

Topological Causality in Dynamical Systems

Daniel Harnack,^{*} Erik Laminski,[†] Maik Schünemann,[‡] and Klaus Richard Pawelzik[§]

University of Bremen, 28359 Bremen, Germany and Center for Cognitive Science (ZKW), 28359 Bremen, Germany

(Received 10 April 2017; revised manuscript received 8 June 2017; published 1 September 2017)

Determination of causal relations among observables is of fundamental interest in many fields dealing with complex systems. Since nonlinear systems generically behave as wholes, classical notions of causality assuming separability of subsystems often turn out inadequate. Still lacking is a mathematically transparent measure of the magnitude of effective causal influences in cyclic systems. For deterministic systems we found that the expansions of mappings among time-delay state space reconstructions from different observables not only reflect the directed coupling strengths, but also the dependency of effective influences on the system's temporally varying state. Estimation of the expansions from pairs of time series is straightforward and used to define novel causality indices. Mathematical and numerical analysis demonstrate that they reveal the asymmetry of causal influences including their time dependence, as well as provide measures for the effective strengths of causal links in complex systems.

DOI: 10.1103/PhysRevLett.119.098301

Introduction.—The concept of causality has a long history ranging back to the beginnings of natural philosophy [1]. In recent formalizations it refers to situations where states \mathbf{x}_1 of one system part influence states \mathbf{x}_2 of another part [2]. It is often assumed that some aspects of \mathbf{x}_1 vary independently of \mathbf{x}_2 , and that the flow of information in the overall system is essentially unidirectional. This premise is at odds with nonlinear dynamical systems studied in, e.g., physics, ecology, economy, and neuroscience: generally, two system parts, e.g., two brain areas, will have bidirectional interaction and cyclic information flow. The classical notion of causality becomes problematic here since cause and effect are entangled.

This entanglement is reflected in Takens' theorem [3,4], which proves that in generic smooth deterministic dynamical systems the overall state is reconstructable from any measured observable. In other words, if \mathbf{x}_1 and \mathbf{x}_2 interact bidirectionally, each $\mathbf{x}_1(t)$ and $\mathbf{x}_2(t)$ alone contain the full information about the whole system constituted by \mathbf{x}_1 and \mathbf{x}_2 . That is, the system cannot be separated into subsystems and rather behaves as a whole. In consequence, the question for causal relations cannot be answered by classifying component systems as cause or effect, but rather asks for the asymmetry and strength of influence among these components.

Here we present a mathematical definition of directed influence tailored to entangled dynamical systems. As a key insight we discovered that local distortions in the mappings between state-space reconstructions based on different observables reflect the time dependent efficacy of causal links among the underlying system components. Causality measures derived from this relation are analytically accessible for simple systems and for more complicated ones can be estimated in a model free, data driven manner.

Topological causality.—Our approach relies on extensions of Takens' theorem, which is reviewed shortly: Let the (multidimensional) states \mathbf{x}_1 and \mathbf{x}_2 of two system components be governed by

$$\begin{aligned}\dot{\mathbf{x}}_1 &= f_1(\mathbf{x}_1, w_{12}\mu_2(\mathbf{x}_2)), \\ \dot{\mathbf{x}}_2 &= f_2(\mathbf{x}_2, w_{21}\mu_1(\mathbf{x}_1)),\end{aligned}$$

where $\mu_i(\mathbf{x}_i)$ denote fixed scalar functions and w_{ij} coupling constants.

The trajectories $(\mathbf{x}_1(t), \mathbf{x}_2(t))$ form an invariant manifold in the phase space of the joint dynamical system. A manifold in a delay coordinate space visited by $\mathbf{r}_{\phi_i}^{\mathbf{x}_i}(t) = (\phi_i(\mathbf{x}_i(t)), \phi_i(\mathbf{x}_i(t + \tau)), \phi_i(\mathbf{x}_i(t + 2\tau)), \dots, \phi_i(\mathbf{x}_i(t + (m-1)\tau)))$ is topologically equivalent (a homeomorphic mapping between both manifolds exists) if $w_{ij} \neq 0$ and the embedding dimension m is sufficient. Here, ϕ_i is a measurement function depending on a scalar component of \mathbf{x}_i , which is omitted to simplify the notation ($\mathbf{r}^{\mathbf{x}_i} := \mathbf{r}_{\phi_i}^{\mathbf{x}_i}$). By transitivity, a unique mapping from $\mathbf{r}^{\mathbf{x}_i}$ to $\mathbf{r}^{\mathbf{x}_j}$, denoted by $M_{i \rightarrow j}^t$, exists iff $w_{ij} \neq 0$ [5].

Assuming that these mappings are differentiable, $M_{i \rightarrow j}^t$ denotes the local linearization (Jacobian matrix) of $M_{i \rightarrow j}$ at the reference point t : Given that $\{t_1^{\mathbf{x}_i}, \dots, t_k^{\mathbf{x}_i}\}$ are the time indices of the nearest neighbors on $\mathbf{r}^{\mathbf{x}_i}$ to the reference point $\mathbf{r}^{\mathbf{x}_i}(t)$, $M_{i \rightarrow j}^t$ is the linear approximation of the mapping that projects $\{\mathbf{r}^{\mathbf{x}_i}(t_1^{\mathbf{x}_i}), \dots, \mathbf{r}^{\mathbf{x}_i}(t_k^{\mathbf{x}_i})\}$ to $\{\mathbf{r}^{\mathbf{x}_j}(t_1^{\mathbf{x}_i}), \dots, \mathbf{r}^{\mathbf{x}_j}(t_k^{\mathbf{x}_i})\}$. In practice, we analyze the expansion $e_{i \rightarrow j}^t$ of $M_{i \rightarrow j}^t$, which is determined by the singular values $\sigma_k^t(M_{i \rightarrow j}^t)$ of $M_{i \rightarrow j}^t$ larger than 1,

$$e_{i \rightarrow j}^t = \prod_k \max(1, \sigma_k^t(M_{i \rightarrow j}^t)).$$

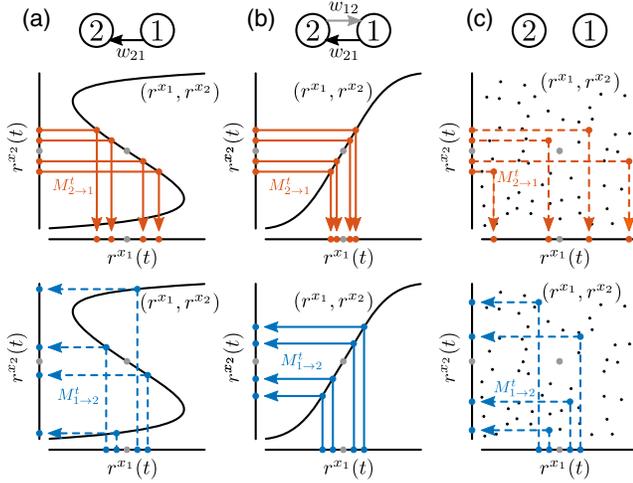


FIG. 1. The relation of points \mathbf{r}^{x_1} and \mathbf{r}^{x_2} on multidimensional manifolds illustrated in one dimension. The joint manifold represented by $(\mathbf{r}^{x_1}, \mathbf{r}^{x_2})$ can be interpreted as the function mediating the mappings $M_{i \rightarrow j}$ between both spaces, and local linearizations $M_{i \rightarrow j}^t$ of the mappings as the slope around a reference point. (a) When only $w_{21} \neq 0$, a one-to-one mapping $M_{2 \rightarrow 1}$ from \mathbf{r}^{x_2} to \mathbf{r}^{x_1} exists, but not in the reverse direction: $\mathbf{r}^{x_2}(t)$ is not uniquely determined for all states $\mathbf{r}^{x_1}(t)$. Locally, $M_{1 \rightarrow 2}^t$ can be attributed a diverging expansion property, since close neighbors of a given point $\mathbf{r}^{x_1}(t)$ correspond to distant parts of the joint density $(\mathbf{r}^{x_1}, \mathbf{r}^{x_2})$. The dashed lines visualize nonuniqueness. (b) Here, both couplings are nonzero, but $w_{21} > w_{12}$. Larger independence of x_1 implies a stronger expansion by $M_{1 \rightarrow 2}^t$ than by $M_{2 \rightarrow 1}^t$ at most reference points, which is indicated by the higher slope of $(\mathbf{r}^{x_1}, \mathbf{r}^{x_2})$ when seen from \mathbf{r}^{x_1} . (c) If no coupling exists, expansion diverges in both directions.

To illustrate how the expansions of these mappings between reconstructions relate to directed effective influence, consider the following thought experiment: First, a system with unidirectional interaction is observed, i.e., $w_{21} \neq 0$; $w_{12} = 0$. By virtue of Takens' theorem, a unique mapping $M_{2 \rightarrow 1}$ from reconstruction \mathbf{r}^{x_2} to \mathbf{r}^{x_1} exists. However, $M_{1 \rightarrow 2}$ does not exist, since \mathbf{x}_1 has no information on \mathbf{x}_2 . This is illustrated in Fig. 1(a) by a joint manifold $(\mathbf{r}^{x_1}, \mathbf{r}^{x_2})$ lying “folded” over \mathbf{r}^{x_1} but uniquely over \mathbf{r}^{x_2} . Note here that somewhat counterintuitively the influence from \mathbf{x}_1 to \mathbf{x}_2 is reflected in the “backward” mapping $M_{2 \rightarrow 1}$: the existence of $M_{2 \rightarrow 1}$ implies coupling from \mathbf{x}_1 to \mathbf{x}_2 .

Now increasing w_{12} while keeping $w_{21} > w_{12}$ leads to mutual but asymmetric interaction. Reconstructions \mathbf{r}^{x_1} and \mathbf{r}^{x_2} both reveal the same global system state. However, the weaker coupling from \mathbf{x}_2 to \mathbf{x}_1 implies that the region of \mathbf{r}^{x_2} states consistent with a small region of \mathbf{r}^{x_1} states around $\mathbf{r}^{x_1}(t)$ is larger than vice versa at most reference points t : Both \mathbf{r}^{x_1} and \mathbf{r}^{x_2} are driven away from their state at time t by a combination of internal dynamics and the external influence from the other variable, but \mathbf{r}^{x_2} is more so due to the stronger coupling w_{21} . This entails that $e_{1 \rightarrow 2}^t > e_{2 \rightarrow 1}^t$, and a joint manifold $(\mathbf{r}^{x_1}, \mathbf{r}^{x_2})$ lying uniquely over both

reconstruction spaces, but more “steeply” over \mathbf{r}^{x_1} [Fig. 1(b)].

If w_{12} is now decreased again to approach 0, $e_{1 \rightarrow 2}^t$ increases until it diverges at the point where $(\mathbf{r}^{x_1}, \mathbf{r}^{x_2})$ folds in on itself as seen from \mathbf{r}^{x_1} [Fig. 1(a)]. This happens at $w_{12} = 0$, where the map $M_{1 \rightarrow 2}^t$ loses uniqueness and corresponding points to neighbors in \mathbf{r}^{x_1} lie scattered over the whole dynamical range of \mathbf{r}^{x_2} . Thus we equate infinite expansion to the nonexistence of the corresponding mapping.

Consequently, when the couplings among x_1 and x_2 vanish altogether, both component systems behave independently and the density of the resulting joint manifold factorizes. When observed from reference states $\mathbf{r}^{x_1}(t)$ and $\mathbf{r}^{x_2}(t)$, the mappings can be considered infinitely expanding in both directions, since for most reference points close neighbors correspond to distant points in the respective other space [Fig. 1(c)].

In these expositions we assumed that the scalar observables $x_i = \phi_i(\mathbf{x}_i)$ have been transformed to their quantiles $q(x_i) = F(x_i)$ prior to time-delay embedding, where $F(x_i) = P[X_i \leq x_i]$ is the cumulative density function for the invariant measure of x_i . This eliminates expansions not caused by directed influences but arising from the numerical representation of the scalar time series [[6] (A)]. In the following we routinely perform quantile transformations by kernel density estimations of the invariant densities $p(x_i)$ [7].

Following these topological considerations, we hypothesize that local expansions of the mappings between reconstruction manifolds of two observables can be utilized for graded measures of the directed causal influences between component systems represented by these observables, where $e_{i \rightarrow j}^t$ is inversely related to the strength of the causal influence $j \rightarrow i$.

The relation between expansion and causal influence can be made fully explicit with coupled one-dimensional time-discrete maps for which two-dimensional time-delay embeddings are sufficient. Consider the system given by

$$\begin{aligned} x_1(t+1) &= x_1(t)[3.8(1-x_1(t)) - w_{12}x_2(t)], \\ x_2(t+1) &= x_2(t)[3.6(1-x_2(t)) - w_{21}x_1(t)], \end{aligned} \quad (1)$$

which has served as a model of ecological systems [8]. Here, $M_{i \rightarrow j}^t$ based on raw data is accessible analytically for every point in state space. For weak coupling weights $e_{i \rightarrow j}^t$ is dominated by $1/|x_i(t)w_{ij}|$ [[6] (B)]. This example illustrates two important points. First, the weights are inversely proportional to the respective expansions and $\lim_{w_{ij} \rightarrow 0} e_{i \rightarrow j}^t = \infty$, in line with the heuristic considerations. And second, the causal influence does not only depend on the coupling weights, but also on the system's state: $M_{1 \rightarrow 2}^t$ is strongly expansive for low x_1 values, and $M_{2 \rightarrow 1}^t$ for low values of x_2 , so that different regions in state space could be characterized by different causal dominance.

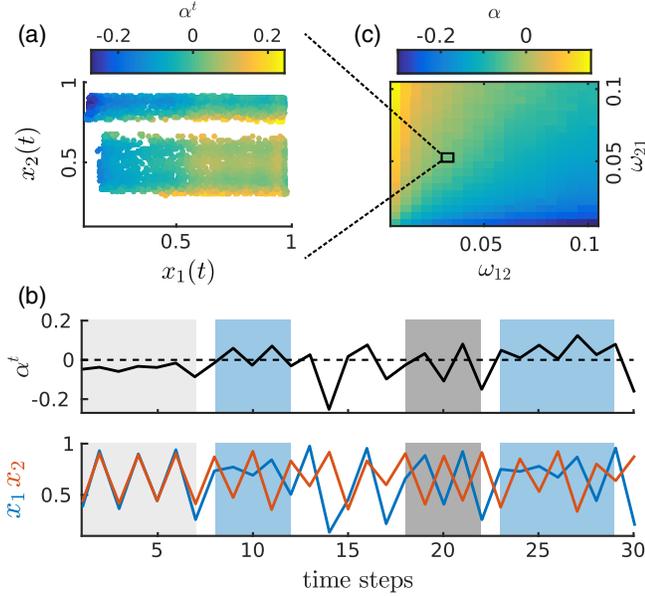


FIG. 2. (a) The state-space dependent α^t in a system given by Eqs. (1) is shown for $w_{21} = 0.05$, $w_{12} = 0.02$. (b) A segment of α^t and the corresponding time courses of x_1 and x_2 for the same coupling weights. Different regimes of dominant causal direction give rise to different dynamical motifs. If α^t is close to 0 for subsequent time points (light gray), x_1 and x_2 synchronize. If α^t varies strongly around 0 (dark gray), x_1 and x_2 desynchronize. When the causal influence from one variable to the other is dominant, here from 1 to 2 (blue), the trajectory of x_2 shows higher amplitude excursions than the one of x_1 . (c) The mean asymmetry index α for the same system with varying coupling strengths.

To measure such state dependent asymmetry of causal influence, we define an index $-1 \leq \alpha^t \leq 1$ as

$$\alpha^t = \frac{\log(e_{1 \rightarrow 2}^t) - \log(e_{2 \rightarrow 1}^t)}{\log(e_{1 \rightarrow 2}^t) + \log(e_{2 \rightarrow 1}^t)}.$$

This definition is motivated by the relation of the log expansions to loss of certainty in information theoretical terms [[6] (C)]. Note, however, that depending on the particular system and the purpose of analysis other choices can be more useful [[6] (G) and (H)]. Figure 2(a) shows that the asymmetry index α^t in this example fluctuates considerably over time as the system explores the state space. This change of causal dominance gives rise to various dynamical regimes among the time courses of \mathbf{r}^{x_1} and \mathbf{r}^{x_2} , which are also obvious to see in $x_1(t)$ and $x_2(t)$ proper since the dimensionality of the system is low [Fig. 2(b)]. Specifically, when, e.g., the influence from x_1 to x_2 is stronger than in the reverse direction (blue region), i.e., $e_{1 \rightarrow 2}^t > e_{2 \rightarrow 1}^t$, the trajectory of x_2 shows stronger fluctuations than the one of x_1 .

To define the mean asymmetry index we average over the states visited during the dynamics,

$$\alpha = \frac{\langle \log(e_{1 \rightarrow 2}^t) - \log(e_{2 \rightarrow 1}^t) \rangle_t}{\langle \log(e_{1 \rightarrow 2}^t) + \log(e_{2 \rightarrow 1}^t) \rangle_t}.$$

For the case of bidirectional coupling, enforcing $M_{1 \rightarrow 2}^t = (M_{2 \rightarrow 1}^t)^{-1}$, this simplifies to

$$\alpha = \frac{\langle \log(\det(M_{1 \rightarrow 2}^t (M_{1 \rightarrow 2}^t)^T)) \rangle_t}{2 \sum_k \langle |\log(\sigma_k^t(M_{1 \rightarrow 2}^t))| \rangle_t},$$

which underlines the dissimilarity to approaches using $\log(|\det(M)|)$ [[9,6 (C)]]. Figure 2(c) shows α for different combinations of coupling weights. The fact that $\alpha \neq 0$ for $w_{12} = w_{21}$ reflects the difference between the dynamical equations for x_1 and x_2 and highlights again that the expansion is not a mere proxy of the coupling weight, but actually measures the effective influence exerted along the causal link. However, the log determinant in α does not differentiate between the qualitatively different situations of balanced strong and balanced weak influence. For this purpose we define the topological causality (TC)

$$C_{i \rightarrow j}^t = \frac{1}{1 + \log(e_{j \rightarrow i}^t)},$$

$$C_{i \leftarrow j}^t = \frac{1}{1 + \langle \log(e_{j \rightarrow i}^t) \rangle_t}.$$

$C \in [0, 1]$ satisfies the following fundamental intuitions about causality: TC from component system i to j vanishes if no causal link exists ($w_{ji} = 0$), and for small couplings it is a monotonic function of the coupling weight. Also here alternative definitions with the same properties are possible. In general, C and C^t depend on the coupling weights as well as on the current state of the system.

In cases where the dynamical system model does not allow for an analytical linearization of the mappings between reconstructed spaces, or the model itself is not known, the local mappings and hence their expansion can be estimated in a purely data-driven manner. After optimal embedding parameters m and τ are chosen [10], one finds the time indices $\{t_i^{x_i}, \dots, t_k^{x_i}\}$ of the k nearest neighbors on \mathbf{r}^{x_i} to a reference point $\mathbf{r}^{x_i}(t)$. From these sets of points, expansions and chance levels can be estimated as elaborated in [6] (D). In [6] (E) we demonstrate the robustness of these estimations towards variations in analysis parameters and time series length. Notably, the direction of asymmetry can still be correctly identified for very short time series. Estimated quantities are marked by a $\hat{\cdot}$.

As an example of a more complex case that is not analytically tractable we adduce a system of coupled Rössler equations [11] described by

$$\dot{x}_i(t) = -f_i y_i(t) - z_i(t) + \sum_j \Omega_{ij},$$

$$\dot{y}_i(t) = f_i x_i(t) + 0.1 y_i(t),$$

$$\dot{z}_i(t) = 0.1 + z_i(t)[x_i(t) - 14]; \quad i = 1, \dots, n, \quad (2)$$

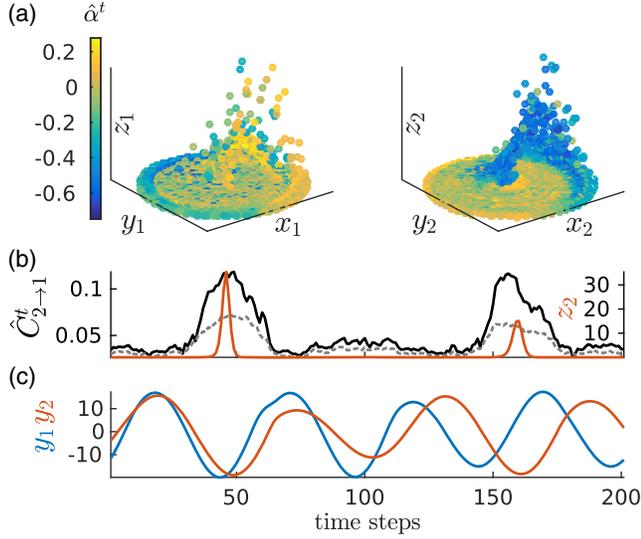


FIG. 3. Two bidirectionally coupled Rössler systems [Eqs. (2)] with $w_{21} = 0.2$, $w_{12} = 0.6$, and $\Omega_{ij} = w_{ij}z_j(t)$. The time series of 10^5 data points were embedded with dimension $m = 13$. $\hat{\alpha}^t$ and $\hat{C}_{2 \rightarrow 1}^t$ are shifted to be aligned with the temporal mean $t + 1/2(m-1)\tau$ of the corresponding reconstructions $r^{z_i}(t)$. (a) Local asymmetry $\hat{\alpha}^t$ of 10^4 points shown on projections of the attractor to each system. (b) $\hat{C}_{2 \rightarrow 1}^t$ (black) for 200 consecutive time steps and the corresponding time series of z_2 (orange). The gray dashed line marks chance level. (c) Time series y_1 and y_2 used to estimate $\hat{C}_{2 \rightarrow 1}^t$.

with coupling functions Ω_{ij} . If not stated otherwise, $\{f_1, f_2, f_3\}$ were set to $\{0.99, 0.85, 0.67\}$ and $q(y_i)$ were used as measurements from the individual systems. Figures 3(a) and 3(b) show causality measures for a bidirectionally coupled system ($n = 2$) with $\Omega_{ij} = w_{ij}z_j(t)$. When choosing the coupling function in this way, strong causal influence $i \rightarrow j$ is only expected if the driving z_i component deviates from 0. Both $\hat{\alpha}^t$ and $\hat{C}_{i \rightarrow j}^t$ capture this temporal structure, which is not obvious from the time courses of the used measurements y_1, y_2 [Fig. 3(c)].

In order to serve as satisfactory definitions, the proposed indices of causality must meet fundamental requirements that can be demonstrated by examining simple network motifs.

One prerequisite is transitivity, meaning that “if 1 causes 2 and 2 causes 3, then 1 causes 3.” Since $M_{3 \rightarrow 1}^t = M_{2 \rightarrow 1}^t M_{3 \rightarrow 2}^t$, it can be shown that

$$\begin{aligned} C_{1 \rightarrow 3}^t &\geq C_{1 \rightarrow 2}^t C_{2 \rightarrow 3}^t \quad \text{if } w_{21} \neq 0 \wedge w_{32} \neq 0 \\ C_{1 \rightarrow 3}^t &= 0 \quad \text{else,} \end{aligned}$$

meaning that transitivity is mathematically guaranteed. Figure 4(a) shows $\hat{C}_{1 \rightarrow 3}$ for a system of three coupled Rössler equations. Only w_{21} and w_{32} were varied and other coupling weights fixed to 0.

Another required property is the ability to distinguish shared input from true interaction. This is formally guaranteed if the receiving systems both retain independent

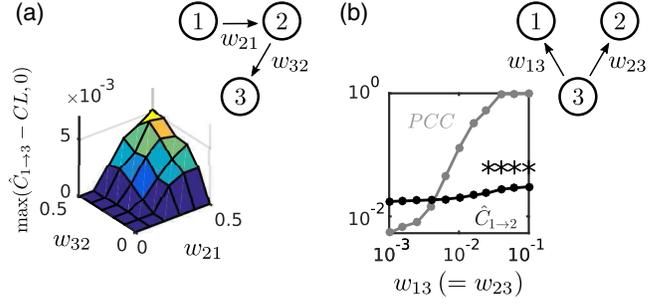


FIG. 4. (a) Transitivity. A unidirectionally coupled chain is realized by a system of Eqs. (2) ($n = 3$) with only $w_{21} \neq 0$ and $w_{32} \neq 0$ and $\Omega_{ij} = w_{ij}x_j(t)$. \hat{C} was averaged from 10^3 randomly selected data points from time series of 10^5 points with an embedding dimension $m = 13$. A bivariate increase of the excess of $\hat{C}_{1 \rightarrow 3}$ over chance level (CL) is observed, consistent with the theoretical prediction. (b) Common input investigated in a system of Eqs. (2) ($n = 3$) with $\Omega_{ij} = w_{ij}[x_j(t) - x_i(t)]$ and only $w_{13} = w_{23} \neq 0$. $\hat{C}_{i \rightarrow j}$ is estimated with an embedding dimension $m = 7$. $\hat{C}_{1 \rightarrow 2}$ and $\hat{C}_{2 \rightarrow 1}$ (not shown) depend weakly on the common input and only become significant (marked by *) in the presence of high redundancy between 1 and 2, signified by a high Pearson correlation coefficient (PCC).

degrees of freedom, i.e., do not synchronize completely. To show that also estimated TC reflects this property we consider a system described by Eqs. (2) ($n = 3$) with coupling functions $\Omega_{ij} = w_{ij}[x_j(t) - x_i(t)]$, where only $w_{13} \neq 0$ and $w_{23} \neq 0$, generating a divergent network motif. Here, $\{f_1, f_2, f_3\}$ were set to $\{0.99, 0.97, 0.98\}$. With increasing coupling from x_3 to x_1 and x_2 , the latter two synchronize more strongly, masking the actually absent interaction. Figure 4(b) shows that $\hat{C}_{1 \rightarrow 2}$ is nearly independent of the common drive, and becomes significant only in the presence of substantial correlations.

To demonstrate the applicability to real experimental time series we analyzed EEG data [12], for which a predominant information flow from frontal to dorsal channels was identified [13]. The same pattern is revealed by the asymmetry index $\hat{\alpha}$ [6] (F).

Discussion.—The concept of TC is based on the insight that expansions of mappings between time-delay state-space reconstructions from different observables systematically reflect effective, state-space dependent influences among parts of nonseparable deterministic systems.

Being tailored to such systems, which preserve information between coupled components, TC seems complementary to methods for determining causal influences in stochastic systems. The most prominent examples are Wiener-Granger causality (WGC) [14] and transfer entropy (TE) [15], which are conceptually related [16]. Both are based on the reduction of uncertainty in one time series by including past information from the other. However, the approaches of WGC/TE and TC are not independent: In stochastic linear systems, the observed dynamics in a

time-delay coordinate space can be interpreted as samples from a probability density of consecutive sequences of length m (if $1/\tau$ equals the sampling rate) in the observables. And an expansive mapping between probability densities induces information loss. Formally, expansion can be directly related to loss of certainty between states measured with finite precision [[6] (C)]. In stark contrast to the usual applications of WGC/TE, however (but see [17]), TC exploits the expansion of the backward mapping from effect x_2 to cause x_1 for determining the causal influence from x_1 to x_2 . In preliminary investigation we observed that TC can indeed detect effective influences in predominantly stochastic systems, raising the intriguing possibility that it is well suited for both deterministic and stochastic systems [[6] (H)].

To overcome the limitations of WGC when dealing with nonseparable dynamical systems, several approaches have been based on relations among state-space reconstructions. For example, tests for the existence of directed unique mappings between reconstructed manifolds can be used as an all-or-nothing criterion to detect causal links [5,8,18–21], to which TC represents an extension since it allows for gradual quantification of the influence. TC is most closely related to the empirical procedure of convergent cross mapping (CCM) [8] that yields interesting results in a range of applications, e.g., [8,22–24]. The CCM measure relies on errors when predicting one reconstruction manifold from another: the slower the convergence of the prediction error of r^{x_i} from r^{x_j} with increasing time series length, the weaker the causation x_i to x_j . We suspect that this effect is a consequence of the expansion: the more expansive the mapping $M_{j \rightarrow i}$ locally is, the more its nonlinearities hamper predictions with a given finite number of data points (see also [25]). In other words, we believe CCM evaluates deviations from the assumption that the mapping $\{r^{x_i}(t_1^{x_i}), \dots, r^{x_i}(t_k^{x_i})\}$ to $\{r^{x_j}(t_1^{x_j}), \dots, r^{x_j}(t_k^{x_j})\}$ is linear and therefore is an indirect estimate of the underlying effective influence [[6] (G)], which TC measures directly. Supporting this we find that CCM convergence speeds do not share the ability of TC to detect influences in linear stochastic systems [[6] (H)].

D. H. was funded by the Bundesministerium für Bildung und Forschung (Bernstein Award Udo Ernst, Grant No. 01GQ1106) and thanks U. A. Ernst for support during the project. We thank D. Rotermund and U. A. Ernst for helpful comments on the manuscript.

*daniel@neuro.uni-bremen.de

†e.laminski@uni-bremen.de

‡mschuene@uni-bremen.de

§pawelzik@neuro.uni-bremen.de

- [1] Aristotle, *Metaphysics* (350 BC).
- [2] N. Ay and D. Polani, *Adv. Complex Syst.* **11**, 17 (2008).
- [3] F. Takens, in *Dynamical Systems and Turbulence*, Springer Lecture Notes in Mathematics Vol. 898, edited by D. A. Rand and L.-S. Young (Springer-Verlag, Berlin, 1981).
- [4] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, *Phys. Rev. Lett.* **45**, 712 (1980).
- [5] B. Cummins, T. Gedeon, and K. Spendlove, *SIAM J. Appl. Dyn. Syst.* **14**, 335 (2015).
- [6] See Supplemental Material <http://link.aps.org/supplemental/10.1103/PhysRevLett.119.098301> for further analytics, numerical methods, application to experimental data and comparisons to other measures.
- [7] A. W. Bowman and A. Azzalini, *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations* (Oxford University Press, Oxford, 1997), Vol. 18.
- [8] G. Sugihara, R. M. May, H. Ye, C. Hsieh, E. R. Deyle, and M. Fogarty, *Science* **338**, 496 (2012).
- [9] D. Janzing, J. Mooij, K. Zhang, J. Lemeire, J. Zscheischler, P. Daniušis, B. Steudel, and B. Schölkopf, *Artif. Intell.* **182–183**, 1 (2012).
- [10] E. Bradley and H. Kantz, *Chaos* **25**, 097610 (2015).
- [11] O. E. Rössler, *Phys. Lett.* **57A**, 397 (1976).
- [12] P. L. Nunez, R. B. Silberstein, Z. Shi, M. R. Carpenter, R. Srinivasan, D. M. Tucker, S. M. Doran, P. J. Cadusch, and R. S. Wijesinghe, *Clin. Neurophysiol.* **110**, 469 (1999).
- [13] G. Nolte, A. Ziehe, V. V. Nikulin, A. Schlögl, N. Krämer, T. Brismar, and K.-R. Müller, *Phys. Rev. Lett.* **100**, 234101 (2008).
- [14] C. W. J. Granger, *Econometrica* **37**, 424 (1969).
- [15] T. Schreiber, *Phys. Rev. Lett.* **85**, 461 (2000).
- [16] L. Barnett, A. B. Barrett, and A. K. Seth, *Phys. Rev. Lett.* **103**, 238701 (2009).
- [17] S. Haufe, V. V. Nikulin, K.-R. Müller, and G. Nolte, *NeuroImage* **64**, 120 (2013).
- [18] W. Liebert, K. R. Pawelzik, and H. G. Schuster, *Europhys. Lett.* **14**, 521 (1991).
- [19] D. Chicharro and R. G. Andrzejak, *Phys. Rev. E* **80**, 026217 (2009).
- [20] H. Ma, K. Aihara, and L. Chen, *Sci. Rep.* **4**, 7464 (2014).
- [21] L. M. Pecora, T. L. Carroll, and J. F. Heagy, *Phys. Rev. E* **52**, 3420 (1995).
- [22] X. Wang, S. Piao, P. Ciais, P. Friedlingstein, R. B. Myneni, P. Cox, M. Heimann, J. Miller, S. Peng, T. Wang, H. Yanga, and A. Chen, *Nature (London)* **506**, 212 (2014).
- [23] S. Tajima, T. Yanagawa, N. Fujii, and T. Toyozumi, *PLoS Comput. Biol.* **11**, e1004537 (2015).
- [24] E. H. van Nes, M. Scheffer, V. Brovkin, T. M. Lenton, H. Ye, E. Deyle, and G. Sugihara, *Nat. Climate Change* **5**, 445 (2015).
- [25] J.-J. Jiang, Z.-G. Huang, L. Huang, H. Liu, and Y.-C. Lai, *Sci. Rep.* **6**, 24088 (2016).