Measuring Information-Transfer Delays

Michael Wibral[1]*, **Nicolae Pampu[2]**, **Viola Priesemann[3]**, **Felix Siebenhühner[1,4]**, **Hannes Seiwert[1]**, **Michael Lindner[5]**, **Joseph T. Lizier[6,7]**, **Raul Vicente[1,3,8]**

1 MEG Unit, Brain Imaging Center, Goethe University, Frankfurt, Germany, 2 Center for Cognitive and Neural Studies (Coneural), Cluj-Napoca, Romania, 3 Frankfurt Institute for Advanced Studies (FIAS), Goethe University, Frankfurt, Germany, 4 Department of Physics, University of California Santa Barbara, Santa Barbara, California, United States of America, 5 Center for Economics and Neuroscience, Friedrich-Wilhelms University, Bonn, Germany, 6 Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany, 7 CSIRO Information and Communication Technologies Centre, Marsfield, New South Wales, Australia, 8 Max-Planck Institute for Brain Research, Frankfurt, Germany

Abstract

In complex networks such as gene networks, traffic systems or brain circuits it is important to understand how long it takes for the different parts of the network to effectively influence one another. In the brain, for example, axonal delays between brain areas can amount to several tens of milliseconds, adding an intrinsic component to any timing-based processing of information. Inferring neural interaction delays is thus needed to interpret the information transfer revealed by any analysis of directed interactions across brain structures. However, a robust estimation of interaction delays from neural activity faces several challenges if modeling assumptions on interaction mechanisms are wrong or cannot be made. Here, we propose a robust estimator for neuronal interaction delays rooted in an information-theoretic framework, which allows a model-free exploration of interactions. In particular, we extend transfer entropy to account for delayed source-target interactions, while crucially retaining the conditioning on the embedded target state at the immediately previous time step. We prove that this particular extension is indeed guaranteed to identify interaction delays between two coupled systems and is the only relevant option in keeping with Wiener's principle of causality. We demonstrate the performance of our approach in detecting interaction delays on finite data by numerical simulations of stochastic and deterministic processes, as well as on local field potential recordings. We also show the ability of the extended transfer entropy to detect the presence of multiple delays, as well as feedback loops. While evaluated on neuroscience data, we expect the estimator to be useful in other fields dealing with network dynamics.

Introduction

Many phenomena in the world around us, such as traffic systems, gene regulatory networks, neural circuits and the Internet can be best understood in terms of complex networks. Understanding such networks requires knowledge about the existence and direction of the interactions in the network. Often, the network function also depends on the interaction timing. For example, understanding of the railway system is incomplete if only the location of train tracks and the direction in which they are used is known. At least information on train travel times is necessary to catch a glimpse of how the network serves its purpose, and only a timetable enables one to use this network efficiently. As in this example, interaction delays may have a pivotal role in understanding the function of complex networks.

In neuroscience, interaction delays arise mainly due to propagation of action potentials ('spikes') along axonal processes and can amount to several tens of milliseconds. The presence of axonal delays is of particular importance for coordinated neural activity (e.g. synchronization, Hebbian learning) because they add an intrinsic component to the relative timing between spikes. For example, two neurons projecting to a downstream neuron will be observed to spike simultaneously by this downstream neuron only when their relative timing of spikes compensates the difference in their axonal delays and in the dendritic delays to the soma of the target neuron. Indeed, disruption of coordinated activity by the pathological modification of axonal delays is thought to account for some deficits in diseases such as multiple sclerosis [1], schizophrenia [2], and autism [3]. Thus, the estimation of both, interaction delays and interaction strengths from multichannel brain recordings are needed to better resolve the dynamic coordination between different areas. In this paper we propose an extension of an information-theoretic *functional*, transfer entropy, to determine both the information transfer and interaction delays between processes.

In the following, we review the key concepts of Wiener causality and transfer entropy, and describe the outline of the paper.

Causality, Transfer Entropy, and the Estimation of Interaction Delays

Ideally, in explorative analyses both the presence of directed interactions between two physical systems, as well as their timing

should be detected without any *a priori* knowledge of the coupled systems or their specific interaction mechanism, i.e. a model free analysis is required. To keep our analysis as model-free as possible, we assume that the coupled physical systems $\mathcal{X},\mathcal{Y}$ produce the observed time series $\{x_1,\ldots,x_T\},\{y_1,\ldots,y_T\}$ via measurements at discrete times $t\in1\ldots T$. These time series are understood as realizations x_t,y_t of stationary random processes X,Y for mathematical treatment. The stationarity assumption for the random processes is convenient here as it allows to replace ensemble averages by time averages, but the proposed method will also work for ensemble averaging. In the remainder of the text, upper case letters X,Y refer to these random processes, X_t,Y_t to the random variables the processes are composed of, while lower case letters with subscript indices x_t,y_t refer to scalar realizations of these random variables. Bold case letters $\mathbf{X},\mathbf{Y},\mathbf{X_t},\mathbf{Y_t},\mathbf{x_t},\mathbf{y_t}$ refer to the corresponding processes, random variables, and their realizations in a state space representation (see the methods section for the construction of these state spaces).

The structure of directed interactions can be analyzed by assigning a causal influence from a process X to another one Y, if knowledge about the past of realizations of X and Y together allows one to predict the future of Y better than knowledge about the past of Y alone. This is known as Norbert Wiener's 'principle of causality', and does not by itself entail a modeling approach. Although this principle paved the way for a formal analysis of directed interactions, we note that today for an analysis of truly causal interactions more stringent requirements have to be met [4,5]. If only Wiener's principle is met, we speak of *predictive information transfer* [6]. However, predictive information transfer may often be exactly the quantity of interest when analyzing directed interactions in networks, especially when these networks actively process information [6,7].

Wiener's principle can be directly translated into an information-theoretic framework by reformulating it as the question: "*What information does the past of X provide about the future of Y, that the past of Y did not already provide?*". Schreiber [8] formalized this question in terms of a conditional mutual information $I(\cdot;\cdot|\cdot)$ between the involved quantities:

$$TE(X\rightarrow Y)=I(Y^+;\mathbf{X}^-|\mathbf{Y}^-),\qquad(1)$$

where Y^+ is a future random variable of the process Y, whereas $\mathbf{X}^-$ and $\mathbf{Y}^-$ denote suitably reconstructed past *state variables* of the processes X and Y, respectively. The corresponding quantity has been described several times in the literature (e.g. [8,9]) and is most often refered to as *transfer entropy* [8].

The use of transfer entropy or related methods [10,11] for model free analyses of directed interactions has seen a dramatic surge of interest recently, both in neuroscience [9,12–31], physiology [32–34], as well as in the general theories of computation [7,24,35] and causality [5]. For specific application scenarios transfer entropy has proven to be clearly superior compared to alternative analyses of interactions [22].

Schreiber originally defined the transfer entropy functional for random processes X and Y, with discrete-valued time index t, as [8]:

$$TE(X\rightarrow Y)=I(Y_{t+1};\mathbf{X_t}|\mathbf{Y_t}),\qquad(2)$$

Yet, it was noted early on that in real world physical systems information from $\mathcal{X}$ needs a finite time δ to arrive at $\mathcal{Y}$. Because of this, past states variables $\mathbf{X_t},\mathbf{Y_t}$ and future future random variables Y_{t+1}, that replace the abstract quantities $\mathbf{X}^-,\mathbf{Y}^-$ and Y^+ in the functional from equation 1 in the calculation of a specific estimator, such as in equation 2, have to be redefined appropriately to reflect this fact. Therefore, two suggestions were made to adapt transfer entropy:

- First, we and others suggested to use the following formula to take efficiently into account the possibility of a non-vanishing interaction delay [9,12,13,29],

$$TE(X\rightarrow Y,u)=I(Y_{t+u};\mathbf{X_t}|\mathbf{Y_t}),\qquad(3)$$

  where the parameter u is the time which an influence needs to propagate from $\mathcal{X}$ to $\mathcal{Y}$ [9,12,13,29]. A scanning approach for the parameter u was suggested to recover the delay with the largest predictive information transfer and, thereby, recover the dominant interaction delay [12].

- Second, Pompe and Runge [36] suggested a similar scanning approach to adjust the TE to consider transfer from a previous source variable X_t to a future target variable Y_{t+u}, while conditioning on the source state variable $\mathbf{Y_{t+u-1}}$, and additionally conditioning on the previous source state variable $\mathbf{X_{t-1}}$:

$$MIT(X\rightarrow Y,u)=I(Y_{t+u};X_t|\mathbf{Y_{t+u-1}},\mathbf{X_{t-1}}),\qquad(4)$$

This measure is known as the *momentary information transfer* (MIT) (see Methods section part on MIT for full details). While the idea to scan u in order to maximize MIT is similar to the first approach, the conditioning on a past source state was thought to ensure that the specific delay was identified where the transferred information appeared in the source first.

While the first approach in equation 3 seems like a natural extension of transfer entropy, we will show in the next section that equation 3 violates Wiener's principle of causality. This is because the above functional violates the requirement of an optimal self prediction from the past of Y to the future of Y that is implicit in Wiener's principle. In short, using a state $\mathbf{Y_t}$ that is not obtained *immediately prior* to the future Y_{t+u} – the state one is trying to predict – ignores potentially relevant predictive information (see below for details). We will provide a simple example where the above functional (equation 3) from references [9,12,13,29] does not recover the correct interaction delay.

We will also show by counter-example that the second suggestion, the MIT, is not able to reconstruct the correct interaction delay in a simple example in the presence of memory in the source.

Therefore, we present in this study an improved transfer entropy functional that honors the requirement of an optimal self prediction and that successfully recovers the correct interaction delay. We formulate a mathematical theorem that the improved transfer entropy functional is maximal when its delay parameter coincides with the underlying interaction delay and give the corresponding proof. To further validate our approach on finite data we run extensive simulations of stochastic and deterministic delay-coupled systems. Local field potentials are also used to test the recovery of interaction delays in electrophysiological recordings.

Furthermore, we will demonstrate below that our novel approach allows to test the presence of self-feedback activity in a single recorded signal. Finally, we discuss how information about interaction delays can be used to enhance the power of effective connectivity analyses.

Results

Comparison of Transfer Entropy Functionals in Relation to Wiener's Principle

Wiener's principle asks for information about the future of Y that the past of X can provide *in addition* to the information already provided by the past of Y. If, however, this latter information is underestimated, we may potentially obtain an erroneously high value of predictive information transfer from X to Y. From this it follows that information provided by the past of Y to the future of Y must be estimated optimally, a fact that we will refer to as the *self prediction optimality* (SPO) requirement from here on. Equivalently, from an information-theoretic computational perspective, we can view the 'self prediction' as an *information storage* by Y [37], and so underestimating the information storage in Y's dynamics can lead us to overestimate the information transfer from X to Y [7,35].

For practical applications of Wiener's principle this means that we have to guarantee an optimal self prediction at least within the limits of our respective prediction framework. For the most general case of TE where no a priori model-based knowledge can be used, this means we have to resort to model free prediction, as it is for example provided by the local predictor [38,39].

We will first show that the transfer entropy functional given in equation 3 is not self prediction optimal and then provide a new transfer entropy functional that is self prediction optimal.

To show that the functional from equation 3 is indeed not self-prediction optimal, we rewrite this equation by substituting t by $t-u$:

$$TE(X \to Y, u) = I(Y_t; \mathbf{X_{t-u}} | \mathbf{Y_{t-u}}). \tag{5}$$

We see that the self prediction of Y, or equivalently the conditioning, is done based on the state $\mathbf{Y}$ at time $t-u$. Let us define u_0 as the value of the parameter u that optimizes self prediction of Y_t by $\mathbf{Y_{t-u_0}}$. If we now scan u in search of the maximum predictive information transfer from $\mathbf{X_{t-u}}$ to Y_t, we will potentially condition our mutual information on states $\mathbf{Y_{t-u}}$ with $u \neq u_0$, i.e. we condition on states that are not optimal for self prediction. This suboptimal conditioning may artificially inflate transfer entropy values in a u-dependent manner. Thus, maximum transfer entropy values do not only depend on the true information flow from X to Y and its delay, but also on the quality of the self-prediction (conditioning). As a consequence, the maximal apparent information transfer estimated by equation 5 may be found at values of u that do not represent the true delay δ, and the attempt to identify the true interaction delay by maximizing predictive information transfer with the functional in equation 5 may lead to erroneous results.

From the above it follows that conditioning on the past state $\mathbf{Y_{t-u}}$ should always be done with respect to the optimal state $\mathbf{Y_{t-u_0}}$. Hence, a modified functional reads:

$$TE'(X \to Y, u) = I(Y_t; \mathbf{X_{t-u}} | \mathbf{Y_{t-u_0}}), \tag{6}$$

where u_0 in principle would have to be identified by a model-free prediction scheme, such as the one proposed by Ragwitz and Kantz [39]. However, we can abbreviate this procedure and formally prove that u_0 must be 1 sample if equation 6 is to represent a *causal relationship* (see next subsection). Furthermore, using $u_0 = 1$ properly eliminates any *information storage* from the past of Y that could otherwise be mistaken as information transfer from

X. And finally, the use of $u_0 = 1$ allows us to take a dynamical systems view of the *state transition* $\mathbf{Y_{t-1}} \to \mathbf{Y_t}$, and consider the TE as measuring how much information $\mathbf{X_{t-u}}$ provides about this state transition.

Result. *The predictive information transfer from X to Y over a time delay u is properly captured (aligning with Wiener's principle) by:*

$$TE_{SPO}(X \to Y, u) = I(Y_t; \mathbf{X_{t-u}} | \mathbf{Y_{t-1}}). \tag{7}$$

This functional fulfills the self prediction optimality requirement and we chose the subscript 'SPO' to reflect this. We note that in the case $u=1$ $TE(X \to Y, u=1)$ the old estimator (equation 3) is equal to $TE_{SPO}(X \to Y, u=1)$ (and this equality also holds for the original formulation of the TE from equation 2); however, this does not hold for other u in general and as such using equation 6 with $u > 1$ and $u_0 \neq 1$ does not satisfy Wiener's principle. To see this, we rewrite from equation 6 : $TE'(X \to Y, u) = I(Y_t; \mathbf{X_{t-u}}, \mathbf{Y_{t-u_0}}) - I(Y_t; \mathbf{Y_{t-u_0}})$, which allows us then to see explicitly that both, (1) the joint information term $I(Y_t; \mathbf{X_{t-u}}, \mathbf{Y_{t-u_0}})$ supplied by the source and the past of the target, and (2) the information storage term $I(Y_t; \mathbf{Y_{t-u_0}})$, differ under the cases $u_0 = u$ (old delayed TE estimator, equation 3) and $u_0 = 1$ (new delay TE estimator, equation 7).

In the next section we provide a theorem that formally states that TE_{SPO} is maximal when the parameter u is equal to the true interaction delay δ, and give a proof. Thus, TE_{SPO} can be used to recover an interaction delay δ in coupled systems as:

$$\delta = \arg\max_u (TE_{SPO}(X \to Y, u)). \tag{8}$$

A Theorem on the Identifiability of the True Delay δ of an Interaction

Part of this study is a proof of the fact that the new proposed functional assumes its maximal value when the delay parameter u in TE_{SPO} (equation 7) is equal to the true delay δ. The main finding can be summarized in the following theorem:

Theorem 1. *For two discrete-time random processes X, Y with a state space representation $\mathbf{X}, \mathbf{Y}$, coupled from X to Y via a non-zero delay δ, $TE_{SPO}(X \to Y, u) = I(Y_t; \mathbf{X_{t-u}} | \mathbf{Y_{t-1}})$ is maximal for $u = \delta$. This also holds in the presence of additional coupling from Y to X.*

The main ideas behind delay reconstruction via maximizing TE_{SPO} are illustrated in Figure 1. By scanning the delay parameter u we shift the considered state $\mathbf{X_{t-u}}$ of the source process X in time. If this state $\mathbf{X_{t-u}}$ is in the relative future of the observation to be predicted for Y, i.e. Y_t, its influence has not arrived at Y yet. As a consequence, the state is uninformative and we get low TE_{SPO}. If the state has a time delay $u = \delta$, such that the influence arrives exactly at Y_t, then TE_{SPO} is maximal. If the state has too long a delay, then its influence has arrived before Y_t and is already taken into account via conditioning on the past state $\mathbf{Y_{t-1}}$; again we obtain low TE_{SPO}. In the following we will present our proof. Since it is of a technical nature the reader may safely skip ahead if not interested in this material.

Proof of Theorem 1

Outline. We start by showing that the three random variables $\mathbf{X_{t-\delta-\xi}}$, $(\mathbf{X_{t-\delta}}, \mathbf{Y_{t-1}})$, and Y_t form a Markov chain for $\forall \xi \neq 0$. To this end we first demonstrate the *d-separation* of $\mathbf{X_{t-\delta-\xi}}$ and $\mathbf{Y_t}$ by $(\mathbf{X_{t-\delta}}, \mathbf{Y_{t-1}})$ which is equivalent to conditional independence of $\mathbf{X_{t-\delta-\xi}}$ and $\mathbf{Y_t}$, given $(\mathbf{X_{t-\delta}}, \mathbf{Y_{t-1}})$ [4]. This, in turn, is equivalent
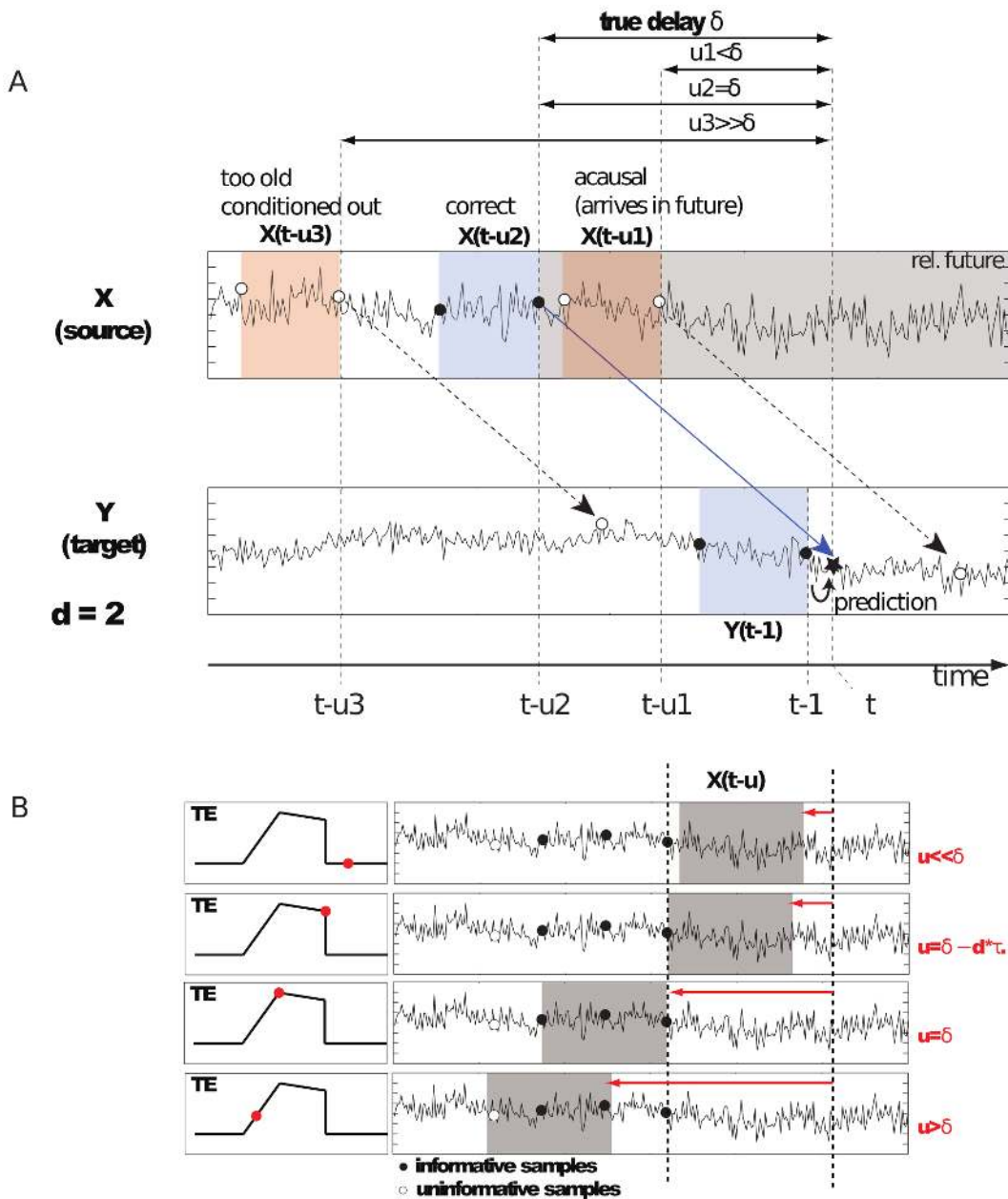
Figure 1. Illustration of the main ideas behind interaction delay reconstruction using the TE$_{SPO}$ estimator. (A) Scalar time courses of processes X,Y coupled $X \to Y$ with delay δ, as indicated by the blue arrow. Colored boxes with circles indicate data belonging to a certain state of the respective process. The star on the Y time series indicates the scalar observation $Y(t)$ to be predicted in Wiener's sense. Three settings for the delay parameter u are depicted: (1) $u < \delta$ with – u is chosen such that influences of the state $X(t - u_1)$ on Y arrive in the future of the prediction point. Hence, the information in this state is useless and yields no transfer entropy. (2) $u = \delta$ – u is chosen such that influences of the state $X(t - u_2)$ arrive exactly at the prediction point, and influence it. Information about this state is useful, and we obtain nonzero transfer entropy. (3) $u > \delta$ – u is chosen such that influences of the state $X(t - u_3)$ arrive in the far past of prediction point. This information is already available in the past of the states of Y that we condition upon in TE_{SPO} Information about this state is useless again, and we obtain zero transfer entropy. (B) Depiction of the same idea in a more detailed view, depicting states (gray boxes) of X and the samples of the most informative state (black circles) and noninformative states (white circles). The the curve in the left column indicates the approximate dependency of TE_{SPO} versus u. The red circles indicates the value obtained with the respectzive states on the right.
doi:10.1371/journal.pone.0055809.g001

to $\mathbf{X}_{t-\delta-\xi} \to (\mathbf{X}_{t-\delta}, \mathbf{Y}_{t-1}) \to \mathbf{Y}_t$ forming a Markov chain. Using the *decomposition property* of conditional independence (e.g. [4]), we see that this result also holds if we replace the state $\mathbf{Y}_t$ by the corresponding scalar observation Y_t. We then use this Markov chain for statements on the relation of mutual information terms

built from its variables and rearrange terms to arrive at the statements on TE formulated in theorem 1.

d-separation. From figure 2, representing the causal graph of the two random processes X, Y, we see that:

- all *sequential* paths from $\mathbf{X}_{t-\delta-\xi}$ into $\mathbf{Y}_t$ are blocked by the joint random variable $(\mathbf{X}_{t-\delta}, \mathbf{Y}_{t-1})$ (sequential paths of this kind

only exist for $\xi \geq 0$, i.e. states of $\mathbf{X}$ that are further back in time than the true interaction delay δ.),

- whereas *none* of the *convergent* paths between $\mathbf{X_{t-\delta-\xi}}$ and $\mathbf{Y_t}$ are blocked, because we do not condition on $\mathbf{Y_{t+1}}$ or any other future value of $\mathbf{Y}$,
- and all of the *divergent* paths between $\mathbf{X_{t-\delta-\xi}}$ and $\mathbf{Y_t}$ are blocked by conditioning on $\mathbf{X_{t-\delta}}$.

Hence, $\mathbf{X_{t-\delta-\xi}}$ and $\mathbf{Y_t}$ are indeed *d-separated* by $(\mathbf{X_{t-\delta}}, \mathbf{Y_{t-1}})$.

Conditional probability distributions and Markov chain property. Given this *d-separation*, we can state for the corresponding conditional and joint probability distributions $p(\cdot)$ that:

$$
\begin{aligned}
&p(\mathbf{X_{t-\delta-\xi}}, \mathbf{Y_t} | \mathbf{X_{t-\delta}}, \mathbf{Y_{t-1}}) \\
&= p(\mathbf{X_{t-\delta-\xi}} | (\mathbf{X_{t-\delta}}, \mathbf{Y_{t-1}})) p(\mathbf{Y_t} | (\mathbf{X_{t-\delta}}, \mathbf{Y_{t-1}})),
\end{aligned}
\tag{9}
$$

which is equivalent to $\mathbf{X_{t-\delta-\xi}} \rightarrow (\mathbf{X_{t-\delta}}, \mathbf{Y_{t-1}}) \rightarrow \mathbf{Y_t}$ being a Markov chain. From this it follows via the *decomposition property* of conditional independence that the following is also a Markov chain:

$$
\mathbf{X_{t-\delta-\xi}} \rightarrow (\mathbf{X_{t-\delta}}, \mathbf{Y_{t-1}}) \rightarrow Y_t.
\tag{10}
$$

From the markov chain property to delay reconstruction. Building on equation 10 we see that the Markov property still holds if we form a new random variable from $\mathbf{X_{t-\delta-\xi}}$ by considering it jointly with $\mathbf{Y_{t-1}}$, and we obtain:

$$
(\mathbf{X_{t-\delta-\xi}}, \mathbf{Y_{t-1}}) \rightarrow (\mathbf{X_{t-\delta}}, \mathbf{Y_{t-1}}) \rightarrow Y_t
\tag{11}
$$

Using the data processing theorem, this leads to an inequality for mutual information terms between the variables as:

$$
I(Y_t; \mathbf{X_{t-\delta-\xi}}, \mathbf{Y_{t-1}}) \leq I(Y_t; \mathbf{X_{t-\delta}}, \mathbf{Y_{t-1}})
\tag{12}
$$

By subtracting the *active information storage* [37], $I(Y_t; \mathbf{Y_{t-1}})$, from both sides of this inequality we get:

$$
\begin{aligned}
&I(Y_t; \mathbf{X_{t-\delta-\xi}}, \mathbf{Y_{t-1}}) - I(Y_t; \mathbf{Y_{t-1}}) \\
&\leq I(Y_t; \mathbf{X_{t-\delta}}, \mathbf{Y_{t-1}}) - I(Y_t; \mathbf{Y_{t-1}})
\end{aligned}
\tag{13}
$$

and, by using the chain rule for conditional mutual information:

$$
I(Y_t; \mathbf{X_{t-\delta-\xi}} | \mathbf{Y_{t-1}}) \leq I(Y_t; \mathbf{X_{t-\delta}} | \mathbf{Y_{t-1}}) \Rightarrow
\tag{14}
$$

$$
TE_{SPO}(\mathbf{X_t} \rightarrow Y_t, u = \delta - \xi) \leq TE_{SPO}(\mathbf{X_t} \rightarrow Y_t, u = \delta) \, \forall \xi \in \mathbb{R}
\tag{15}
$$

Hence, the value of TE_{SPO} is indeed maximal when the parameter u is equal to the true delay δ – q.e.d.

Remarks. Since the above derivation does not hold if we replace the conditioning set of variables $(\mathbf{X_{t-\delta}}, \mathbf{Y_{t-1}})$ by $(\mathbf{X_{t-\delta}}, \mathbf{Y_{t-u_0}})$ with $u_0 \neq 1$ (because the relevant d-separation is not maintained), then we must set $u_0 = 1$ to obtain an estimator with a potential causal interpretation.

Similarly, we note that the use of $u_0 = 1$ is necessary to eliminate information storage $I(Y_t; \mathbf{Y_{t-1}})$ from the past of Y being attributed to the predictive information transfer $TE_{SPO}(\mathbf{X_t} \rightarrow Y_t, u)$.

Importantly, since this proof only relied on a proper conditioning of the probability distributions of the parents of Y_t in the causal graph, it also holds for bidirectional coupling, as the parents of Y_t in the causal graph do not change by adding coupling from Y to X.
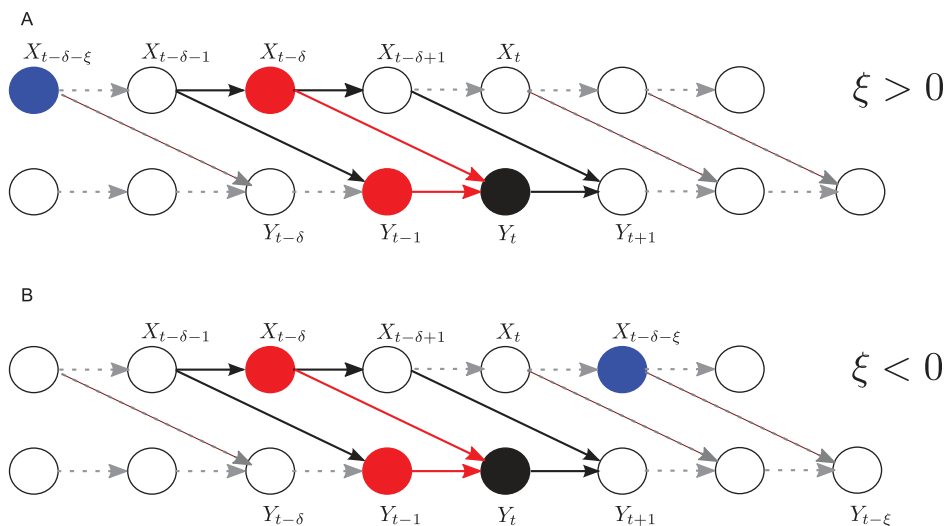

Figure 2. Causal graph for two coupled systems $X \rightarrow Y$. Illustration of d-separation of $\mathbf{X_{t-\delta-\xi}}$ and $\mathbf{Y_t}$ by $(\mathbf{X_{t-\delta}}, \mathbf{Y_{t-1}})$. Arrows indicate a causal influence (directed interaction). Solid lines indicate a single time step, broken lines an arbitrary number of time steps. The black circle is the state to be predicted in Wiener's sense, the red circles indicate the states that form its set of parents in the graphs. These states are also the ones conditioned upon in the novel estimator TE_{SPO}. The blue circle indicates the state in the graph for which we want to determine that forms a Markov chain: $\mathbf{X_{t-\delta-\xi}} \rightarrow (\mathbf{X_{t-\delta}}, \mathbf{Y_{t-1}}) \rightarrow \mathbf{Y_t}$. For $\xi \neq 0$ all sequential paths from $\mathbf{X_{t-\delta-\xi}}$ into $\mathbf{Y_t}$ are blocked, as are the divergent paths between these nodes. All convergent paths (e.g. via $\mathbf{Y_{t-\delta}}$ in (B)) are not blocked. This holds for $\xi > 0$ (A) and $\xi < 0$ (B).
doi:10.1371/journal.pone.0055809.g002

We also note that some constraints apply: (a) If the two systems cannot be directly observed, but only noisy observations $\tilde{X}, \tilde{Y}$ of the true systems dynamics are available, theorem 1 cannot be proven for these noisy observations $\tilde{X}, \tilde{Y}$, but may hold for many cases – see the examples given int the results section. We can also provide a proof (not given here) for autoregressive, linearly, unidirectionally coupled systems, where at most one of the two observations is noisy, i.e. we have $X, \tilde{Y}$. However, counter examples can be found for very pathological structures of the noise on the two systems, that are not expected in physical systems (multi modal combinations of δ-distributions). (b) Simulations show that for Gaussian noise of low amplitude, delays can still be identified correctly. (c) If the bidirectional coupling leads to full synchronization of the two systems, such that the probability distributions in equation 17 are delta distributions, transfer entropy is not defined. (d) There must be no other indirect sequential paths $\mathbf{X}_{t-\delta-\xi} \rightarrow Z_{t-v} \rightarrow Y_t$ (for some $\delta + \xi > v \geq 1$) via some other variable Z which are not blocked by $(\mathbf{X}_{t-\delta}, \mathbf{Y}_{t-1})$ since this would violate the d-separation here. Extensions of the proof to this situation are conceivable, but are the topic of future studies.

Inability of Momentary Information Transfer to Reconstruct Coupling Delay

In order to contrast $TE_{SPO}(Y \rightarrow X, u)$ with the momentary information transfer (MIT) $MIT(X \rightarrow Y, u) = I(Y_t; X_{t-u} | \mathbf{Y}_{t-1}, \mathbf{X}_{t-u-1})$ [36], we next examine two test cases. Case (Ia) contains noisy short-term source memory which leads to an erroneous delay estimation by MIT; in contrast, case (Ib) is an example that was reported to produce erroneous results in TE_{SPO} in [36], but we show here that this was due to the use of a symbolic preprocessing step in [36] and that TE_SPO is perfectly capable to reconstruct the correct delay.

As explained in the methods section, in test case (Ia) Y_t is a direct function of X_{t-1}, while X_{t-1} itself is a noisy mapping from X_{t-2} (with noise parameter η, also see table 1). As such, $u=1$ should be identified as the correct interaction delay here, although the source memory makes $u=2$ a potential candidate for an incorrect identification. The dynamics for test case (Ia) were run to provide 10^6 observations for estimating the required probability distribution functions. The estimation of the modified transfer entropy functional and the momentary information transfer was performed for this test case with the open-source "Java Information Dynamics Toolkit" [40] as detailed in the Methods section.

Figure 3 shows the results of measuring $TE_{SPO}(Y \rightarrow X, u)$ and $MIT(X \rightarrow Y, u)$, with delays $u=1$ and 2, as a function of the source noise parameter η. We see that, in line with our earlier proof regarding this situation of unidirectional coupling, $TE_{SPO}(Y \rightarrow X, u)$ consistently identifies the correct delay $u=1$, since $TE_{SPO}(Y \rightarrow X, u=1) > TE_{SPO}(Y \rightarrow X, u=2)$ for all $\eta > 0$. On the other hand, for a significant range of η, $MIT(X \rightarrow Y, u)$ is

deceived by the source memory into incorrectly identifying $u=2$ as the relevant delay.

Certainly $MIT(X \rightarrow Y, u)$ fulfills its design in identifying the lag to the time step of the source where the relevant information in the target first appeared. As we see in this example however, the existence of some information regarding the target variable in the source at a certain lag does not mean that this relevant information was transferred to the target at that particular lag. Here, the memory in the source makes the source strongly correlated to the target over lag $u=2$. This drives both measures to high values for $u=2$ with low noise η, yet while $TE_{SPO}(Y \rightarrow X, u=1)$ remains higher still, $MIT(X \rightarrow Y, u=1)$ conditions out this correlated information and so falls below $MIT(X \rightarrow Y, u=2)$. MIT returns the wrong result here because the conditioning on previous source state is not necessary and removes relevant information.

Finally, we examine test case (Ib), a bidirectionally coupled logistic map, for which Pompe and Runge [36] found that the $TE_{SPO}(X \rightarrow Y, u)$ incorrectly identified the interaction lag as $u=1$ instead of $u=2$, when this quantity was estimated using symbolic mapping (capturing ordinal relationships). In contrast to the findings in [36], our analysis described in the methods section using Kraskov-Grassberger-Stögbauer estimators [41] is able to accurately identify the correct lag $u=2$ as having larger $TE_{SPO}(X \rightarrow Y, u)$ (2.123 bits for $u=2$ as compared to 0.826 bits for $u=1$). This result is in line with our proof that $TE_{SPO}(X \rightarrow Y, u)$ is maximized at the correct delay even in the case of bidirectional coupling. It is also in line with the more detailed empirical results we obtained for bidirectionally coupled processes as presented below. The opposite finding in [36] appears to simply be an artifact of their symbolic mapping approach: symbolic mapping may be a useful technique to handle small data sets, but it certainly removes parts of the information about the processes, and this information may well be relevant. Certainly, this is the case with coupled logistic maps, where examining ordinal relationships will miss many of the subtleties regarding how consecutive states are updated by the map.

Estimating Interaction Delays from Simulated Data

Here, we test the capability of TE_{SPO} to detect the interaction delays from a series of simulated and experimental time series. The different cases cover stochastic, deterministic and real time series, representing different interaction configurations and delay ranges, and are described in detail in the methods section. The estimation of the modified transfer entropy functionals in these test cases (II-IX) was performed with the open-source MATLAB toolbox TRENTOOL [42] as detailed in the Methods section. State space reconstruction was performed using the Ragwitz criterion [39] in TRENTOOL, to obtain states that allow optimal self prediction, given the data. Throughout this section the *estimated* delays are

Table 1. Definition of stochastic self-mapping updates f with memory and noise for variable X in test case (Ia).

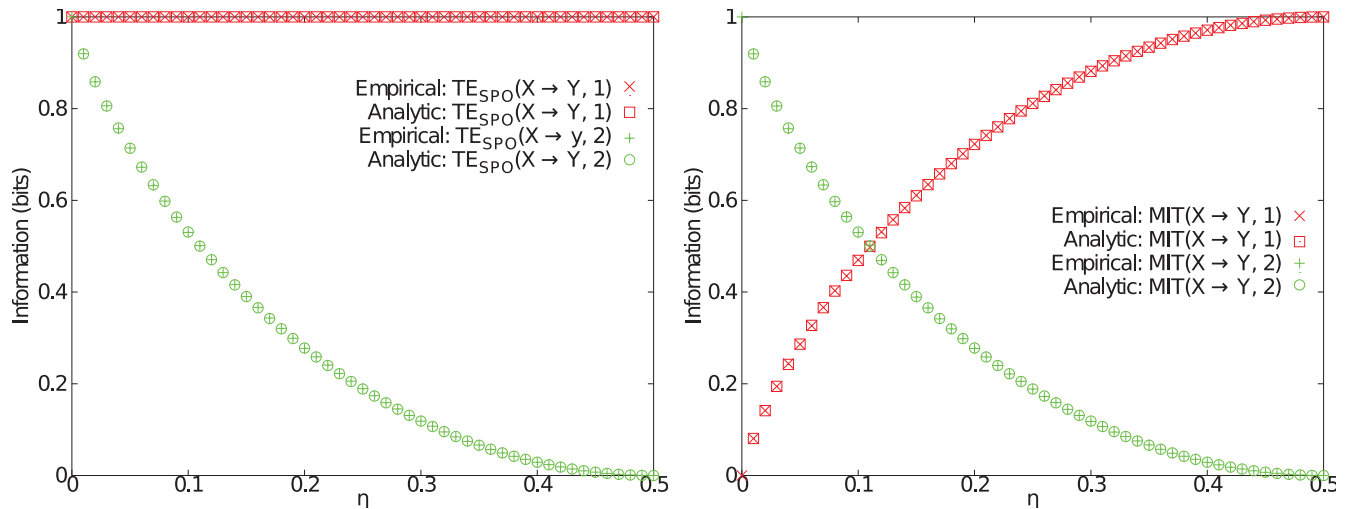| $X(t-1)$	$p(X(t)=f(X(t-1)))$			
	$X(t)=0$	$X(t)=1$	$X(t)=2$	$X(t)=3$
0	$(1-\eta)/2$	$\eta/2$	$(1-\eta)/2$	$\eta/2$
1	$(1-\eta)/2$	$\eta/2$	$(1-\eta)/2$	$\eta/2$
2	$\eta/2$	$(1-\eta)/2$	$\eta/2$	$(1-\eta)/2$
3	$\eta/2$	$(1-\eta)/2$	$\eta/2$	$(1-\eta)/2$

Figure 3. Test case (Ia), comparison of MIT and TE. Analytic and empirical measurements of (a) Transfer entropy $TE_{SPO}(X \to Y, u)$ and (b) Momentary information transfer $MIT(X \to Y, u)$ as a function of memory noise parameter η for the discrete-valued process with short-term source memory and a delay $\delta = 1$. Each measure is plotted for delays $u = 1$ (red) and 2 (green). The correct causal interaction delay coorsponds $u = 1$ and therefore we expect an appropriate measure to always return a higher value with $u = 1$ than with $u = 2$, i.e the red curve should always be at higher values than the green curve. Nevertheless, there is potential for $u = 2$ to be identified erroneously as the delay due to the presence of memory in the source X, and MIT indeed finds this result for a range of the memory noise parameter η (below 0.1).
doi:10.1371/journal.pone.0055809.g003

indicated by $\hat{\delta}$, whereas the true, simulated delays are indicated by δ.

Overview of simulated test cases II-IX. Figure 4 presents the general structure of test cases (II-VII,IX). All these cases comprise two systems labeled as X, Y, which are either both autoregressive order 10 processes (AR(10), equation 29), or both Lorenz systems (Lorenz, equation 32). For the Lorenz systems, the second coordinate (V - see equation 32) was used as the observable producing the time series used for analysis. The systems may interact in the direction $X \to Y$, with either a single delay δ_{XY}, or a set of delays $\{\delta_{XY}\}$, with coupling strengths γ_{XY} or $\{\gamma_{XY}\}$, respectively. In the reverse direction $Y \to X$ we only consider the case of single interactions with parameters δ_{YX}, and γ_{YX}. Additionally, in some of the cases delayed self feedback is present from process X to process X, with delay δ_{XX}, and strength γ_{XX}. All simulated interactions, including self-feedback, were non-linear (quadratic) functions. One additional case (VIII) investigates delay reconstruction from a unidirectionally coupled ring structure of three Lorenz systems; the last case (IX) simulates the effects of observation noise on delay reconstruction. Details of the test cases are presented in table 2. For each test cases 50 data segments (trials) of 3000 sampling points each were simulated, resulting in a total of 150.000 data points. A full description of the generating equations for the system dynamics and the simulation details can be found in the subsection on the test cases in the methods section. In the following, we present results for these eight test cases (II-IX), with test case (V) serving as an example for the inability of the 'old' estimator [9,12,13,29] to recover the correct interaction delays.

Recovery of a single interaction delay. In test case (II) we investigated two unidirectionally coupled autoregressive (AR) processes where a single interaction delay $\delta_{XY} = 20$ was present. We investigated TE_{SPO} as a function of the assumed interaction delay u. Figure 5 shows the results of computing TE_{SPO} and its statistical significance (with a null hypothesis of no source-target coupling, see Methods) for the two possible directions of interaction, $X \to Y$ and $Y \to X$. $TE_{SPO}^{X \to Y}$ shows a maximal value for $u = 20$ units, which matches the nominal value of 20 sampling

steps. $TE_{SPO}^{X \to Y}$ is statistically significant across a certain interval of delays around the maximum (14 to 23 sampling points) even when corrected for multiple comparisons. This blurring of the statistical significance of the predictive information transfer can be partly explained by *memory* in the source X (via autoregressive terms) meaning the predictive value of the actual directly influential scalar observation $x_{t+1-\delta}$ of the source is detectable in states $\mathbf{x_{t+1-u}}$ of X both *before and after* the actual delay $u = \delta$ (compare the extension of sources states indicated by shaded boxes in figure 1). An additional factor here is that examination of the source states $\mathbf{x_{t-u}}$ (instead of scalar observations x_{t-u}) means that full information about the directly influential observation $x_{t+1-\delta}$ is contained in several source states *after* $\mathbf{x_{t+1-\delta}}$. Crucially, the opposite direction ($Y \to X$) reveals a flat profile with no statistical significance, in correspondence with the absence of a directed interaction from process Y to X.

Recovery of multiple interaction delays. In the test case (III), we investigated two unidirectionally coupled AR processes where multiple interaction delays δ_{XY} were present,
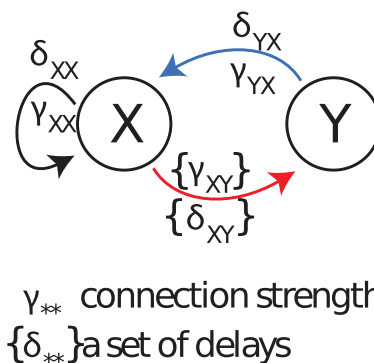

Figure 4. Overview over the structure of simulated test cases II-IX. Note that not all combination of links and parameters are always investigated. For details refer to table 2.
doi:10.1371/journal.pone.0055809.g004

Table 2. Parameter settings used to create the simulated test cases II-IX.

Testcase	System	$\{\delta_{XY}\}$	δ_{YX}	δ_{XX}	$\{\gamma_{XY}\}$	γ_{YX}	γ_{XX}
II	AR(10)	20	n.a.	n.a.	0.1	0	0
III	AR(10)	{15,20,25,30,35}	n.a.	n.a.	0.1	0	0
IV	AR(10)	{18,19,20,21,22}	n.a.	n.a.	0.1	0	0
V	Lorenz	45	75	n.a.	0.1	0.1	0
VI	Lorenz	75	n.a.	45	0.3	0	0.3
VII	Lorenz	45	n.a.	75	0.3	0	0.3
VIII	Lorenz (Ring)	20	$\delta_{YZ}=30$	$\delta_{ZX}=40$	0.15	$\gamma_{YZ}=0.15$	$\gamma_{ZX}=0.15$
IX	Lorenz+Noise	45	75	n.a.	0.1	0.1	0

Lorenz = chaotic Lorenz system; AR = autoregressive processes. For the meaning of the coupling and delays constants $\{\delta_{..}\}$ and $\gamma_{..}$ refer to figure 4.
doi:10.1371/journal.pone.0055809.t002

$\delta_{XY} = 15,20,25,30,35$. Figure 6 reveals that they can be readily detected by scanning TE_{SPO}. Well separated peaks indicate the presence of multiple delays around values of ~ 14, 19, 25, and 30 sampling units for the direction of interaction X to Y. The curve displays an additional shoulder at $u = 35$. Nominal delays in the simulations were 15, 20, 25, 30 and 35, and thus all but the longest delay were correctly detected. The longest delay is most likely not detected because much information from the relevant source state $\mathbf{x}_{t+1-\delta}$ has already been communicated to the target over several shorter delays, due to the inherent memory of the AR(10) process, and there is no longer enough novel information provided by the source given the past state of the target to evoke a clear peak. However, the transfer entropy values $TE_{SPO}^{X \to Y}$ indeed were

statistically significant up to an assumed delay of 35 units, in line with the maximal delay simulated.

A more complex case (IV) is encountered when dealing with a smooth distribution of delays. Figure 7 demonstrates that in this case, a peak of TE_{SPO} is attained near to the mean of the distribution of delays. The width of the peak is proportional to the width of the delay distribution. However, an exact estimation of the range of delays is difficult since single delays are also associated with broad peaks in the TE_{SPO} versus assumed delay curves (see figure 5, but note the different scale of the time axes). We note that the peak of TE_{SPO} is skewed towards the shorter of the actual interaction delays, and this may be due to: (a) the longer delays providing less novel information from the source given that it is already contained in the target state from the shorter delays (as discussed in the above paragraph); and/or (b) the persistence of information of the current influential component $x_{t+1-\delta}$ of the source state $\mathbf{x}_{t+1-\delta}$ in several following source states (as discussed in the preceding subsection).

Recovery of delays of bidirectional interactions. For the analysis of two bidirectionally coupled Lorenz systems in test case (V), with $\delta_{XY} = 45$ and $\delta_{XY} = 75$, transfer entropy values peaked at $u = 46$ and $u = 76$ samples for the interaction from process X to Y,
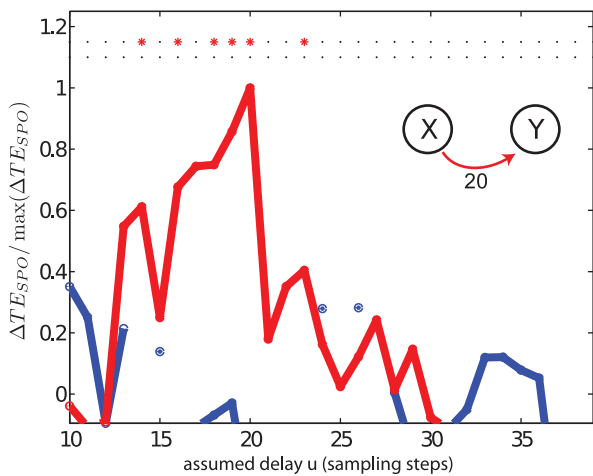

Figure 5. Test case (II). Transfer entropy (TE_{SPO}) values and significance as a function of the assumed delay u for two unidirectionally coupled autoregressive systems. For visualization purposes all values were normalized by the maximal value of the TE between the two systems, i.e. $\max(TE_{SPO}^{X \to Y}(u), TE_{SPO}^{Y \to X}(u))$. Red and blue color indicate normalized transfer entropy values and significances for interactions $X \to Y$ and $Y \to X$, respectively. The nominal interaction delay δ_{XY} used for the generation of the data was 20 sampling units from the process X to Y. Asterisks indicate those values of u for which the p-value < 0.05 once corrected for multiple comparisons. Missing points for $TE_{SPO}^{Y \to X}(u)$ are because the analyses for these u's failed to pass the shift test (a conservative test in TRENTOOL to detect potential instantaneous cross-talk or shared noise between the two time series, see [42]).
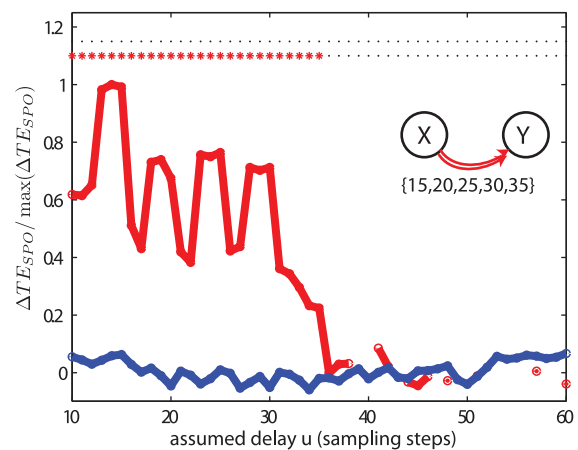doi:10.1371/journal.pone.0055809.g005

Figure 6. Test case (III). Transfer entropy (TE_{SPO}) values and significance as a function of the assumed delay u for two unidirectionally coupled autoregressive systems with multiple delays. The simulated delays $\{\delta_{XY}\}$ were 15, 20, 25, 30 and 35 sampling points. The rest of the parameters and criteria used are the same as those in Figure 5.
doi:10.1371/journal.pone.0055809.g006

and Y to X, respectively (Figure 8). These values differed only by one sample from to the true interaction delays used for simulation. Moreover, the relation between the transfer entropy values for the two coupling directions reversed with increasing delay parameter u: for delay values up to 65, transfer entropy values were larger for the direction from process X to Y, for values of u larger than 65 the opposite was the case. This is an important finding as the difference of the transfer entropies in both directions, also called the net transfer entropy, is often used as an indicator of the effective or dominating interaction structure. However, in our example, this net information transfer changed sign with changing delay parameter u. As an additional result, we show that the cross correlation function between the signals of the two systems was flat (Figure 9), as expected for a quadratic coupling.

Failure of the traditional estimator to recover the correct delays. We also analyzed the case (V) of bidirectionally coupled Lorenz systems with the 'old' estimator ($u_0 = u$) from references [9,12,13,29] (figure 10). As expected on theoretical grounds, this type of estimator did not recover the simulated delays for the two coupling directions ($\delta_{XY} = 45$ and $\delta_{YX} = 75$), but instead delivered erroneous estimates ($\hat{\delta}_{XY} = 28$ and $\hat{\delta}_{YX} = 61$).

Recovery of the delay of a feedback loop. We consider here the cases (VI) and (VII) where a feedback loop is affecting the dynamics of a node. In particular, we investigate first how the presence of feedback can be detected, and second how a feedback loop in a node affects estimation of transfer entropy to a different node.

We note that feedback loops do not pose a principal conceptual problem. Mathematically, a perfect state space reconstruction (see Methods for an explanation of state space reconstruction.) would subsume the feedback activity into the node's dynamics. In practical terms, however, *long* self-feedback delays – in comparison to the intrinsic dynamics of the node – can not be covered practically in Taken's classical state space reconstruction [43]. The reasons for this are twofold: (1) Self prediction performance may become unstable in high dimensional state spaces necessary to recover the delayed self-feedback. Algorithms determining optimal embedding parameters by optimal self prediction may get stuck in local minima this way. (2) Computational cost quickly diverges
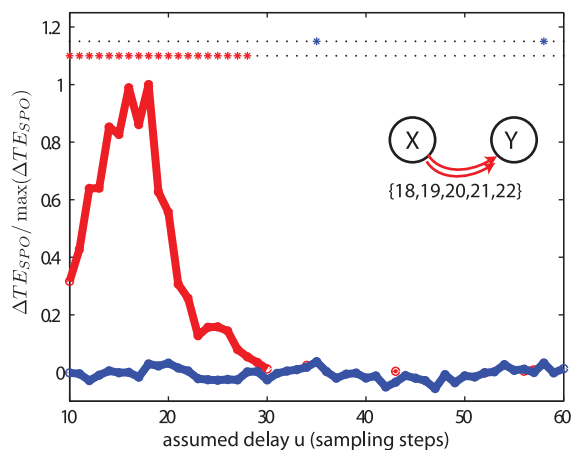

Figure 8. Test case (V). Transfer entropy (TE_{SPO}) values and significance as a function of the assumed delay u for two bidirectionally coupled, chaotic Lorenz systems. The simulated delays were $\delta_{XY} = 45$ and $\delta_{YX} = 75$, and the coupling constants were $\gamma_{XY} = \gamma_{YX} = 0.1$. The delays were recovered as $\hat{\delta}_{XY} = 46$ and $\hat{\delta}_{YX} = 76$. For more parameters see table 2.
doi:10.1371/journal.pone.0055809.g008

with a growing number of dimensions in state space. Hence, one may not even include the necessary number of dimensions and the necessary range of embedding delays in the parameter ranges that are searched to get the optimal embedding values. Given an imperfect state space reconstruction, even our new estimator is no longer self-prediction optimal, then. In the following we show that this estimator is nevertheless useful *to detect* self-feedback under conditions of long self-feedback delays.

Thus, we start by considering the dynamics of a single Lorenz system (process X in test case (VI)) subject to a non-linear delayed feedback loop. This is, its own past output is fed back after it has undergone a nonlinear (quadratic) transformation (see the section 'test cases' in Methods). To detect the presence of feedback activity we evaluate $TE_{SPO}(X \to X, u)$ between the system's original time series and its own past u time steps back. To this end we choose embedding parameters such that the reconstructed states do not

Figure 7. Test case (IV). Transfer entropy (TE_{SPO}) values and significance as a function of the assumed delay u for two unidirectionally coupled autoregressive systems with multiple delays. The simulated delays $\{\delta_{XY}\}$ were 18, 19, 20, 21 and 22 sampling points. The rest of the parameters and criteria used are the same as those in Figure 5.
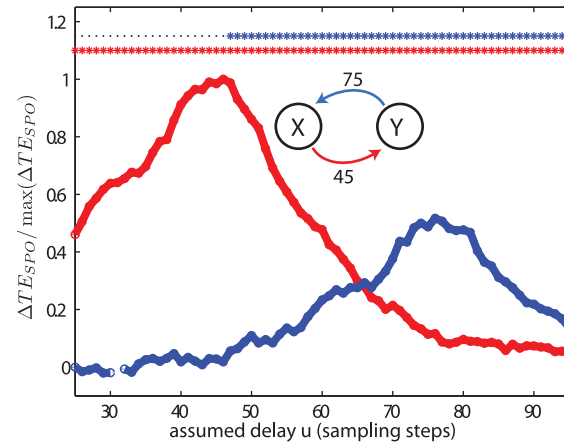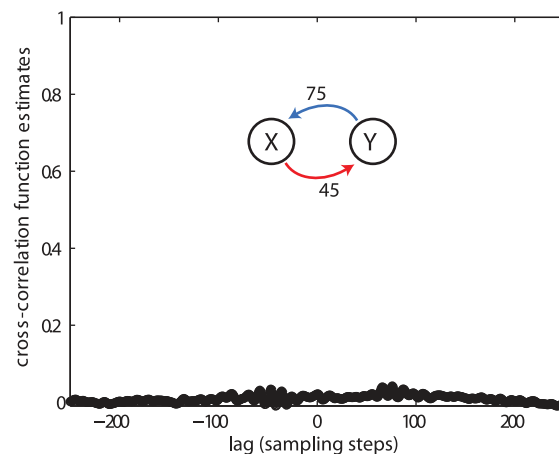doi:10.1371/journal.pone.0055809.g007

Figure 9. Test case (V). Crosscorrelation function for the two quadratically coupled chaotic Lorenz systems from figure 8.
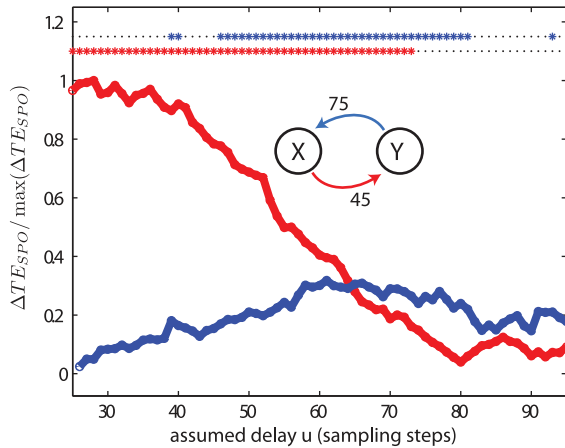doi:10.1371/journal.pone.0055809.g009

Figure 10. Test case (V) analyzed with the old estimator. Transfer entropy (TE) values and significance estimated by the old estimator from references [9,12,13,29] as a function of the assumed delay u for two bidirectionally coupled, chaotic Lorenz systems. The simulated delays were $\delta_{XY} = 45$ and $\delta_{YX} = 75$. These delays were recovered erroneously as $\hat{\delta}_{XY} = 28$ and $\hat{\delta}_{YX} = 61$. For more parameters see table 2.

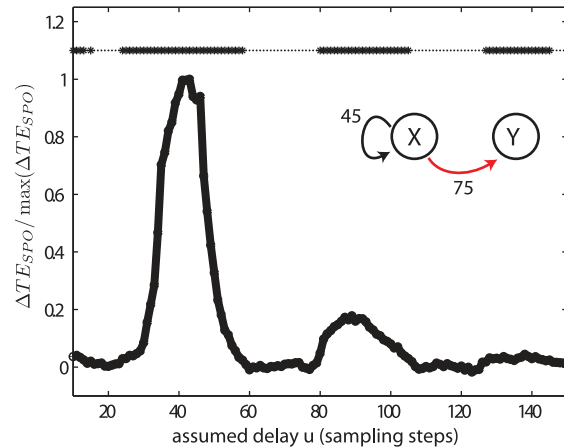doi:10.1371/journal.pone.0055809.g010

Figure 11. Test case (VI) - self-feedback analysis. Transfer entropy (TE_{SPO}) values between past and present of one of two Lorenz systems (X) and significances as a function of the assumed delay u. The analyzed chaotic Lorenz system was subject to a feedback loop with delay $\delta_{XX} = 45$, and an outgoing interaction $X \rightarrow Y$ with delay $\delta_{XY} = 75$, but no incoming interaction. The recovered delay for the self feedback was $\hat{\delta}_{XX} = 43$, with a sidepeak at around two times this value. For the interaction analysis $X \rightarrow Y$ see figure 12. For more parameters see table 2.

doi:10.1371/journal.pone.0055809.g011

cover the feedback delay. The results are shown in Figure 11. Clear peaks are observed at $u = 45$, which corresponds to the simulated feedback delay δ_{XX}, and its integer multiples. This corresponds to information being fed back via multiple rounds of the feedback loop.

The presence of feedback loops can challenge the reliable detection of information transfer between nodes. For example, if a node subject to feedback is unidirectionally coupled to another node, the direct computation of transfer entropy between the two nodes as a function of a delay parameter can lead to wrong inferences. To illustrate this point, we computed the self prediction optimized transfer entropy between the Lorenz system (X) with a feedback loop and another Lorenz (Y) which receives its output.

First, we consider test case (VI) in which the feedback loop time ($\delta_{XX} = 45$) is shorter than the interaction delay between the systems ($\delta_{XY} = 75$). As shown in Figure 12, $TE_{SPO}^{X \rightarrow Y}$ has the highest peak at $u = 75$, corresponding to the simulated interaction delay. However, it also shows clear peaks at $u = 30$ and $u = 120$, values which amount to the difference between the interaction and feedback delays, and their sum respectively. In general, peaks were observed at $u = \delta_{XY} + n\delta_{XX} > 0; n \in \mathbb{Z}$, corresponding to different combinations of cycles around the feedback loop plus the interaction delay.

TE_{SPO} from Y to X, the direction in which no coupling was simulated, also exhibits several weak but significant peaks, i.e. we find false positive results in this case. Peaks were present at multiples of the delay feedback time minus the propagation time between the two Lorenz systems ($u = n\delta_{XX} - \delta_{XY} > 0; n \in \mathbb{Z}$). These peaks in TE_{SPO} from Y to X appear because the feedback loop in process X results in recurrent information in the dynamics of X that can be predicted by the process Y, because thew process Y also receives a copy of them via the connection $X \rightarrow Y$. This information is useful to predict the state of X when the assumed delay u in TE_{SPO} from Y to X is at least as big as $u = n\delta_{XX} - \delta_{XY}$, with n chosen such that u is positive. Notice that the size of the peaks decreases with larger n.

Second, we considered test case (VII) in which the feedback delay δ_{XX} is longer than the interaction delay time δ_{XY} (see

Figures 13 and 14). In this case structure similar to test case (VI) is observed for the location of the peaks of TE_{SPO}. However, $TE_{SPO}^{Y \rightarrow X}$ shows higher and more false positive peaks than in case (VI). This occurs since when $\delta_{XX} > \delta_{XY}$, process Y can predict the
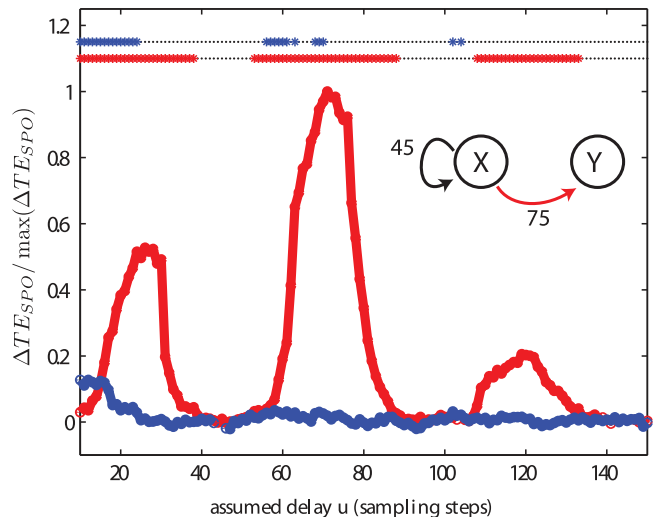

Figure 12. Test case (VI). Transfer entropy (TE_{SPO}) values and significance as a function of the assumed delay u for a unidirectionally coupled chaotic Lorenz systems. The first Lorenz is subject to a feedback loop ($\delta_{XX} = 45$) and unidirectionally couples to a second Lorenz with a interaction delay of $\delta_{XY} = 75$ samples. Recovered delays were $\hat{\delta}_{XX} = 43$ (see figure 11), and $\hat{\delta}_{XY} = 73$. Sidepeaks were observed for $\hat{\delta}_{XY}$ close to $\delta_{XY} \pm n \cdot \delta_{XX}$. Spurious interactions were observed in the reversed direction at $\hat{\delta}_{YX} = n \cdot \delta_{XX} - \delta_{XY}$, as it is expect for a system with self feedback [45]. Considering the positive test for self-feeback (figure 11) and the recovery of the self-feedback delay, the true system connectivity can be derived by combining the analysis of self-feedback and interaction delays.

doi:10.1371/journal.pone.0055809.g012

transitions that will occur in X already after a single delay loop, because even for $n=1$ the condition $u=n\delta_{XX}-\delta_{XY}>0$ is fulfilled, – in contrast to a prediction of two delay loops ahead as in the previous case above. This situation is related to the so-called anticipative synchronization in which a slave system (Y) can anticipate the dynamics of the master system when this is subject to a long feedback loop [44,45].

To be clear: these observations are not a theoretical problem with $TE_{SPO}^{Y\rightarrow X}$ but are a practical issue in estimation (due to the impracticality of adequately forming Taken's embedding of X in the presence of the long delay loop, as described above).

Ring of Lorenz systems. In a network of three Lorenz systems coupled into a unidirectional ring, test case (VIII), our method identified the three simulated delays $\delta_{XY}=20$, $\delta_{YZ}=30$, $\delta_{ZY}=40$ with reasonable precision as $\hat{\delta}_{XY}=21$, $\hat{\delta}_{YZ}=28$, $\hat{\delta}_{ZX}=38$ (figure 15). Analysis of self-feedback (as it is in principle present in a ring structure) for system X resulted in no significant peak at the expected sum of all three simulated delays (90), indicating that the information originally transfered from system 1 into the ring is *effectively* wiped out by the chaotic dynamics of the next nodes in the ring, a phenomenon well known in from coupled chaotic laser systems [46].

Effects of observation noise. In test case (IX) we simulated two bidirectionally, quadratically coupled Lorenz systems with delays $\delta_{XY}=45$, $\delta_{YX}=75$, and added independent, Gaussian, white noise to the time series of the V-coordinate (see equations 32, 33 for details) before the reconstruction of delays. Observation noise did degrade the precision of delay reconstruction to a certain degree: with 1%, 2% and 9% of the total signal variance contributed by noise, the estimated delays were $\hat{\delta}_{XY}=45,40,42$, and $\hat{\delta}_{YX}=70,72,70$ (figure 16). Note that noise amplitude and delay reconstruction error do not seem to be systematically related, suggesting that the effects of particular realizations of finite data cause the reconstruction errors.
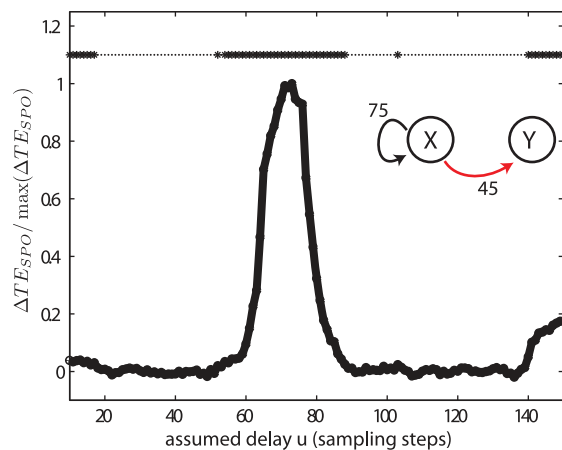

Figure 14. Test case (VII). Transfer entropy (TE_{SPO}) values and significance as a function of the assumed delay u for a unidirectionally coupled chaotic Lorenz systems. The first Lorenz is subject to a feedback loop ($\delta_{XX}=75$) and unidirectionally couples to a second Lorenz with a interaction delay of $\delta_{XY}=45$ samples. Recovered delays were $\hat{\delta}_{XX}=74$ (see figure 13), and $\hat{\delta}_{XY}=44$. Sidepeaks were observed for $\hat{\delta}_{XY}$ close to $\delta_{XY}\pm n\cdot\delta_{XX}$. Spurious interactions were observed in the reversed direction at $\hat{\delta}_{YX}=n\cdot\delta_{XX}-\delta_{XY}$, as it is expect for a system with self feedback [45]. Considering the positive test for self-feeback (figure 13) and the recovery of the self-feedback delay, the true system connectivity can be derived by combining the analysis of self-feedback and interaction delays.
doi:10.1371/journal.pone.0055809.g014

Local Field Potential Data

To demonstrate that interaction delays can be reconstructed from biological time series with sufficient precision, we analyzed recordings of the electroretinogram (R) and local field potentials from the tectum (T) of the turtle brain (*Pseudemys scripta elegans*, figure 17). These data were recorded during stimulation (S) with light flashes at time points determined by a random process and with a duration drawn from a uniform random distribution; this ensured stationarity of the time series. In this experiment, direct physical interactions existed from the light source to the retina and from the retina to the tectum ($S\rightarrow R\rightarrow T$). In addition, there was an indirect interaction from the light source to the tectum, mediated by the retina. This opens the possibility to check the precision of the delay reconstruction despite the fact the we have no precise knowledge of the true biological interaction delays. The evaluation of reconstruction precision is possible because the interaction delays reconstructed from $S\rightarrow R$ and from $R\rightarrow T$ should sum up to the interaction delay reconstructed from $S\rightarrow T$ if reconstruction is precise.

We reconstructed the interaction delays the same way as in all previous test cases by scanning u. Indeed, the reconstructed delays were: $\hat{\delta}_{(S,R)}=30$ms, $\hat{\delta}_{(R,T)}=12$ms, $\hat{\delta}_{(S,T)}=44$ ms, meaning that the reconstructed delays between light source and retina and between retina and tectum added up to the reconstructed delay between light source and tectum with an error of 2 ms or 4.5%.

Discussion

Transfer Entropy Estimation without Violating Wiener's Principle

We have laid out in the introduction why the earlier formulation of transfer entropy with an explicit time-delay as given in equation 3 and as used in [9,12,13,29] is not a precise formulation of

Figure 13. Test case (VII) - self-feedback analysis. Transfer entropy (TE_{SPO}) values between past and present of one of two Lorenz systems (X) and their significances as a function of the assumed delay u for a single chaotic Lorenz system subject to a feedback loop with delay $\delta_{XX}=75$, and an outgoing interaction $X\rightarrow Y$ with delay $\delta_{XY}=45$. The recovered delay for the self feedback was $\hat{\delta}_{XX}=74$, with a sidepeak at two times this value. For the interaction analysis $X\rightarrow Y$ see figure 14. For more parameters see table 2.
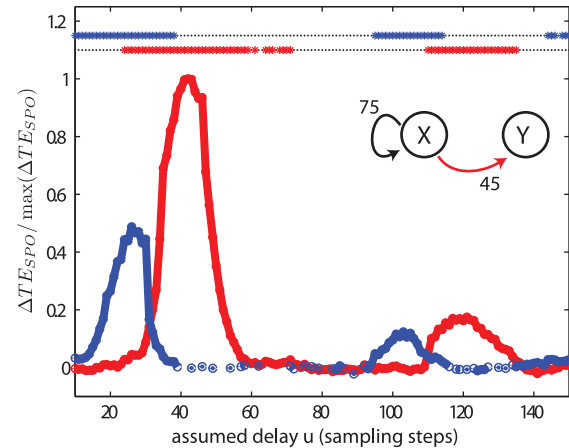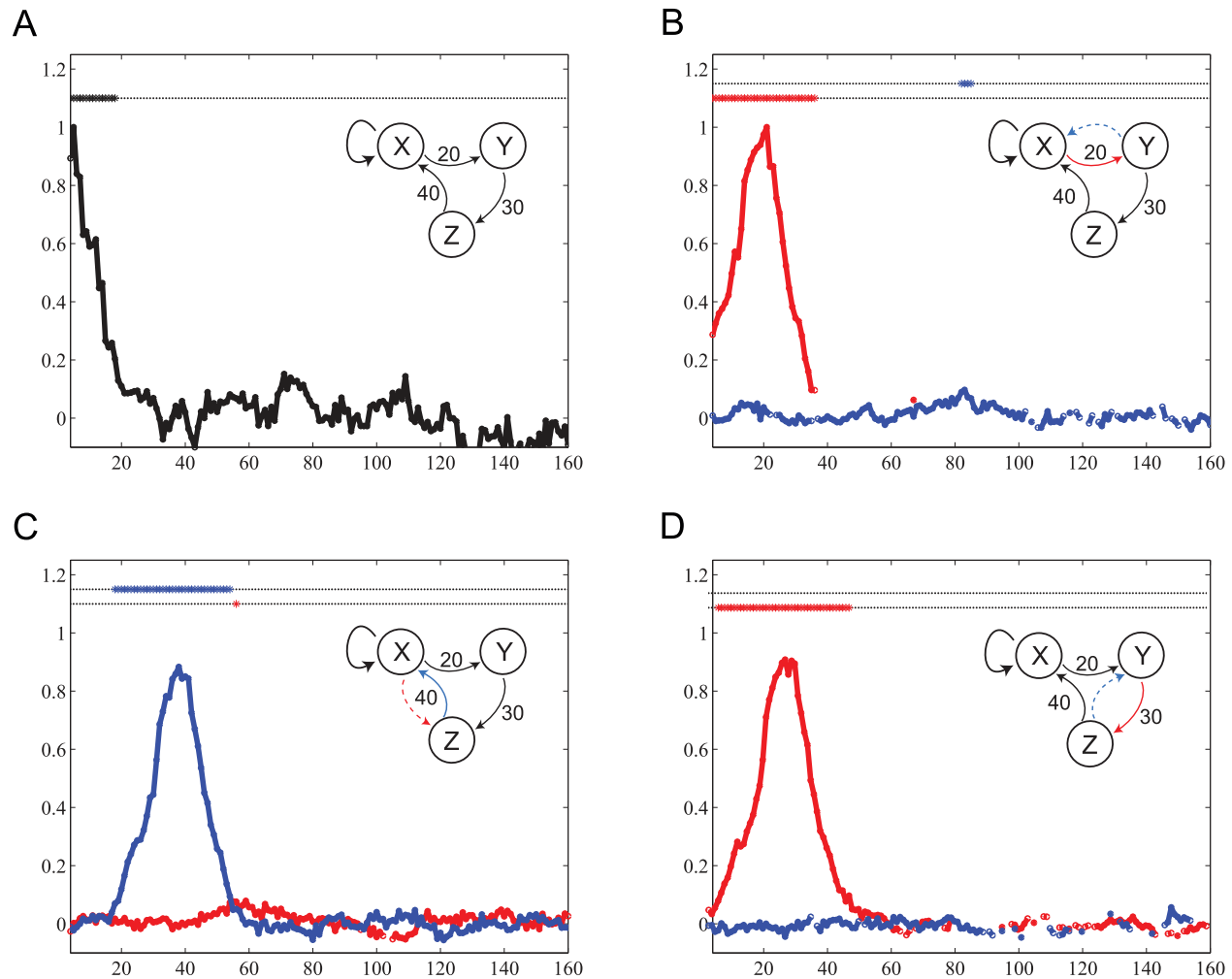doi:10.1371/journal.pone.0055809.g013

Figure 15. Test case VIII. Transfer entropy (TE_{SPO}) values and significance as a function of the assumed delay u for three unidirectionally coupled chaotic Lorenz systems. The First Lorenz couples with the second Lorenz with an interaction delay of $\delta_{XY} = 20$ samples, the second Lorenz is unidirectionally coupled with the third Lorenz at a delay of $\delta_{YZ} = 30$ samples and the third Lorenz is unidirectionally coupled with the first Lorenz at an interaction delay of $\delta_{ZX} = 40$ samples. The reconstruction of the simulated delays were: (A) self feedback, $\hat{\delta}_{XX} = 2$, this value may be due to insufficient embedding, (B) $\hat{\delta}_{XY} = 21$, (C) $\hat{\delta}_{YZ} = 28$, and (D)$\hat{\delta}_{ZX} = 38$.
doi:10.1371/journal.pone.0055809.g015

Wiener's principle of causality in information-theoretic terms, as it violates the requirement of optimal self-prediction of the target time series. Accordingly, we were able to construct a test case were this functional gives a wrong estimate of the interaction delay in a system (figure 10), while the novel functional proposed here (equation 7), gives the correct result (figure 8). We therefore suggest to use this novel functional wherever interaction delays are expected. Accordingly, the new functional has been implemented in version 2.0 of our open source toolbox for transfer entropy estimation, TRENTOOL [42], and has been made the default option.

Interaction Delay Reconstruction by Maximizing Predictive Information Transfer

In this study we demonstrated that it is possible to reconstruct the delays of interactions between two systems by finding the maximum of the predictive information transfer estimated by a novel transfer entropy functional, TE_{SPO}, with a parametric dependence on the interaction delay. Our work complements

earlier, anecdotal reports of delay-sensitivity of the predictive information transfer [12,13,42], by presenting a new formulation of the transfer entropy functional rigorously based on Wiener's principle of causality and *backed by a formal proof* (see results section). Crucially, our experimental results identify the precise interaction delay for coupled systems with a single interaction delay, validating the formal proof which was constructed under these conditions. Furthermore though, we show that the reconstruction of interaction delays is possible for a large range of coupling types, multiple interaction delays, complex dynamics of the subsystems, for ring, and bidirectionally coupled systems. This last point is of great importance, as up to now, the analysis of bidirectionally coupled systems has often been discouraged, at least implicitly. Hesitation to analyze bidirectionally coupled systems is based on two observations - on the one hand, bidirectional coupling often leads to complete synchronization and in these cases an analysis of interactions is indeed not possible - on the other hand, it has been shown that the quantification of causal effect *size* is not always possible in these systems [47]. With our finding that in cases where
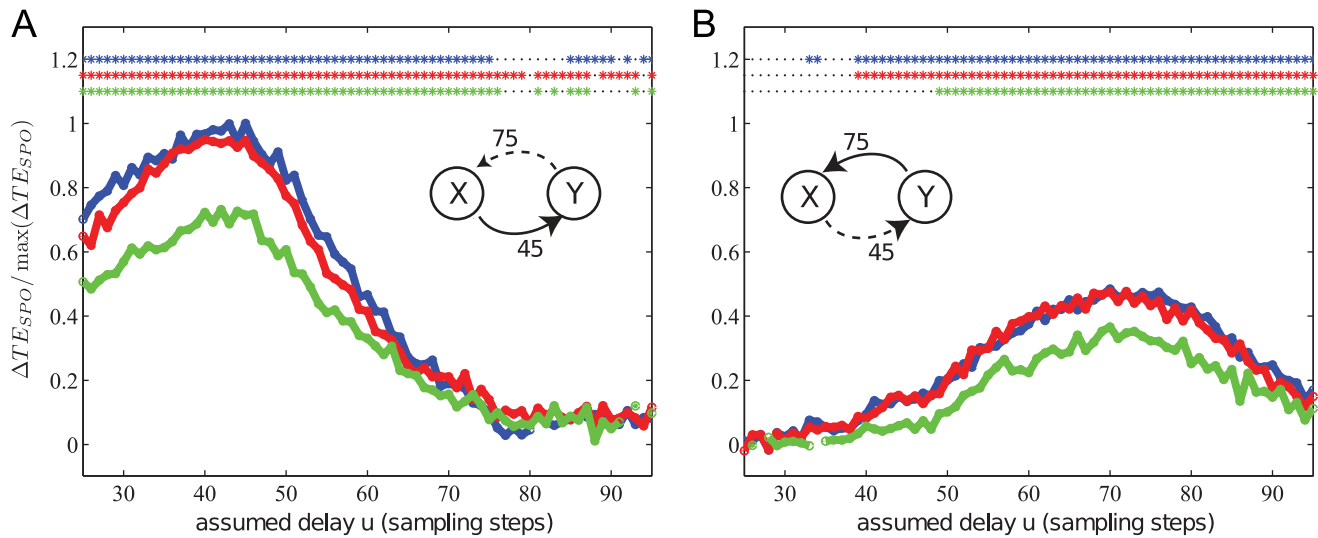
Figure 16. Test case IX. Transfer entropy (TE_{SPO}) values and significance as a function of the assumed delay u for two bidirectionally coupled, chaotic Lorenz systems. The simulated delays were $\delta_{XY} = 45$ and $\delta_{YX} = 75$. Observation noise with different amplitude was added to the simulated time series of the processes. The delays were recovered as (A) $\hat{\delta}_{XY} = 45$ and $\hat{\delta}_{YX} = 70$ for $S/N = 45.88db$ (blue), $\hat{\delta}_{XY} = 40$ and (B) $\hat{\delta}_{YX} = 72$ for $S/N = 37.74db$ (red) and $\hat{\delta}_{XY} = 42$ and $\hat{\delta}_{YX} = 70$ for $S/N = 23.84db$ (green).
doi:10.1371/journal.pone.0055809.g016

Figure 17. Interaction delay reconstruction in the turtle brain. (A) Electroretinogram (green), and LFP recordings (blue), light pulses are marked by yellow boxes. (B) Schematic depiction of stimulation and recording, including the investigated interactions and the identified delays.
doi:10.1371/journal.pone.0055809.g017

there is no complete synchronization we can reconstruct at least the individual delays of bidirectional coupling (again backed by our formal proof), we hope to revive the analysis of such systems - that are abundant in nature and technology.

Formally identical functionals to TE_{SPO} have been independently introduced several times in the literature, first by Nichols and colleagues [48], then by Overbey and Todd [49] – both in the field of structural integrity analysis of mechanical systems. In addition, Ito and colleagues used a formally identical functional to increase the detectability of interactions in spiking neural data [50]. In none of these studies the use of the functional for explicit delay reconstruction has been recognized and as a consequence no proofs for this property have been given. Ito and colleagues did indeed state that the delay parameter in their equation is there to account for finite delays, but they simply assumed maximality of the functional at the correct delay, without proof.

Note that the successful reconstruction of the true interaction delay between two coupled systems depends on reliable and precise enough estimates of the corresponding information-theoretic quantities (see methods section for the algorithms applied here). Obtaining these estimates may become a problem for small sample sizes. In this case, additional statistical testing against the null hypothesis of no coupling should be included when scanning delays and only maxima that show statistically significant coupling should be evaluated.

In addition, stationarity of the time series entering the analysis must be given, because transfer entropy between two random processes is typically defined via a time average that can only be used instead of the proper ensemble average for random processes if stationarity is given. If stationarity cannot be assumed a priori for the time series under investigation, appropriate testing should be performed (see [51,52] and references therein). If at least cyclostationarity can be guaranteed, the proposed method could be used as functional within the ensemble averaging framework described in [53]. Reliably repeated non-stationarities may also be removed using the mathematical methods presented in [54]. In this case, however, additional testing for remaining non-stationarities is recommended. Note that for the local field potential data
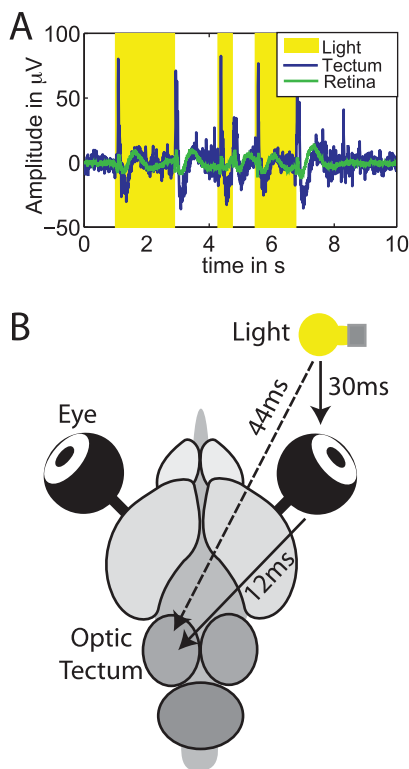
analyzed here, approximate stationarity was guaranteed by stimulation at multiple randomly chosen time points within each epoch of the experiment time series (i.e. "trial"), with stimulation durations also drawn from a random distribution.

Comparison to the Momentary Information Transfer

As detailed in the methods section, *momentary information transfer* (MIT) [36] is an alternate approach to reconstructing the interaction delay, again using a scanning approach to maximize MIT as a function of delay u. Both TE_{SPO} and MIT condition on the immediately previous target state (which is correct in comparison to the old TE estimator), and given that TE_{SPO} uses an embedded source state $\mathbf{X_{t-u}}$, both measures include *synergistic* information that is jointly contributed by $\{X_{t-u}, \mathbf{X_{t-u-1}}\}$ (as opposed to TE computed using only a single source *observation*). Yet a stark difference is the extra conditioning by MIT on $\mathbf{X_{t-u-1}}$, which removes *redundant* information that was already contained in the source. This prevents any of this information from being attributed to transfer at the lag u.

A major conceptual difference between the Pompe and Runge study and ours is that no formal proof of the maximality of their functional MIT at the correct interaction delay is given, and as we argue below cannot be given.

Indeed, we provide a counter-example here – in test case (Ia) – where the momentary information transfer is not necessarily maximized at the correct interaction delay, in direct contrast to our proof of this property for TE_{SPO}. As such, MIT is not always reliably inferring the correct interaction delay.

The intention of the MIT in removing the self-redundant information in the source was to find the delay at which the relevant information about the target's state update first appeared in the source. However, the availability of such information in the source at a specific time point does not mean that it is being transferred at that instant, and the presence of *memory* in the source inevitably leads the MIT to underestimate the influence of the source at the actual interaction delay (because of the removal of *redundant*, though potentially transferred, information). We demonstrated for these reasons that the MIT failed to identify the correct interaction delay in test case (Ia) on a simple, unidirectional coupling, single-interaction delay example with source memory, whereas our TE_{SPO} estimator functioned correctly.

As argued in [55], when investigating effective connectivity, the removal of redundant information by prematurely conditioning on other potential sources reduces our ability to infer connections. Conditioning on the source at another delay(s) should only be performed once one has already established the primary (or additional) coupling delay (see further discussion below and in [34,55]).

Furthermore, the actual MIT analysis in [36] was carried out on a symbolic mapping of the time series (to vectors capturing the ordinal relationships between the variables). This approach, while potentially useful for short time-series realizations in magnifying certain relationships in the data, removes information on absolute values of the variables involved. While this loss of information may be harmless in climatology – the field of research MIT was aimed at –, it might be crucial to retain this information in other fields, such as neuroscience. Indeed, for the example in their study ([36], Section V.A) where the TE_{SPO} was not able to infer the correct interaction delay but *MIT* was, we have shown in test case (Ib) that this result is an artifact of the symbolic mapping in [36] rather than the measures themselves. Specifically, when we estimated in the continuous domain (retaining much information that the symbolic mapping removed), TE_{SPO} correctly identifies the interaction lag in this coupled logistic map process.

Interaction Delay Fingerprints in Systems with Self-feedback

The phenomenon of self feedback is present in many complex networks. This self feedback can arise genuinely in the nodes of the network, e.g. by mechanisms such as autapses in neural networks [56], or because the systems receives (transformed) self-feedback via an unobserved part of the network. The analysis of interaction delays via TE_{SPO} may offer valuable hints with respect to the presence of absence of relevant self-feedback.

Here we showed that TE_{SPO} can detect the presence of feedback loops when applied to the time series of a single system and its own past – even if the system is chaotic and the feedback loops entail nonlinear transformations of the systems output. The information on the delay of a feedback loop can then be used to disentangle the potentially complex delay fingerprints, consisting of multiple peaks, that arise if such a system with self-feedback is coupled to other systems.

In principle, the multiple spurious peaks in the direction $X \to Y$ in our example, and all of the peaks in the direction $Y \to X$ should vanish, given a state space reconstruction with states long enough to cover the delay feedback time. However, very long reconstruction lengths might lead to instable estimates due to the 'curse of dimensionality' [57]. In such a case, where the practically feasible reconstruction lengths for states are too short to cover the full dynamics of a system, the identification of feedback loops by TE_{SPO} helps to better interpret the estimated pairwise information transfers between the nodes of a network. Time points around the self-feedback delay could in addition be included into the embedding states of a *non-uniform*, data-efficient embedding scheme [34], which would assist in removal of information storage and more accurate assessment of the transfer entropy as per the principles outlined in subsection on Wiener's principle above.

TE_{SPO} differs from a simple application of the lagged auto-mutual information functional (AMIF) in that active information storage $I(X_t; \mathbf{X_{t-1}})$ [37] contained in the most recent reconstructed state $\mathbf{X_{t-1}}$ is removed. This will accentuate the presence of peaks in the delay spectrum compared to AMIF.

Information Transfer Delays from Noisy Time Series

The proof for the identifiability of the true delay in the information transfer between two time series holds strictly only for the case of zero *observation* noise. Indeed the fundamental differences between time series obtained from Markov systems and those obtained from hidden Markov systems (i.e. from noisy observations) make it difficult to extend our proof without specifying the noise explicitly in each case. Nevertheless, our simulations of typical noise influences – such as independent, Gaussian, white noise – show that our approach works well in practice. While noise does indeed degrade precision, the random relation between observation noise amplitude and delay reconstruction error suggest that these errors are due to the combined effects of finite data and noise, and can be alleviated by increasing sample size.

Relation to Linear Granger Causality and Corresponding Time Delay Reconstruction Procedures

Recently, Barnett and colleagues [19] demonstrated that transfer entropy and linear Granger causality are equivalent for the case of data with a Gaussian distribution. This result greatly simplifies the computation of interactions for data of this class. Neural data, however, do most likely not have a Gaussian distribution. This can for example be seen when comparing brain electrical source signals from physical inverse methods with the

time course of corresponding ICA components [58]. Given that ICA components are as non-Gaussian as possible by definition, and given the fact that ICA components and brain source signals extracted by inverse methods closely match, we can interpret this as evidence for a non-Gaussian nature of brain signals. For these signals, TE may have an advantage for the analysis of directed interactions. On the other hand, the methodology presented here should be transferable to the domain of linear Granger causality in Gaussian data by virtue of the proof by Barrett and colleagues. Hence, the approach presented here may be seen as an alternative to earlier attempts to infer timing delays via linear Granger causality by inspecting time-dependent model-coefficients and using large model orders of 200 and more. For Gaussian data the scanning approach presented here would be equivalent to setting the first n model coefficients to zero without having to estimate them, and scanning n, estimating the k next model coefficients, where k is the embedding dimension, resulting from optimization via the Ragwitz criterion [39].

Relation of Delay Reconstruction and Multivariate TE Analyses

In systems composed of more than two interacting subsystems a pairwise, bivariate transfer entropy analysis as given by equation 7 may lead to wrong inferences with respect to the presence of an interaction between two subsystems. This happens when either a third subsystem drives the two subsystems under investigation with differential delays ('common drive'), or when the two subsystems under investigation are connected indirectly via a third system acting as a relay ('cascade effect'). The *potential presence* of interaction configurations of this kind can be detected by looking at timing relations across the graph of bivariate interactions. Here, this was demonstrated by our reconstruction of delays in LFP data in the turtle, where the interaction delay between light source and tectum was equal to the sum of delays on a route connecting the same end points (light source to tectum, via the retina). While both a common drive (light source → retina, light source → tectum) or cascade scenario (light source → retina → tectum) could explain the manifestations of these delays, what is important here is that this result is consistent with the known biological indirect interaction via the retina. If the concern is just to avoid false positive detection of interactions, a simple delay analysis approach may be a data efficient alternative to fully multivariate treatments of TE [7,34,35,55]. That said, such multivariate treatments could yield further insights, e.g. if the *conditional* TE [7,35] from the retina to the tectum conditioned on the light source (with appropriate delays incorporated) were statistically non-zero, then this would eliminate the possibility that the inferred retina → tectum relationship was a result of common drive by the light source. Combining delay analysis with multivariate treatments is feasible but more complicated (delays must be determined in an appropriate order, in the same way that the self-conditioning delay u_0 was determined here before the source delay u was explored - see some relevant discussion in [34,55]); and will be the subject of a future publication.

Delay Estimation Versus Significance Testing

We would like to stress here that inference on the presence of information transfer is a task separate from reconstructing the delay of the information transfer. For the former task we employed nonparametric permutation testing of TE_{SPO} against surrogate data obtained from exchanging data epochs in the source. For the latter we searched the peak of TE_{SPO}. Our simulation results demonstrate that the existence of information transfer can typically be assessed over a wide range of assumed delays. On

the one hand, this fact underlines the robustness of the TE_{SPO}-functional against misspecification of the delay parameter. On the other hand it is a warning not to conclude from the presence of significant information transfer at a certain delay parameter that the true delay is close to the assumed delay, if no scanning of the delay parameter was performed.

A Practical Note on the Combination of Delay Reconstruction and Shift Testing

In practical applications, linear mixing or instantaneous cross-talk between signals occurs and may bias interaction measures based on Wiener's principle [12,59]. To detect such crosstalk, we have proposed a so-called shift-test [12,13,42]. This test determines if shifting the source time series into the past by the time that represents the assumed delay, u, increases TE. This way two predictive information transfer terms get compared: on the one hand the predictive information transfer from the past of the source to the target, on the other hand the instantaneous information transfer from source to target. This procedure works very well as demonstrated in [42], but gives rise to minor a technical problem for real world data, because in these data instantaneous cross-talk is never truly zero. For assumed delays u that are much larger than the true delay, the information transfer will be arbitrarily small, due to the combined effects of conditioning on the past of the target and the finite memory of the source. In contrast, cross-talk will always be non-zero. Hence, in situations with some finite cross-talk, there will a largest delay parameter u_{max}, beyond which the presence of crosstalk will always be reported by the shift test. This, however, can be easily seen by comparing TE values at large u to those at smaller u. In contrast to a mere nuisance effect, relevant cross-talk should outweigh TE_{SPO} even at the optimal delay.

Conclusion

We present a novel transfer entropy functional, which is a rigorous formulation of Wiener's principle of causality in information-theoretic terms, respecting the condition of optimal self-prediction of the target time series from its own past. This functional has an explicit parametric representation of interaction delays between interacting systems. Scanning this parameter in search of the maximal predictive information transfer allows one to reconstruct interaction delays from a wide variety of systems.

Methods

Ethics Statement

Local field potential data were taken from experiments published elsewhere [42]. These animal experiments were approved by the German local authorities (Regierungspraesidium, Hessen, Darmstadt).

Practical Transfer Entropy Estimation

In this section we outline the particular estimator of the TE_{SPO} functional as provided in our toolbox TRENTOOL [42], and used in all analysis in this study. This realization relies on three steps: (1) state space reconstruction from scalar time series, (2) reformulation of the conditional mutual information in terms of four Shannon entropies, and (3) subsequent entropy estimation by a modified Kraskov-Stoegbauer-Grassberger estimator [41,42].

As causality and interactions are defined as properties of systems, not scalar time series, we first have to reconstruct the corresponding state space of the interacting systems from the scalar time series. For this purpose we used Takens delay

embedding [43] and optimized embedding parameters (d and τ, see below) according to Ragwitz' criterion [39] for the target signal of each interaction pair. The use of Ragwitz' criterion yields delay embedding states that provide optimal self prediction for a large class of systems, either deterministic or stochastic in nature. Delay embedding states of the systems under investigation can be written as delay vectors of the form:

$$\mathbf{x}_t^d = (x_t, x_{t-\tau}, x_{t-2\tau}, \ldots \\ , x_{t-(d-1)\tau}),$$ (16)

where d and τ denote the embedding dimension and Taken's embedding delay, respectively.

Using the states obtained by delay embedding we can rewrite transfer entropy as:

$$TE_{SPO}(X \to Y, u) = \sum_{y_t, \mathbf{y}_{t-1}^{dy}, \mathbf{x}_{t-u}^{dx}} p\left(y_t, \mathbf{y}_{t-1}^{dy}, \mathbf{x}_{t-u}^{dx}\right) \\ \log \frac{p\left(y_t | \mathbf{y}_{t-1}^{dy}, \mathbf{x}_{t-u}^{dx}\right)}{p\left(y_t | \mathbf{y}_{t-1}^{dy}\right)},$$ (17)

or, using a representation in the form of four Shannon (differential) entropies, as:

$$TE_{SPO}(X \to Y, u) = S\left(\mathbf{y}_{t-1}^{dy}, \mathbf{x}_{t-u}^{dx}\right) - S\left(y_t, \mathbf{y}_{t-1}^{dy}, \mathbf{x}_{t-u}^{dx}\right) \\ + S\left(y_t, \mathbf{y}_{t-1}^{dy}\right) - S\left(\mathbf{y}_{t-1}^{dy}\right).$$ (18)

Thus, TE_{SPO} estimation amounts to computing a combination of different joint and marginal differential entropies. Shannon differential entropies can be estimated by nearest-neighbor techniques that exploit the statistics of distances between neighboring data points in a given embedding space in a data efficient way. This efficiency is necessary to estimate entropies in high-dimensional spaces from limited real data [60,61]. Nearest-neighbor estimators are as local as possible given the available data. The assumption behind nearest-neighbor estimators is only a certain smoothness of the underlying probability distribution. Nearest-neighbor estimators can therefore be considered as non-parametric techniques, as desired for a model-free approach to transfer entropy estimation. Unfortunately, it is problematic to estimate TE by simply applying a nearest-neighbor estimator (e.g. Kozachenko-Leonenko estimator) separately to each of the terms appearing in equation 5. The reason is that the dimensionality of the spaces involved in equation 18 can differ largely across terms. Thus, fixing a given number of neighbors for the search will set very different spatial scales (range of distances) for each term. Since the error bias of each term is dependent on these scales, the errors would not cancel each other but accumulate. We therefore used the Kraskov-Grassberger-Stögbauer estimator which handles this problem by only fixing the number of neighbors k in the highest dimensional space and by projecting the resulting distances to the lower dimensional spaces as the range to look for neighbors there [41]. After adapting this technique to the TE formula [53], the estimator we use can be written as

$$TE(X \to Y, u) = \psi(k) + \left\langle \psi\left(n_{\mathbf{y}_{t-1}^{\mathbf{dy}}} + 1\right)\right. \\ - \psi\left(n_{y_t \mathbf{y}_{t-1}^{\mathbf{dy}}} + 1\right) \\ \left. - \psi\left(n_{\mathbf{y}_{t-1}^{\mathbf{dy}} \mathbf{x}_{t-u}^{\mathbf{dx}}} + 1\right)\right\rangle_t,$$ (19)

where ψ denotes the digamma function, while the angle brackets $\langle \cdot \rangle_t$ indicate an averaging over different time points. The distances to the k-th nearest neighbor in the highest dimensional space (spanned by $y_t, \mathbf{y}_{t-1}^{dy}, \mathbf{x}_{t-u}^{dx}$) define the radius of the spheres for the counting of the number of points $n_{(\cdot)}$ in these spheres around each state vector in all the marginal spaces $(\cdot)$ involved.

Non-parametric Statistical Testing Against Surrogate Data

Even using Kraskov's kernel estimation techniques as described above does not guarantee zero bias of the resulting estimator. Thus, the obtained TE values have to be compared against suitable surrogate data using non-parametric statistical testing to infer the presence or absence of directed interactions [12]. In short, the surrogate data must be produced under a null hypothesis of no source-target coupling, while retaining as many other statistical properties as possible (in particular the state transition probabilities $p(y_t | \mathbf{y}_{t-1}^{dy})$). To this end we simulated and recorded data in an epoch based way and constructed surrogate data by shifting the time series of one of the two signals of each pair by one epoch, trying to preserve as many data features as possible (see detailed descriptions of the statistical routines in [42]). TE values were quantified as excess TE values with respect to surrogate data:

$$\Delta TE_{SPO}(X \to Y, u) = TE_{SPO}(X \to Y, u) - TE_{SPO}(X' \to Y, u),$$ (20)

where X' denotes the surrogate data. With respect to these surrogate data we also obtained significance values using permutation testing against suitable surrogate data as detailed in [42], to minimize the potential effects of bias introduced by noise and small sample size. Note that we assess statistical significance and quantify the excess transfer entropy ΔTE_{SPO} in both possible directions of interactions ($X \to Y, Y \to X$), although interactions in some cases were unidirectional. By testing both directions nevertheless, we can also characterize the behavior of the proposed estimator with respect to false positive detection of interactions.

Momentary Information Transfer

Pompe and Runge [36] recently proposed to reconstruct interaction delays using an information-theoretic functional, called *momentary information transfer (MIT)*. In their functional the interaction delay between two systems is also introduced in the form of a parameter of a conditional mutual information term. As for TE_{SPO} this parameter is scanned in order to maximize the value of their functional, MIT. In contrast to our approach, conditioning of the mutual information in the method of Pompe and Runge is done with respect to the joint history of the *two* variables in question:

$$MIT(X \to Y, u) = I(Y_t; X_{t-u} | \mathbf{Y}_{t-1}, \mathbf{X}_{t-u-1}),$$ (21)

That is, while it retains the conditioning on the immediately previous state of the target $\mathbf{Y_{t-1}}$ that we use in TE_{SPO}, it *additionally* conditions on the state of the source $\mathbf{X_{t-u-1}}$ previous to the scalar source observation under consideration. The essence of Pompe and Runge's argument is that their conditioning on $\mathbf{X_{t-u-1}}$ seeks to find the delay over which the transferred information is first available in the source, though we note that, as explored in earlier sections, the availability of such information in the source does not equate to it being transferred at that point. We note, that for our study, MIT was measured using discrete probability distribution functions, not the symbolic mapping (capturing ordinal relationships) that it is often associated with in the case of continuous data.

Test Cases

We used simulated data and electrophysiological recordings to test the ability of the described methodology to detect interaction delays. The generation and characteristics of the studied time series are described below; note that we chose a bracket $(x(t))$ instead of a subscript notation (x_t) for time dependencies for this section to avoid cluttered subscripts. All analyses were performed in TRENTOOL (version 2.0.3, [42]) unless otherwise noted. The Ragwitz criterion [39] was used to determine the embedding dimension d and lag τ. We used a significance level of 0.05, and corrected for multiple comparisons via false discovery rate (FDR [62]), to assess significance of the coupling. To identify interaction delays, we scanned the source delay parameter u from 10 up to 150 time steps in steps of 1 sample.

Discrete-value process with short-term source memory. Test case (Ia) is formed from coupled discrete-valued processes X, Y, where $X \in \{0,1,2,3\}$ and $Y \in \{0,1\}$, which were generated according to the equations:

$$Y(t+1) = X(t) \bmod 2, \tag{22}$$

$$X(t) = f(X(t-1), \eta), \tag{23}$$

where f is a noisy self-mapping of X to its next value defined in Table 1 with noise parameter η. Note that f incorporates some randomness and some stochastic short-term memory in the next state. The update function f can be explained very simply if we consider X as a joint variable of upper and lower bits $\{X_2, X_1\}$, and understand that $X_2(t)$ is randomly determined at each time step, while $X_1(t)$ is a copy of $X_2(t-1)$ with probability $1-\eta$, otherwise it is the inverse of $X_2(t-1)$. Using this interpretation, we have $Y(t+1) = X_1(t)$.

Here, the true causal delay $X \to Y$ is 1 time step, though the source memory in X means that X and Y will be strongly correlated over a 2 time step delay also. In this case, we only examined these two candidate delays. Clearly, the Ragwitz criterion is satisfied here with embedding dimension $d=0$ (there is no auto-correlation between values of Y).

As this system is discrete valued, MIT was measured using discrete probability distribution functions here, not the symbolic mapping (capturing ordinal relationships) that it is often associated with for continuous data.

Bidirectionally coupled logistic map. Test case (Ib) is taken from Section V.A of [36]. It is formed from the bidirectionally coupled logistic map processes X, Y, where we have X and $Y \in [0,1]$, which were generated according to the equations:

$$X(t) = f(g_{y \to x} \bmod 1), \tag{24}$$

$$g_{y \to x} = c_{y \to x} Y(t - \delta_{y \to x}) + (1 - c_{y \to x}) X(t-1), \tag{25}$$

$$Y(t) = f(g_{x \to y} \bmod 1), \tag{26}$$

$$g_{x \to y} = c_{x \to y} X(t - \delta_{x \to y}) + (1 - c_{x \to y}) Y(t-1), \tag{27}$$

$$f(a) = 4a(1-a). \tag{28}$$

The process is run using the same parameters as in [36]: $T = 512$ samples, 1000 repeated trials (results averaged over trials), $\delta_{y \to x} = 5$, $\delta_{x \to y} = 2$, $c_{y \to x} = 0.2$ and $c_{x \to y} = 0.5$. We run the system from random initial states and discard $100T$ samples before collecting T observations for our measures, to ensure the removal of transient effects. The embedding dimension $d = 1$ and lag $\tau = 1$ are selected to match those used in [36].

We examine $TE_{SPO}(X \to Y, u)$ for delays $u = 1$ and 2; clearly the correct delay should be measured as $u = \delta_{x \to y} = 2$. All calculations for test cases (Ia) and (Ib) were made using the open-source "Java Information Dynamics Toolkit" [40]; the results here can be reproduced using the demos/octave/DetectingInteractionLags demo of this toolkit. TE was measured using a Kraskov-Grassberger-Stögbauer estimator here, to contrast the results with those obtained for TE from symbolic mapping (capturing ordinal relationships) in [36] (which incorrectly inferred $u = 1$ as the interaction delay).

Autoregressive (AR) processes. Coupled autoregressive processes X, Y were generated according to the equations:

$$X(t+1) = \sum_{k=0}^{m} \alpha_k X(t-k) + \sigma \eta_X(t), \tag{29}$$

$$Y(t+1) = \sum_{k=0}^{m} \beta_k Y(t-k) + \sigma \eta_Y(t) \tag{30}$$

$$+ \frac{\gamma_{XY}}{|\{\delta_{XY}\}|} \sum_{\delta \in \{\delta_{XY}\}} X^2(t+1-\delta). \tag{31}$$

where m $= 10$ is the order of the autoregressive processes, $\sigma = 0.1$ is the dynamic noise amplitude of uncorrelated, unit-variance, zero-mean Gaussian noise terms $\eta_X(t)$ and $\eta_Y(t)$, $|\{\delta_{XY}\}|$ denotes the number of elements in the set of delays $\{\delta_{XY}\}$, and specific values for the delays (δ_{XY}) and coupling strengths (γ_{XY}) are listed in table 2 for test cases II-IV; the values for α_k and $beta_k$ where constructed from roots of the characteristic polynomial of the the AR process, that were chosen at random on the unit circle to guarantee a stationary AR process.

Chaotic dynamical systems. As a more complex case we investigated two Lorenz systems with non-linear (quadratic) coupling and potential self-feedback according to:

$$\dot{U}_i(t) = \sigma(V_i(t) - U_i(t))\, \dot{V}_i(t) =$$

$$U_i(t)(\rho_i - W_i(t)) - V_i(t) + \sum_{i,j=X,Y} \gamma_{ij} V_j^2(t - \delta_{ij}), \dot{W}_i(t) = \quad (32)$$

$$U_i(t)V_i(t) - \beta W_i(t),$$

with $i,j = X,Y$ ($i,j = X,Y,Z$ for teat case IX); parameters as indicated in table 2 for test cases IV-IX; σ, ρ and β, are the *Prandtl number*, the *Rayleigh number*, and a geometrical scale; γ_{ij} represent the coupling strengths from system i to j, with $i=j$ indicating (delayed) self-feedback. Note that always $\gamma_{YY} = 0$ (no self-feedback $Y \to Y$). The δ_{ij} are the respective delays of the coupling or of the self-feedback. Numerical solutions to these differential equations were computed using the *dde*23 solver in MATLAB and results were resampled such that the delays amounted to the values given in table 2. For analysis purposes we analyzed the V-coordiantes of the systems.

Noisy Lorenz systems. While the proof for our approach holds strictly only for noise-free systems, in practice the proposed procedure works well for the noise profile encountered in many technical or life-sciences applications. To demonstrate this, we simulated coupled Lorenz systems as in equation 32 and afterwards added independent,Gaussian, white observation noise of varying amplitude according to:

$$\tilde{V}_i(t) = X(t) + \xi_i \eta_{V_i}(t) \quad i = X,Y, \quad (33)$$

where V_i was simulated as above, $\eta_{V_i}(t)$ was unit variance Gaussian white noise and ξ_i chosen such that 1%, 2% and 9% of the final signal variance were contributed by noise.

Ring of Lorenz systems. We also coupled three Lorenz systems into a uni-directional ring using equations 32 above, however, this time setting $\delta_{XY} = 20$, $\delta_{YZ} = 30$, $\delta_{ZX} = 40$, and $\gamma_{XY} = \gamma_{YZ} = \gamma_{ZX} = 0.15$.

Electrophysiological data. In the last test case we used data which were recorded from the turtle (Pseudemys scripta elegans) to determine interaction delays between brain areas. This experiment was described previously in [42].

Preparation. Experiments were approved by the German local authorities (Regierungspraesidium, Hessen, Darmstadt). One turtle (Pseudemys scripta elegans) was anesthetized with 15 mg Ketamine, and 2 mg Medetomidinhydrochloride and decapitated. The entire brain with the eyes attached was removed as described in [63]. The brain was placed in a petri dish and superfused with oxygenated ringer. The ringer consisted of (in mM) 96.5 NaCl, 2.6 KCl, 2.0 MgCl2, 31.5 NaHCO3, 20 D-glucose, 4 CaCl2 at pH 7.4 and was administered at room temperature (22°C).

Electrophysiological recordings. The electroretinogram was recorded with a chlorided silver wire in a Vaseline well that was built around the right eye. The contralateral tectal signal was recorded in a superficial layer at the center of the left tectum with a quartz/platinum-tungsten electrode (Thomas Recordings, Giessen, Germany) with impedance 1 MΩ at 1 kHz. Data were amplified and filtered (1 Hz to 6 kHz) before being digitized at 32 kHz. For the analysis, the continuous data were low-pass filtered with 240 Hz, down-sampled to 500 Hz and cut into 60 trials with 50 s each.

Visual stimulation. A sequence of red LED light pulses with random duration (uniform distribution between 1 ms and 2 s) and random inter pulse interval (uniform distribution between 1 ms and 5 s) was triggered via the parallel port using MATLAB and the Psychophysics Toolbox extension [64,65]. A light guide projected the full field flashes onto the retina.

Author Contributions

Designed the software used in analysis: ML MW RV NP. Proved the theorem: MW HS. Conceived and designed the experiments: MW VP RV JL. Performed the experiments: MW VP RV NP ML. Analyzed the data: MW VP RV NP FS ML. Wrote the paper: MW RV VP NP.

References

1. Felts PA, Baker TA, Smith KJ (1997) Conduction in segmentally demyelinated mammalian central axons. J Neurosci 17: 7267–7277.
2. Whitford TJ, Ford JM, Mathalon DH, Kubicki M, Shenton ME (2012) Schizophrenia, myelination, and delayed corollary discharges: a hypothesis. Schizophr Bull 38: 486–494.
3. Sun L, Grützner C, Bölte S, Wibral M, Tozman T, et al. (2012) Impaired gamma-band activity during perceptual organization in adults with autism spectrum disorders: evidence for dysfunctional network activity in frontal-posterior cortices. J Neurosci 32: 9563–9573.
4. Pearl J (2000) Causality: models, reasoning, and inference. Cambridge University Press.
5. Ay N, Polani D (2008) Information ows in causal networks. Adv Complex Syst 11: 17.
6. Lizier JT, Prokopenko M (2010) Differentiating information transfer and causal effect. Eur Phys J B 73: 605–615.
7. Lizier JT, Prokopenko M, Zomaya AY (2008) Local information transfer as a spatiotemporal filter for complex systems. Phys Rev E 77: 026110.
8. Schreiber T (2000) Measuring information transfer. Phys Rev Lett 85: 461–464.
9. Paluš M (2001) Synchronization as adjustment of information rates: detection from bivariate time series. Phys Rev E 63: 046211.
10. Pecora L, Caroll T, Heagy JF (1996) Nonlinear dynamics and Time Series: Building a Bridge between natural and statistical sciences, Fields Institute Communications, American Mathematical Society, chapter Statistics for continuity and differentiability: An application to attractor reconstruction from time series. 49–62.
11. Terry JR, Breakspear M (2003) An improved algorithm for the detection of dynamical interdependence in bivariate time-series. Biol Cybern 88: 129–136.
12. Vicente R, Wibral M, Lindner M, Pipa G (2011) Transfer entropy – a model-free measure of effective connectivity for the neurosciences. J Comput Neurosci 30: 45–67.
13. Wibral M, Rahm B, Rieder M, Lindner M, Vicente R, et al. (2011) Transfer entropy in magnetoencephalographic data: Quantifying information ow in cortical and cerebellar networks. Prog Biophys Mol Biol 105: 80–97.
14. Vakorin VA, Kovacevic N, McIntosh AR (2010) Exploring transient transfer entropy based on a group-wise ica decomposition of eeg data. Neuroimage 49: 1593–1600.
15. Vakorin VA, Krakovska OA, McIntosh AR (2009) Confounding effects of indirect connections on causality estimation. J Neurosci Methods 184: 152–160.
16. Chávez M, Martinerie J, Le Van Quyen M (2003) Statistical assessment of nonlinear causality: application to epileptic EEG signals. J Neurosci Methods 124: 113–128.
17. Freiwald WA, Valdes P, Bosch J, Biscay R, Jimenez JC, et al. (1999) Testing non-linearity and directedness of interactions between neural groups in the macaque inferotemporal cortex. J Neurosci Methods 94: 105–119.
18. Amblard PO, Michel OJ (2011) On directed information theory and Granger causality graphs. J Comput Neurosci 30: 7–16.
19. Barnett L, Barrett AB, Seth AK (2009) Granger causality and transfer entropy are equivalent for Gaussian variables. Phys Rev Lett 103: 238701.
20. Besserve M, Scholkopf B, Logothetis NK, Panzeri S (2010) Causal relationships between frequency bands of extracellular signals in visual cortex revealed by an information theoretic analysis. J Comput Neurosci 29: 547–566.
21. Bühlmann A, Deco G (2010) Optimal information transfer in the cortex through synchronization. PLoS Comput Biol 6: e1000934.
22. Garofalo M, Nieus T, Massobrio P, Martinoia S (2009) Evaluation of the performance of information theory-based methods and cross-correlation to estimate the functional connectivity in cortical networks. PLoS One 4: e6482.
23. Gourevitch B, Eggermont JJ (2007) Evaluating information transfer between auditory cortical neurons. J Neurophysiol 97: 2533–2543.
24. Lizier JT, Heinzle J, Horstmann A, Haynes JD, Prokopenko M (2011) Multivariate informationtheoretic measures reveal directed information structure and task relevant changes in fmri connectivity. J Comput Neurosci 30: 85–107.
25. Lüdtke N, Logothetis NK, Panzeri S (2010) Testing methodologies for the nonlinear analysis of causal relationships in neurovascular coupling. Magn Reson Imaging 28: 1113–1119.

26. Neymotin SA, Jacobs KM, Fenton AA, Lytton WW (2011) Synaptic information transfer in computer models of neocortical columns. J Comput Neurosci 30: 69–84.

27. Sabesan S, Good LB, Tsakalis KS, Spanias A, Treiman DM, et al. (2009) Information ow and application to epileptogenic focus localization from intracranial EEG. IEEE Trans Neural Syst Rehabil Eng 17: 244–53.

28. Staniek M, Lehnertz K (2009) Symbolic transfer entropy: inferring directionality in biosignals. Biomed Tech (Berl) 54: 323–8.

29. Vakorin VA, Mii B, Krakovska O, McIntosh AR (2011) Empirical and theoretical aspects of generation and transfer of information in a neuromagnetic source network. Front Syst Neurosci 5: 96.

30. Leistritz L, Hesse W, Arnold M, Witte H (2006) Development of interaction measures based on adaptive non-linear time series analysis of biomedical signals. Biomed Tech (Berl) 51: 64–69.

31. Hadjipapas A, Hillebrand A, Holliday IE, Singh KD, Barnes GR (2005) Assessing interactions of linear and nonlinear neuronal sources using meg beamformers: a proof of concept. Clin Neurophysiol 116: 1300–1313.

32. Faes L, Nollo G (2006) Bivariate nonlinear prediction to quantify the strength of complex dynamical interactions in short-term cardiovascular variability. Med Biol Eng Comput 44: 383–392.

33. Faes L, Nollo G, Porta A (2011) Non-uniform multivariate embedding to assess the information transfer in cardiovascular and cardiorespiratory variability series. Comput Biol Med 42: 290–297.

34. Faes L, Nollo G, Porta A (2011) Information-based detection of nonlinear granger causality in multivariate processes via a nonuniform embedding technique. Phys Rev E Stat Nonlin Soft Matter Phys 83: 051112.

35. Lizier JT, Prokopenko M, Zomaya AY (2010) Information modification and particle collisions in distributed computation. Chaos 20: 037109.

36. Pompe B, Runge J (2011) Momentary information transfer as a coupling measure of time series. Phys Rev E Stat Nonlin Soft Matter Phys 83: 051122.

37. Lizier JT, Prokopenko M, Zomaya AY (2012) Local measures of information storage in complex distributed computation. Information Sciences 208: 39–54.

38. Kantz H, Schreiber T (2003) Nonlinear Time Series Analysis. Cambridge University Press, 2 edition.

39. Ragwitz M, Kantz H (2002) Markov models from data by simple nonlinear time series predictors in delay embedding spaces. Phys Rev E Stat Nonlin Soft Matter Phys 65: 056201.

40. Lizier JT (2012) JIDT: An information-theoretic toolkit for studying the dynamics of complex systems. URL https://code.google.com/p/information-dynamics-toolkit/. Accessed 2013 Jan. 24.

41. Kraskov A, Stoegbauer H, Grassberger P (2004) Estimating mutual information. Phys Rev E Stat Nonlin Soft Matter Phys 69: 066138.

42. Lindner M, Vicente R, Priesemann V, Wibral M (2011) Trentool: A matlab open source toolbox to analyse information ow in time series data with transfer entropy. BMC Neurosci 12(119): 1–22.

43. Takens F (1981) Dynamical Systems and Turbulence, Warwick 1980, Springer, volume 898 of *Lecture Notes in Mathematics*, chapter Detecting Strange Attractors in Turbulence. 366–381.

44. Voss HU (2000) Anticipating chaotic synchronization. Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics 61: 5115–5119.

45. Hahs DW, Pethel SD (2011) Distinguishing anticipation from causality: anticipatory bias in the estimation of information ow. Phys Rev Lett 107: 128701.

46. Soriano MC, der Sande GV, Fischer I, Mirasso CR (2012) Synchronization in simple network motifs with negligible correlation and mutual information measures. Phys Rev Lett 108: 134101.

47. Chicharro D, Ledberg A (2012) When two become one: the limits of causality analysis of brain dynamics. PLoS One 7: e32466.

48. Nichols JM, Seaver M, Trickey ST, Todd MD, Olson C, et al. (2005) Detecting nonlinearity in structural systems using the transfer entropy. Phys Rev E Stat Nonlin Soft Matter Phys 72: 046217.

49. Overbey L, Todd MD (2009) Dynamic system change detection using a modification of the transfer entropy. Journal of Sound and Vibration 322: 438–453.

50. Ito S, Hansen ME, Heiland R, Lumsdaine A, Litke AM, et al. (2011) Extending transfer entropy improves identification of effective connectivity in a spiking cortical network model. PLoS One 6: e27431.

51. Csorgo M, Horvath L (1998) Limit Theorems in Change-Point Analysis. Wiley.

52. Basseville M, Nikiforov I (1993) Detection of Abrupt changes: Theory and Application. Prentice Hall.

53. Gomez-Herrero G, Wu W, Rutanen K, Soriano MC, Pipa G, et al. (2010) Assessing coupling dynamics from an ensemble of time series. arXiv 1008.0539.

54. Turi G, Gotthardt S, Singer W, Vuong TA, Munk M, et al. (2012) Quantifying additive evoked contributions to the event-related potential. Neuroimage 59: 2607–2624.

55. Lizier JT, Rubinov M (2012) Multivariate construction of effective computational networks from observational data. Technical Report 25/2012, Max Planck Institute for Mathematics in the Sciences.

56. Van der Loos H, Glaser EM (1972) Autapses in neocortex cerebri: synapses between a pyramidal cell's axon and its own dendrites. Brain Res 48: 355–360.

57. Marimont RB, Shapiro MB (1979) Nearest neighbour searches and the curse of dimensionality. Journal of the Institute of Mathematics and its Applications 24: 59–70.

58. Wibral M, Turi G, Linden DEJ, Kaiser J, Bledowski C (2008) Decomposition of working memoryrelated scalp erps: crossvalidation of fmri-constrained source analysis and ica. Int J Psychophysiol 67: 200–211.

59. Nolte G, Ziehe A, Nikulin VV, Schlogl A, Kramer N, et al. (2008) Robustly estimating the ow direction of information in complex physical systems. Phys Rev Lett 100: 234101.

60. Kozachenko L, Leonenko N (1987) Sample estimate of entropy of a random vector. Probl Inform Transm 23: 95–100.

61. Victor J (2005) Binless strategies for estimation of information from neural data. Phys Rev E 72: 051903.

62. Genovese CR, Lazar NA, Nichols T (2002) Thresholding of statistical maps in functional neuroimaging using the false discovery rate. Neuroimage 15: 870–878.

63. Rosenberg AF, Ariel M (1990) Visual-response properties of neurons in turtle basal optic nucleus in vitro. J Neurophysiol 63: 1033–45.

64. Brainard DH (1997) The Psychophysics Toolbox. Spat Vis 10: 433–6.

65. Pelli DG (1997) The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spat Vis 10: 437–42.