decompositions grows combinatorially with the number of subsystems studied. As we mentioned in Section III A, one should rely on side information to identify potentially informative decompositions. For the brain this information can be based on anatomical information, or on the function each part is expected to perform [see 52, and references therein]. Another strategy that can be helpful is to determine the order of conditioning based on some criterion like selecting first those variables that are more informative. Along this line, an algorithm based on mutual information maximization is described in [90] to choose the most important variables in which to condition when calculating conditional transfer entropies.

Apart from the application to experimental data the framework can be more accessibly applied to characterize models fitted to data, from which one can generate larger amounts of samples. In [24] it was proposed to consider the data-driven Granger causality analysis as an exploratory approach, which can provide useful information, complementary to the information from structural connectivity, to reduce the space of possible models compatible with the data. Subsequently, a hypothesis-driven confirmatory analysis would consist in the fitting and comparison of some model of effective connectivity like the ones proposed in Dynamical Causal Modeling (DCM) [91, 92]. In these models the causal structure is strongly constrained *a priori*, and the fitting determines the strength of the nonzero connections. However, this use of the Granger causality measures for an exploratory analysis is significantly impaired by the strong assumptions for their applicability for causal inference and how if bivariate measures are used indirect connections and common drivers lead to positive values indicating false positive connections.

Therefore, we think that the framework here presented can be more useful not as a previous exploratory approach to determine the model, but as a posterior tool for the analysis of the dynamics generated by a model and the comparison of different models. The analysis of the relation between anatomical structure and dynamics could profit from replacing particular measures of functional connectivity by a set of interconnected descriptors of the dynamic dependencies. There is a vast line of research applying biologically-inspired models to learn about the relation between structural connectivity and neural dynamics [see 93, 94, for a review]. Transfer entropy has been applied also to these type of models as a measure of information flow or causal influence [e. g. 22, 95–97]. This dynamic analysis contrasts with the focus in the analysis of DCM models on evaluating the changes in the parameters related to effective connectivity [91, 92, 98]. But also for DCM models the analysis of the parameters of effective connectivity can be complemented by the analysis of the dynamics generated by the model. The framework can be useful to investigate the capacity of the model to reproduce specific dependencies found in the experimental data. Of course, the fact that the model has been fitted to the data assures that those properties captured by the fitting criteria are well reproduced. However these criteria may not capture other dynamic dependencies. More ambitiously, assuming that the model is good enough, one can use the framework here proposed to predict changes in the properties of the dynamics caused for example by a change in effective connectivity.

Altogether we believe that the framework we have presented contributes to provide a solid theoretical framework for data analysis of dynamical dependencies in networks of interacting processes. Despite that this framework is not specific for studying brain dynamics and can be applied to any complex system, the increasing amount of research about neural functional and effective connectivity renders this framework especially relevant for neuroscience. In the same way that recent efforts aim to provide an integrative perspective of the different modeling approaches of brain dynamics [8, 99], it is necessary to develop the adequate tools to analyze the data generated from these models.

V. ACKNOWLEDGMENTS

This work was supported by the European Union Seventh Framework Programme (FP7/2007- 2013) under grant agreement no. 269921 (BrainScaleS). DC is supported by the Italian Institute of Technology (IIT). DC thanks Stefano Panzeri and Ariadna Soy for critically reading a draft of this manuscript. DC and AL thank the anonymous reviewers for their useful comments that helped to improve the quality and readability of this work.

[1] C. W. J. Granger, Econometrica **37**(3), 424 (1969).
[2] C. W. J. Granger, J. Econ. Dynamics and Control **2**(1), 329 (1980).
[3] T. Schreiber, Phys. Rev. Lett. **85**, 461 (2000).
[4] C. W. J. Granger, Information and Control **6**, 28 (1963).
[5] N. Wiener, in *Modern Mathematics for Engineers* (McGraw-Hill, New York, 1956).
[6] H. Marko, IEEE T. Commun. **12**, 1345 (1973).
[7] J. Rissanen and M. Wax, IEEE Transactions on information theory **33**(4), 598 (1987).
[8] P. Valdes-Sosa, A. Roebroeck, J. Daunizeau, and K. Friston, Neuroimage **58**(2), 339 (2011).
[9] O. Kwon and J. S. Yang, Europhys. Lett. **82**(6), 68003 (2008).
[10] J. M. Nichols, M. Seaver, S. T. Trickey, M. D. Todd, C. Olson, and L. Overbey, Phys. Rev. E **72**(4), 046217 (2005).
[11] R. Kleeman, J. Atmos. Sci. **64**(3), 1005 (2007).
[12] M. Bauer, J. Cox, M. Caveness, J. Downs, and N. Thornhill, IEEE Trans. Control Syst. Techn. **15**(1), 12 (2007).
[13] E. Pereda, R. Quian Quiroga, and J. Bhattacharya, Prog. Neurobiol. **77**, 1 (2005).
[14] B. Gourevitch, R. Le Bouquin-Jeannes, and G. Faucon, Biol. Cybern. **95**(4), 349 (2006).
[15] S. L. Bressler and A. K. Seth, Neuroimage **58**(2), 323 (2011).
[16] B. Gourevitch and J. J. Eggermont, J. Neurophysiol. **97**(3),

2533 (2007).
[17] M. Chávez, J. Martinerie, and M. Le Van Quyen, J. Neurosci. Meth. **124**(2), 113 (2003).
[18] R. Vicente, M. Wibral, M. Lindner, and G. Pipa, J. Comput. Neurosci. **30**, 45 (2010).
[19] A. Brovelli, M. Ding, A. Ledberg, Y. Chen, R. Nakamura, and S. L. Bressler, P. Natl. Acad. Sci. USA **101**, 9849 (2004).
[20] A. Roebroeck, E. Formisano, and R. Goebel, Neuroimage **25**(1), 230 (2005).
[21] C. Gourieroux, A. Monfort, and E. Renault, Annales d'Economie et de Statistique **6-7**, 470 (1987).
[22] C. Honey, R. Kötter, M. Breakspear, and O. Sporns, Proc. Natl. Acad. Sci. USA **104**(24), 10240 (2007).
[23] K. Friston, PLoS Biol. **7**(2), e1000033 (2009).
[24] A. Roebroeck, E. Formisano, and R. Goebel, NeuroImage **58**(2), 296 (2011).
[25] D. Chicharro and A. Ledberg, PLoS One **7**(3), e32466 (2012).
[26] J. Pearl, *Causality: Models, Reasoning, Inference* (Cambridge University Press, New York, 2009), 2nd ed.
[27] A. Sillito, J. Cudeiro, and H. Jones, Trends Neurosci. **29**(6) (2006).
[28] J. F. Geweke, J. Am. Stat. Assoc. **77**(378), 304 (1982).
[29] M. Ding, Y. Chen, and S. L. Bressler, in *Handbook of Time Series Analysis: Recent Theoretical Developments and Applications.* (Wiley-VCH Verlag, 2006), pp. 437–460.
[30] V. Solo, in *Proceedings of the 47th IEEE Conference on Decision and Control* (2008), pp. 4639–4944.
[31] P. O. Amblard and O. Michel, J. Comput. Neurosci. **30**, 7 (2011).
[32] J. F. Geweke, J. Am. Stat. Assoc. **79**(388), 907 (1984).
[33] A. B. Barrett, L. Barnett, and A. K. Seth, Phys. Rev. E **81**(4), 041907 (2010).
[34] G. Tononi, O. Sporns, and G. Edelman, Proc. Natl. Acad. Sci. USA **91**(11), 5033 (1994).
[35] S. Watanabe, IBM Journal of Research and Development **4**, 66 (1960).
[36] M. Studený and J. Vejnarová, in M. Jordan, ed., *Learning in Graphical models* (Kluwer Academic Publishers, Norwell, MA, USA, 1998), pp. 261–298.
[37] D. Marinazzo, M. Pellicoro, and S. Stramaglia, Phys. Rev. Lett. **100**(14), 144103 (2008).
[38] M. Palus and M. Vejmelka, Phys. Rev. E **75**(5), 056211 (2007).
[39] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (John Wiley and Sons, 2006), 2nd ed.
[40] S. Kullback, *Information Theory and Statistics* (Dover, Mineola,NY, 1959/1997).
[41] C. Hiemstra and J. D. Jones, J. Financ. **49**(5), 1639 (1994).
[42] M. G. Rosenblum and A. S. Pikovsky, Phys. Rev. E **64**, 045202(R) (2001).
[43] N. Ancona, D. Marinazzo, and S. Stramaglia, Phys. Rev. E **70**(5), 056221 (2004).
[44] G. Kramers, PhD dissertation, Swiss Federal Institute of Technology, Zurich (1998).
[45] H. Permuter, Y. Kim, and T. Weissman, IEEE Trans. Inf. Theory **57**(3), 3248 (2009).
[46] L. Barnett, A. B. Barrett, and A. K. Seth, Phys. Rev. Lett. **103**(23), 238701 (2009).
[47] R. Dahlhaus, Metrika **51**, 157 (2000).
[48] M. Eichler, Phil. Trans. R. Soc. B. **360**, 953 (2005).
[49] M. Eichler, J. Econometrics **137**, 334 (2007).
[50] H. Nalatore, M. Ding, and G. Rangarajan, Phys. Rev. E **75**(3), 031123 (2007).
[51] G. Deshpande, K. Sathian, and X. Hu, Neuroimage **52**(3), 884 (2010).
[52] O. Sporns, in H. Johansen-Berg and T. Behrens, eds., *Diffusion MRI: From quantitative measurement to in vivo neuroanatomy* (Academic Press, Amsterdam, 2009), pp. 309–332.
[53] A. Kaiser and T. Schreiber, Physica D **166**, 43 (2002).
[54] L. Barnett and A. K. Seth, J. Neurosci. Meth. **201**(2), 404 (2011).
[55] D. Chicharro, Biol. Cybern. **105**(5-6), 331 (2011).
[56] J. T. Lizier, M. Prokopenko, and A. Y. Zomaya, Phys. Rev. E **77**(2), 026110 (2008).
[57] M. Besserve, B. Schoelkopf, N. K. Logothetis, and S. Panzeri, J. Comput. Neurosci. **29**(3), 547 (2010).
[58] J. Dufour and E. Renault, Econometrica **66**(5), 1099 (1998).
[59] K. Friston, Neuroimage **58**(2), 303 (2011).
[60] S. Panzeri, N. Brunel, N. K. Logothetis, and C. Kayser, Trends Neurosci. **33**(3), 111 (2010).
[61] M. Siegel, T. Donner, and A. Engel, Nature Rev. Neurosci. **13**, 121 (2012).
[62] S. Guo, A. K. Seth, K. M. Kendrick, C. Zhou, and J. Feng, J. Neurosci. Meth. **172**(1), 79 (2008).
[63] S. Guo, J. Wu, M. Ding, and J. Feng, PLoS Comput. Biol. **4**(5) (2008).
[64] N. Ay and D. Polani, Advances in complex systems **11**, 17 (2008).
[65] M. Lungarella, K. Ishiguro, Y. Kuniyoshi, and N. Otsu, Int. J. Bifurcat. Chaos **17**(3), 903 (2007).
[66] S. L. Bressler, W. Tang, C. M. Sylvester, G. L. Shulman, and M. Corbetta, J. Neurosci. **28**(40), 10056 (2008).
[67] C. Kayser and N. K. Logothetis, Frontiers in Integrative Neuroscience **3**, article 7 (2009).
[68] E. Salinas and T. Sejnowski, Nature Rev. Neurosci. **2**, 539 (2001).
[69] M. Rubinov and O. Sporns, NeuroImage **52**(3), 1059 (2010).
[70] E. Bullmore and O. Sporns, Nature Rev. Neurosci. **10**(3), 186 (2009).
[71] O. Sporns, G. Tononi, and G. Edelman, Neural Networks **13**(8-9), 909 (2000).
[72] C. J. Honey, J.-P. Thivierge, and O. Sporns, Neuroimage **52**(3), 766 (2010).
[73] D. Marinazzo, M. Pellicoro, and S. Stramaglia, Phys. Rev. E **77**, 056215 (2008).
[74] O. Stetter, D. Battaglia, J. Soriano, and T. Geisel, arXiv:1201.0732v3 [q-bio.NC] (2012).
[75] G. Tononi, O. Sporns, and G. Edelman, Proc. Natl. Acad. Sci. USA **96**(6), 3257 (1999).
[76] A. K. Seth, A. B. Barrett, and L. Barnett, Phil. Trans. R. Soc. A. **369**(1952), 3748 (2011).
[77] A. K. Seth, E. Izhikevich, G. N. Reeke, and G. M. Edelman, P. Natl. Acad. Sci. USA **103**(28), 10799 (2006).
[78] G. Zamora-López, C. Zhou, and J. Kurths, Front. Neurosci. **5**(83), 1 (2011).
[79] L. Barnett, C. L. Buckley, and S. Bullock, Phys. Rev. E **83**, 041906 (2011).
[80] A. K. Seth, Artificial life **16**(2), 179 (2010).
[81] J. T. Lizier, M. Prokopenko, and A. Y. Zomaya, Chaos **20**(3), 037109 (2010).
[82] J. T. Lizier and M. Prokopenko, Eur. Phys. J. B. **73**, 605 (2010).
[83] J. T. Lizier, S. Pritam, and M. Prokopenko, Artificial Life **17**(4), 293 (2011).
[84] K. Hlaváčkova-Schindler, M. Paluš, M. Vejmelka, and J. Bhattacharya, Phys. Rep. **441**, 1 (2007).
[85] S. Panzeri, R. Senatore, M. A. Montemurro, and R. S. Petersen, J. Neurophysiol. **98**(3), 1064 (2007).
[86] S. L. Bressler, C. G. Richter, Y. Chen, and M. Ding, Stat. Med. **26**(21), 3875 (2007).

[87] Z. Zhou, M. Ding, Y. Chen, P. Wright, Z. Lu, and Y. Liu, Brain Res. **1289**, 22 (2009).
[88] P. Valdés-Sosa, J. Sánchez-Bornot, A. Lage-Castellanos, M. Vega-Hernández, J. Bosch-Bayard, L. Melie-García, and E. Canales-Rodríguez, Phil. Trans. R. Soc. B. **360**(1457), 969 (2005).
[89] H. Lütkepohl, *New introduction to multiple time series analysis* (Springer-Verlag, Berlin, 2006).
[90] D. Marinazzo, M. Pellicoro, and S. Stramaglia, Comput. Math. Meth. Med. 303601 (2012).
[91] K. J. Friston, L. Harrison, and W. Penny, Neuroimage **19**(4), 1273 (2003).
[92] J. Daunizeau, O. David, and K. E. Stephan, NeuroImage **58**(2), 312 (2011).
[93] V. Jirsa, Neuroinformatics **2**, 183 (2004).
[94] M. Breakspear and V. Jirsa, in A. McIntosh and J. VK, eds., *Handbook of brain connectivity* (Springer, Berlin, 2007), pp. 3–64.
[95] M. Lungarella and O. Sporns, PLoS Comput. Biol. **2**, e144 (2006).
[96] A. K. Seth and G. Edelman, Neural Comput. **19**, 910 (2007).
[97] A. Buehlmann and G. Deco, PLoS Comput. Biol. **6**, e10000934 (2010).
[98] O. David, S. J. Kiebel, L. M. Harrison, J. Mattout, J. M. Kilner, and K. J. Friston, Neuroimage **30**, 1255 (2006).
[99] M. Breakspear, V. Jirsa, and G. Deco, Neuroimage **52**, 727 (2010).