

Causal influence in linear Langevin networks without feedback

Andrea Auconi,^{1,2} Andrea Giansanti,^{2,3} and Edda Klipp^{1,*}

¹*Theoretische Biophysik, Humboldt-Universität zu Berlin, Invalidenstraße 42, D-10115 Berlin, Germany*

²*Dipartimento di Fisica, Sapienza Università di Roma, Rome, Italy*

³*INFN, Sezione di Roma 1, Rome, Italy*

(Received 25 October 2016; published 18 April 2017)

The intuition of causation is so fundamental that almost every research study in life sciences refers to this concept. However, a widely accepted formal definition of causal influence between observables is still missing. In the framework of linear Langevin networks without feedback (linear response models) we propose a measure of causal influence based on a new decomposition of information flows over time. We discuss its main properties and we compare it with other information measures like the transfer entropy. We are currently unable to extend the definition of causal influence to systems with a general feedback structure and nonlinearities.

DOI: [10.1103/PhysRevE.95.042315](https://doi.org/10.1103/PhysRevE.95.042315)

I. INTRODUCTION

In the classical description of physical systems, observable objects are interacting in a symmetric way and “causation” means that a particular configuration at time t is followed by (it *causes*) a new configuration at time $t + \tau$ which is univocally determined by the laws of nature [1].

In the case of complex and living systems [2–4], we are rarely able to provide a full mechanistic description because of the overwhelmingly broad range of time scales involved. Moreover, it is often hard even to observe (and define the boundaries of) the objects taking part in these processes and accurately measure their properties. This lack of knowledge leads us to conceive and represent the world *in the practical sense* as if there were some intrinsically free observables, the signals or stimuli, which influence the behavior of other observables, the responses, through asymmetric causal interactions. Signals are *free*, meaning that their dynamics is not influenced by the responses. As an example, the fluctuations of nutrients in the environment are usually modeled as independent random processes, and the cell responses as the consequent activation of biochemical signaling pathways.

We see that the intuition of causal influence between observables originates from a probabilistic description of nature. When we say that a signal provokes a response we mean that the knowledge of the status of the signal $x(t)$ at time t provides some information on the evolution $y(t + \tau)$ at time $t + \tau$ of the response which cannot be extracted solely by the knowledge of the response $y(t)$ at time t itself, i.e., which is not already present in a redundant way in $y(t)$.

The causal influence is a measure of nonredundant information flow over time. It quantifies how, on average, the state of the signal affects the evolution of the state of the response after a time period τ .

In order to make this concept more precise, we first need to find an appropriate picture of signals and responses. It is convenient to use ergodic stochastic processes so that the probability distributions can be calculated as temporal averages on a single trajectory and the initial condition is irrelevant. We considered the independence of the dynamics

of the signals from their responses by excluding the possibility of feedback. Whether a formalization of causal influence is possible outside of the signal-response description remains an open question. We choose to work on linear Langevin networks without feedback (linear response models) because of their analytical tractability and intuitiveness.

Causation is naturally linked with an asymmetry in time, because the effects are defined to be successive to their causes. The arrow of time is not given *a priori* but is empirically understood, therefore the time asymmetry which is proper of causality (the irreversibility of trajectories [5]) cannot be derived from first principles but should be incorporated into the definition of causal influence itself.

We define the causal influence in the partial information decomposition (PID) framework [6] as the unique information that the signal has on the evolution of the response, i.e., the time-lagged Shannon’s mutual information $I(x(t), y(t + \tau))$ minus the redundancy $R(x(t), y(t); y(t + \tau))$:

$$C_{x \rightarrow y}(\tau) \equiv I(x(t), y(t + \tau)) - R(\tau), \quad (1)$$

where $C_{x \rightarrow y}(\tau)$, $R(\tau)$, and $I(x(t), y(t + \tau))$ are functions of the time period τ and not of the time instant t because we consider only ergodic stationary processes.

While Shannon’s mutual information is known [7], the redundancy has to be defined. Some previously defined redundancy measures have been demonstrated all to have the same trivial and unintuitive form in Gaussian systems [8]: they take as redundancy the minimum I_{\min} of the information on the output $[y(t + \tau)]$ given by the sources $[x(t)$ and $y(t)]$ regardless of the mutual information that the sources share, $I_{\min} = \min[I(x(t), y(t + \tau)), I(y(t), y(t + \tau))]$. We define instead the redundancy as a composition of the mutual information between the two sources $x(t)$ and $y(t)$, $I_{xy} \equiv I(x(t), y(t))$, and the total information that they give together on the target $y(t + \tau)$, $I_{\text{tot}} \equiv I[y(t + \tau), (x(t), y(t))]$:

$$R(\tau) \equiv \frac{1}{2} \ln \left(\frac{e^{2(I_{xy} + I_{\text{tot}})}}{e^{2I_{xy}} + e^{2I_{\text{tot}}} - 1} \right). \quad (2)$$

We study and motivate the use of this definition considering the behavior of the resulting measure of causal influence $C_{x \rightarrow y}(\tau)$ in linear response models. However, it results that $C_{x \rightarrow y}(\tau)$ gives an appropriate description of the dynamics of

*edda.klipp@rz.hu-berlin.de

influences only in linear response models. In the case of more general systems of equations with feedback and nonlinearities, $C_{x \rightarrow y}(\tau)$ can assume negative values for which we do not have a clear physical interpretation.

In Sec. II A we review the information processing properties of the bidimensional linear response model. It consists of a fluctuating signal x produced by a unidimensional Ornstein-Uhlenbeck process [9, 10] that linearly influences the dynamics of a response variable y . There we study the properties of the causal influence $C_{x \rightarrow y}(\tau)$ and we compare it to the time-lagged mutual information $I(x(t), y(t + \tau))$, to the transfer entropy $T_{x \rightarrow y} = I(x(t), y(t + \tau) | y(t))$, and to the previously defined unique information measures. Then in Sec. II B we extend the definition to the multidimensional case of directed acyclic graphs and discuss some simple examples. Finally, in the discussion section we consider in more detail the difficulties in the search for a general definition of causal influence for complex systems.

II. RESULTS

A. Basic linear response model

We define the basic linear response model (BLRM) with the two stochastic differential equations

$$\begin{aligned} \frac{dx}{dt} &= -\frac{x}{t_{\text{rel}}} + \sqrt{D} \Gamma(t), \\ \frac{dy}{dt} &= \alpha x - \beta y. \end{aligned} \quad (3)$$

The $x(t)$ is an Ornstein-Uhlenbeck (OU) process [9, 10]. It depends on the realization of the uncorrelated Gaussian noise Γ , which is defined as the $dt \rightarrow 0$ limit of the normal random variable $\mathcal{N}(0, \frac{1}{dt})$. The OU process describes fluctuations around zero mean with an autocorrelation function that decays exponentially, $\langle x(t)x(t + \tau) \rangle = \sigma_x^2 \exp(-\tau/t_{\text{rel}})$, where t_{rel} is the relaxation time, and whose amplitude is proportional to the square root of the diffusion coefficient D , $\sigma_x^2 = D \frac{t_{\text{rel}}}{2}$. One easily shows that the probability density of $x(t + t')$ conditioned on the knowledge of $x(t)$ for generic (positive or negative) time shifts t' is given by the Gaussian distribution:

$$P(x(t + t') | x(t)) = \mathcal{G}[x(t) e^{-\frac{|t'|}{t_{\text{rel}}}}, \sigma_x^2 (1 - e^{-\frac{2|t'|}{t_{\text{rel}}}})]. \quad (4)$$

The second equation in system (2) describes a linear response to the OU process ($\alpha \neq 0$ and $\beta > 0$ are constants), whose formal solution is

$$y(t + \tau) = y(t) e^{-\beta \tau} + \alpha \int_0^\tau dt' x(t + t') e^{-\beta(\tau - t')}. \quad (5)$$

The BLRM is the simplest continuous-time dynamical system with evident causal influence: the dynamics of the variable y is driven by the position of the variable x , which fluctuates around a mean value. The causal network is simply $x \rightarrow y$ and from now on we call x the signal and y the response.

We see (Fig. 1) that the response is not following the signal in its “smallest” fluctuations but is rather integrating it on a time scale defined by the parameters β and t_{rel} as we show. Since the BLRM is a stationary process, the mean of the derivative of the variables products is vanishing $\langle \frac{d(y^2)}{dt} \rangle = \langle \frac{d(xy)}{dt} \rangle = 0$, and

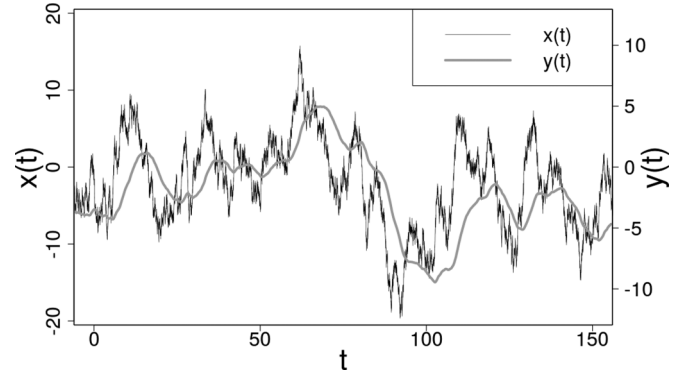


FIG. 1. Stochastic dynamics of the basic linear response model. The parameters are $\alpha = 0.1$, $\beta = 0.2$, $t_{\text{rel}} = 10$, and $D = 10$. All graphs are produced using R [11].

the amplitude of the response fluctuations (driven by the signal) σ_y proves to be proportional to α , $\sigma_y^2 = \frac{\alpha^2 t_{\text{rel}}}{\beta(\beta t_{\text{rel}} + 1)} \sigma_x^2$.

The expectation value of the response y at time $t \pm \tau$ ($\tau > 0$), conditioned on the knowledge of the signal x at time t (Fig. 2), is found using Eq. (4):

$$\langle y(t - \tau) | x(t) \rangle = x(t) \frac{\alpha t_{\text{rel}}}{\beta t_{\text{rel}} + 1} e^{-\frac{\tau}{t_{\text{rel}}}}, \quad (6)$$

$$\langle y(t + \tau) | x(t) \rangle = x(t) \frac{\alpha t_{\text{rel}}}{\beta t_{\text{rel}} - 1} \left(e^{-\frac{\tau}{t_{\text{rel}}}} - \frac{2e^{-\beta \tau}}{\beta t_{\text{rel}} + 1} \right). \quad (7)$$

The Gaussianity of the OU process implies the Gaussianity of the BLRM [12], so that the mutual information [7] that $x(t)$ and $y(t + t')$ share is just a function of their correlation $C(x(t), y(t + t'))$:

$$\begin{aligned} I(x(t), y(t + t')) &= \ln \left(\frac{\sigma_y}{\sigma_{y(t+t')|x(t)}} \right) \\ &= -\frac{1}{2} \ln[1 - C^2(x(t), y(t + t'))]. \end{aligned} \quad (8)$$

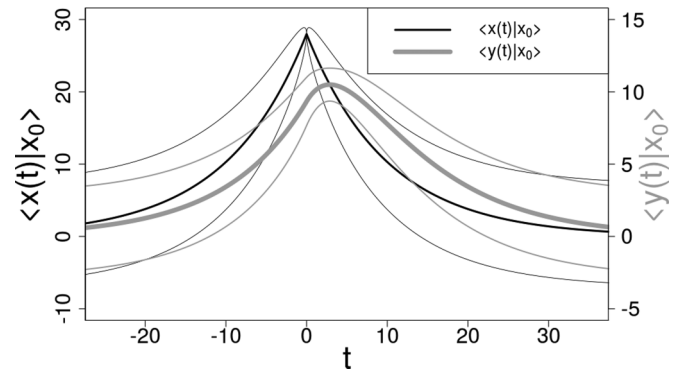


FIG. 2. Conditional probability distributions over time. Given a particular condition (input) at time $t = 0$, $x(0) \equiv x_0 = 28$, we plot the conditioned expectation values $\langle y(t) | x(0) \rangle$ and $\langle x(t) | x(0) \rangle$ with the relative standard deviations $\pm \sigma_{y(t)|x(0)}$ and $\pm \sigma_{x(t)|x(0)}$ (thinner lines) as a function of the time shift t . The parameters are $\alpha = 0.1$, $\beta = 0.2$, $t_{\text{rel}} = 10$, $D = 10$.

The interpretation of the mutual information as the reduction in uncertainty of the variable $y(t + t')$ upon knowledge of the variable $x(t)$ is clearly seen in Eq. (8). We note that in the BLRM, the conditional variance $\sigma_{y(t+t')|x(t)}$ is independent of the particular condition $x(t)$.

Thanks to the fluctuation-dissipation theorem [13] we can also interpret the mutual information as a function of the signal-to-noise ratio $R_{S/N}$, which is the average square deviation from the mean of the estimate of $y(t + t')$ provoked by the knowledge of $x(t)$ divided by the uncertainty that remains $\sigma_{y(t+t')|x(t)}$:

$$I(x(t), y(t + t')) = \frac{1}{2} \ln(1 + R_{S/N}), \quad (9)$$

$$R_{S/N} = \frac{\left(\frac{\partial \langle y(t+t')|x(t) \rangle}{\partial x(t)}\right)^2 \sigma_x^2}{\sigma_{y(t+t')|x(t)}^2}. \quad (10)$$

The time shift which gives the optimal information transmission is given by

$$\tau_{\text{opt}} = \frac{t_{\text{rel}}}{\beta t_{\text{rel}} - 1} \ln\left(\frac{2\beta t_{\text{rel}}}{\beta t_{\text{rel}} + 1}\right), \quad (11)$$

which is always positive as we expect since is the response y that is following the fluctuations of the signal x producing a natural delay in the correlation [14].

The mutual information corresponding to the optimal time shift is

$$I^{\text{opt}} = I(x(t), y(t + \tau_{\text{opt}})) = -\frac{1}{2} \ln\left(1 - 2\left(\frac{\beta t_{\text{rel}} + 1}{2\beta t_{\text{rel}}}\right)^{\frac{\beta t_{\text{rel}} + 1}{\beta t_{\text{rel}} - 1}}\right), \quad (12)$$

which depends just on the product βt_{rel} , which we understand as the ratio of the two time scales of the model: the relaxation time of the signal fluctuations t_{rel} and the response time of the system for a deterministic input $\frac{1}{\beta}$. The limits of high and low information are, respectively, $I^{\text{opt}}(\beta t_{\text{rel}} \rightarrow \infty) = \frac{\ln(\beta t_{\text{rel}})}{2}$ and $I^{\text{opt}}(\beta t_{\text{rel}} \rightarrow 0) = 2\beta t_{\text{rel}}$.

The mutual information is a fundamental characterization of the system, and in general one can use this measure to infer causal relations in network reconstruction [15]. Still, it is not a measure of causal influence. The time-lagged mutual information is greater than 0 in both cases $I(x(t), y(t + \tau))$ and $I(y(t), x(t + \tau))$, meaning that the response at time t gives some information on the signal at successive times $t + \tau$, as we expect since the signal has autocorrelation. Nevertheless, the response does not influence the dynamics of the signal, therefore a measure of causal influence should be 0 in this case.

We could solve this problem considering instead the conditional mutual information, the so-called transfer entropy [16] (equivalent to the Granger causality [17] in linear systems [18]), that is, the additional amount of information that one gets on $y(t + \tau)$ upon knowledge of $x(t)$ when $y(t)$ is already known. For the BLRM it is given by

$$\begin{aligned} T_{x \rightarrow y}(\tau) &= I(x(t), y(t + \tau) | y(t)) = \ln\left(\frac{\sigma_{y(t+\tau)|y(t)}}{\sigma_{y(t+\tau)|x(t), y(t)}}\right) \\ &= \frac{1}{2} \ln\left(1 + \frac{\beta t_{\text{rel}}(e^{-\frac{\tau}{t_{\text{rel}}}} - e^{-\beta\tau})^2}{(1 - \beta t_{\text{rel}})^2 - e^{-2\beta\tau}(1 + \beta t_{\text{rel}}) + e^{-(\beta + \frac{1}{t_{\text{rel}}})\tau} 4\beta t_{\text{rel}} - e^{-\frac{2\tau}{t_{\text{rel}}}} \beta t_{\text{rel}}(1 + \beta t_{\text{rel}})}\right), \end{aligned} \quad (13)$$

$$T_{y \rightarrow x}(\tau) = I(y(t), x(t + \tau) | x(t)) = \ln\left(\frac{\sigma_{x(t+\tau)|x(t)}}{\sigma_{x(t+\tau)|y(t), x(t)}}\right) = 0, \quad (14)$$

where $\sigma_{y(t+\tau)|x(t), y(t)}$ is the standard deviation of $y(t + \tau)$ conditioned on the knowledge of $x(t)$ and $y(t)$.

Here we have $T_{y \rightarrow x} = 0$ according to the fact that the dynamics of the signal is independent of the response. $T_{x \rightarrow y}(\tau)$ is always positive instead and diverges for $\tau \rightarrow 0$ because the knowledge of both $x(t)$ and $y(t)$ synergistically provides information on $y(t + \tau)$ since for small τ the $y_{t+\tau}$ map is (almost always) quasideterministic in x_t and y_t : $y(t + \tau) - y(t) = \tau(\alpha x(t) - \beta y(t)) + o(\tau^2)$. This feature we would call determinism, while the causal influence is rather the visible (macroscopic) effect of the causation of the evolution of the response by the signal, which is obtained gradually over time after the ‘‘cause’’ $x(t)$. It is the information on the evolution of the response $y(t + \tau)$ that we get looking at just the signal $x(t)$, $I(x(t), y(t + \tau))$, minus that part of the information which is already present in a redundant way also in the response $y(t)$.

Following Barrett [8], we define a partial information decomposition [6] of the total information that $x(t)$ and $y(t)$

give on the evolution of the response $y(t + \tau)$,

$$I[y(t + \tau), (x(t), y(t))] = R + U_x + U_y + S, \quad (15)$$

where R is the redundancy, U_x and U_y are the unique information contributions, respectively, of $x(t)$ and $y(t)$ alone, and the synergy S is defined as the information that one gets in addition when considering simultaneously both $x(t)$ and $y(t)$. People are trying to define the information decomposition such that the unique contribution from the signal $U_x(\tau)$ can be interpreted as a measure of the actual information flowing from the signal to the response over time [19]; this is what we call causal influence. Since the input-output mutual information is decomposed in $I(x(t), y(t + \tau)) = R + U_x$ and the transfer entropy in $T_{x \rightarrow y} = U_x + S$, in order to specify the decomposition it is sufficient to give a definition of the redundancy R .

In Gaussian systems [8] the previously defined PID are all equal: they take as redundancy the minimum value between

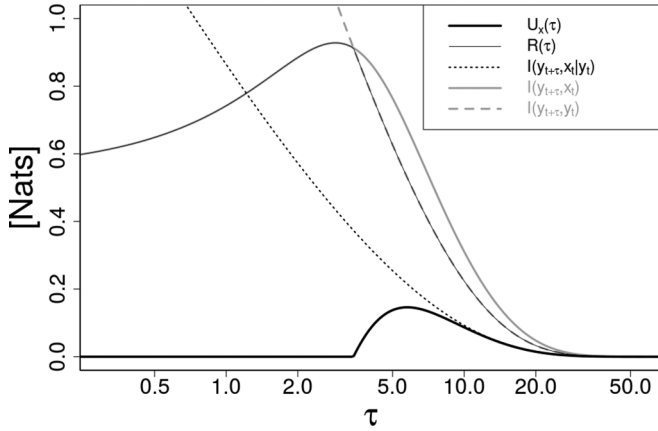


FIG. 3. Previously proposed PIDs where $R(\tau) = I_{\min}$. Information measures are expressed in natural units, Nats = $\frac{\text{bits}}{\ln 2}$. The τ axis is on the logarithmic scale. The parameters are $\beta = 0.2$, $t_{\text{rel}} = 10$.

$I(x(t), y(t + \tau))$ and $I(y(t), y(t + \tau))$, regardless of the information shared between the two sources $I(x(t), y(t))$, $I_{\min} = \min[I(x(t), y(t + \tau)), I(y(t), y(t + \tau))]$. With this definition (Fig. 3) the unique information U_x is 0 until $I(x(t), y(t + \tau))$ is smaller than $I(y(t), y(t + \tau))$, suggesting the existence of an unintuitive activation time for causality. Moreover, let us consider the time shift $\tau = \tau_e$ for which $I(x(t), y(t + \tau_e)) = I(y(t), y(t + \tau_e))$. According to the previously defined PID, where $R(\tau) = I_{\min}$, the mutual information should be totally redundant. However, redundancy does not mean that the two sources give the same information here since the estimate of $y(t + \tau)$ given $x(t)$ is in general different from the estimate of $y(t + \tau)$ given $y(t)$, while only the (reduction in) uncertainty of the estimates is the same; the similarity of these two estimates is given by $I(x(t), y(t))$ and the I_{\min} definition of redundancy is not explicitly dependent on this.

We define instead the redundancy as a composition of the information shared between the two sources $x(t)$ and $y(t)$, $I_{xy} \equiv I(x(t), y(t))$, and the total information that they share with the target $y(t + \tau)$, $I_{\text{tot}} \equiv I_{\text{tot}}(\tau) \equiv I[y(t + \tau), (x(t), y(t))]$ [Eq. (2), Sec. I]:

$$R(\tau) \equiv \frac{1}{2} \ln \left(\frac{e^{2(I_{xy} + I_{\text{tot}})}}{e^{2I_{xy}} + e^{2I_{\text{tot}}} - 1} \right).$$

This definition is inspired by the logic of a linear Markov chain; i.e., we use the formula of the information shared between the variables A and C in the *static* Gaussian linear network $A \rightarrow B \rightarrow C$, namely, $I_{AC} = I_{AB} + I_{BC} - \frac{1}{2} \ln(e^{2I_{AB}} + e^{2I_{BC}} - 1)$. In other words, we define the redundancy as the information that $x(t)$ has on $y(t + \tau)$ “passing through” $y(t)$. The redundant information that the two variables $x(t)$ and $y(t)$ both have on the third variable $y(t + \tau)$ is a fraction of the mutual information between them, I_{xy} , and that fraction is determined by the total information that they have together on $y(t + \tau)$, I_{tot} . The redundancy measure is symmetric in the two sources (causes) $x(t)$ and $y(t)$ and, also, symmetric in I_{xy} and I_{tot} . We call our PID the linear information decomposition (Fig. 4) and we suggest that the proposed definition of redundancy [Eq. (2)] could be a linear approximation of a more general definition to be found.

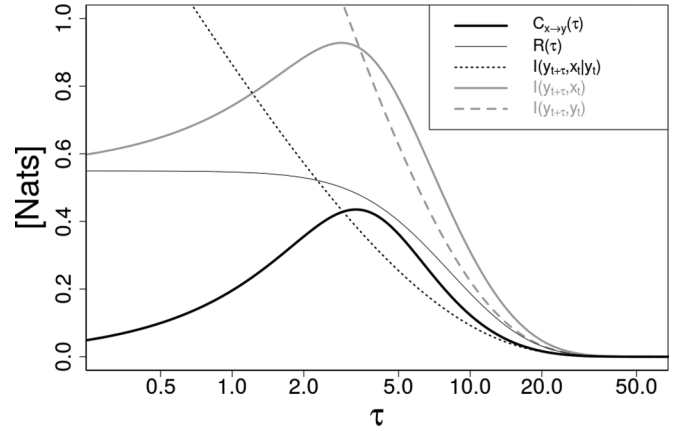


FIG. 4. Linear information decomposition $x \rightarrow y$. The thick black curve represents the unique information that $x(t)$ gives on $y(t + \tau)$, which is our measure of causal influence $C_{x \rightarrow y}(\tau)$. The parameters are $\beta = 0.2$, $t_{\text{rel}} = 10$.

The unique information U_x that results from Eq. (2), $U_x = I(x(t), y(t + \tau)) - R(\tau)$, we claim to be a measure of causal influence for the BLRM [Eq. (1), Sec. I]:

$$C_{x \rightarrow y}(\tau) \equiv I(x(t), y(t + \tau)) - R(\tau).$$

We define the causal influence $C_{x \rightarrow y}(\tau)$ only for positive $\tau \geq 0$ to include the empirical knowledge that the effects are always seen after the causes. $C_{x \rightarrow y}(\tau)$ is a measure of information and, therefore, is measured in natural units, Nats.

Given the realization of the OU process x for a sufficiently long time we can determine with any precision the position of the $y(t + \tau)$, and it is fair to say that this value is totally caused by the sequence of x at previous times. This fact is mirrored in the divergence of the transfer entropy for small time shifts $\tau \rightarrow 0$. Nevertheless, (in common language) we usually consider as causes the single observable facts [$x(t)$ and $y(t)$ in the BLRM] and as the effect a successive observable fact [$y(t + \tau)$], and we wish to quantify the relative strength of these causes in giving the effect. The causal influence $C_{x \rightarrow y}(\tau)$ that the signal has on the response over time is 0 for the time shift $\tau = 0$ and increases with τ (linearly for small τ), meaning that we get the effect of causality (that is the causal influence) gradually over time after the cause $x(t)$. For very long time intervals τ after the cause we can no longer see the effect of the distant past and the causal influence goes to 0. The time shift at which the causal influence peaks τ_{res} is the response time of the system in the probabilistic sense and is slightly different from the maximum correlation time τ_{opt} . In general $\tau_{\text{res}} > \tau_{\text{opt}}$. We note that, as it should be, we get zero causal influence of the response y on the signal x (Fig. 5). This is because the information $I(y(t), x(t + \tau))$ that the response has on the evolution of the signal is gained necessarily via the two steps $y(t) \rightarrow x(t)$ and $x(t) \rightarrow x(t + \tau)$ due to the asymmetry of the interaction and therefore is equal to the redundancy $R(x(t), y(t); x(t + \tau))$. We always get 0 also in the calculation of the self-causal influence $C_{x \rightarrow x} = 0$, meaning that causation is exerted only between different observables, while the autocorrelation of one variable just means memory of the previous states.

The concepts of redundancy and synergy were originally defined (outside of the PID framework) as a single quantity, the

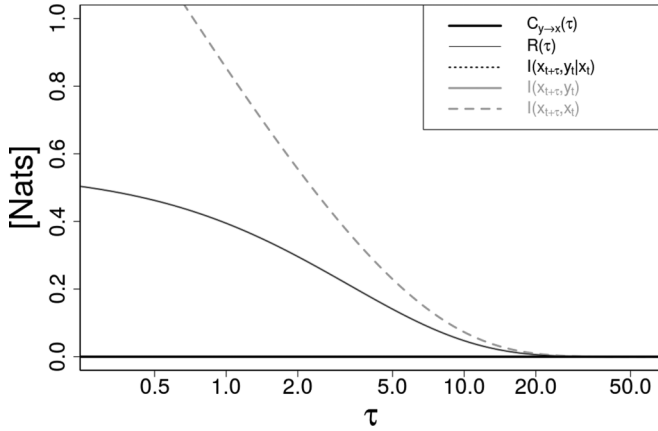


FIG. 5. Linear information decomposition $y \rightarrow x$. The redundant information is equal to the mutual information, meaning that there is no causal influence. The parameters are $\beta = 0.2$, $t_{\text{rel}} = 10$.

“net redundancy/synergy” coinformation measure $I_{R/S}$ [20]:

$$\begin{aligned} I_{R/S}(y(t + \tau), x(t), y(t)) \\ &= I_{\text{tot}} - I(y(t + \tau), x(t)) - I(y(t + \tau), y(t)) \\ &= I(y(t + \tau), x(t) | y(t)) - I(y(t + \tau), x(t)). \end{aligned} \quad (16)$$

Positive and negative values of CoI indicate, respectively, synergy and redundancy. The BLRM is synergistic for small time shifts τ and redundant for larger τ when the mutual information exceeds the transfer entropy. $I_{R/S}$ is symmetric in its three arguments $[x(t), y(t), y(t + \tau)]$ and the relation with the PID measures is simply $I_{R/S}(y(t + \tau); x(t), y(t)) = S - R$.

To understand the behavior of the causal influence in the BLRM as a function of the parameters we study the limits of high and low information (Figs. 6 and 7).

When $\beta t_{\text{rel}} \gg 1$ the mutual information is high and increases with $\ln(\beta t_{\text{rel}})$. The peak of the causal influence also increases but only up to a limit of around $\lim_{\beta t_{\text{rel}} \rightarrow \infty} \max_{\tau} C_{x \rightarrow y} \approx 0.55$ Nat. The position of the peak depends on β : at higher β the response is faster and the effect of causality is seen earlier. When β is fixed, increasing t_{rel} always gives an increase in the mutual information because the slowdown of the dynamics of the signal lets the response

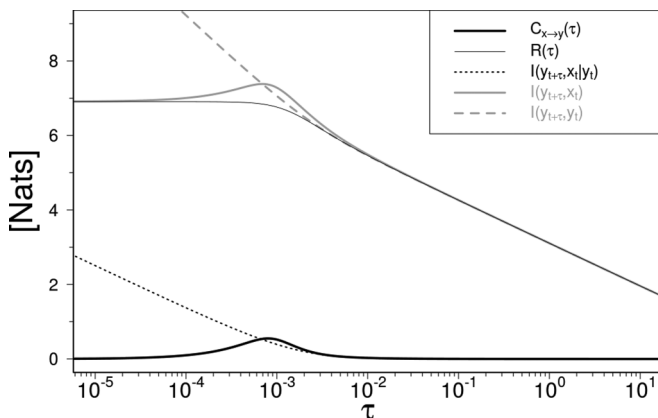


FIG. 6. Linear information decomposition $x \rightarrow y$. High-information scenario: $\beta = 1000$, $t_{\text{rel}} = 1000$.

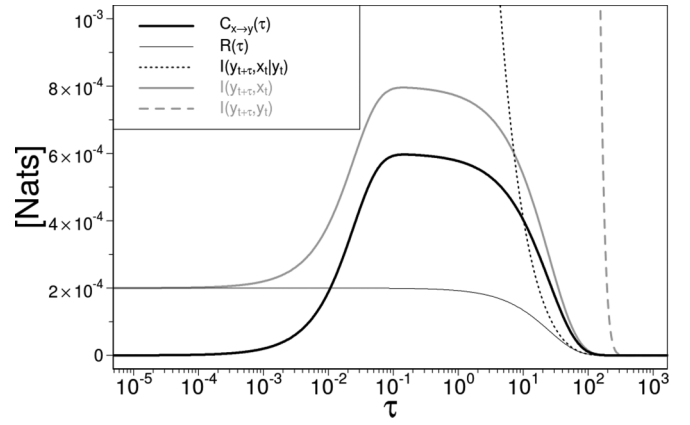


FIG. 7. Linear information decomposition $x \rightarrow y$. Low-information scenario: $\beta = 0.02$, $t_{\text{rel}} = 0.02$.

follow the microscopical structure of the signal with more precision, but at the same time the response is moved slowly (in units of its standard deviation) by the signal, these two effects asymptotically compensating and the peak of the causal influence staying around ≈ 0.55 Nat. This limit we call the causation capacity of the BLRM. In the case of low information $\beta t_{\text{rel}} \ll 1$ the peak of the causal influence is close to 75% of the peak of the mutual information because $\frac{I_{\text{opt}}}{I_{xy}} \rightarrow 4$ for $\beta t_{\text{rel}} \rightarrow 0$. The signal has a fast-decaying autocorrelation, the response is slowly integrating (keeping the memory of) it, and therefore most of the small amount of time-lagged mutual information on the response is causal influence.

Let us now consider a more traditional approach in data analysis. We write the vector autoregressive model for the evolution of the response as $y(t + \tau) = \gamma_{yy}(\tau)y(t) + \gamma_{xy}(\tau)x(t) + \xi(\tau)$, where the γ 's are the linear coefficients of the expansion and $\xi(\tau)$ is the error term. One could consider the coefficient $\gamma_{xy}(\tau) = \frac{\alpha t_{\text{rel}}}{\beta t_{\text{rel}} - 1} (e^{-\frac{\tau}{t_{\text{rel}}}} - e^{-\beta\tau})$ as a measure of the influence of the signal on the response, but then the error term $\xi(\tau)$ would have no explicit role. The intuition of causal influence is based on the concept of information flow and the fact that the uncertainty of the prediction $\langle \xi^2(\tau) \rangle = \sigma_{y(t+\tau)|x(t), y(t)}$ increases with τ should consequently decrease the causal influence. The discrepancy of the coefficient $\gamma_{xy}(\tau)$ with the information measures is clearly shown in Fig. 8.

One dissatisfying feature of our definition, however, is that the redundancy measure does not satisfy the local positivity axiom of the PID; i.e., the synergistic information $S = T_{x \rightarrow y} - C_{x \rightarrow y}$ is negative when the causal influence is greater than the transfer entropy, and this is always the case for long delays τ . This means that part of the “same” information that $x(t)$ and $y(t)$ give on $y(t + \tau)$ is considered as causal influence and not redundancy.

B. Multidimensional case: Networks without feedback

We can extend the causal influence measure for interactions within linear Langevin networks without feedback. Let us define the network of direct influences as the one that has directed links for all the combination of variables (nodes in the network) $(i \rightarrow j)$ for which the variable i appears in the equation for the dynamics of the variable j . The network of the

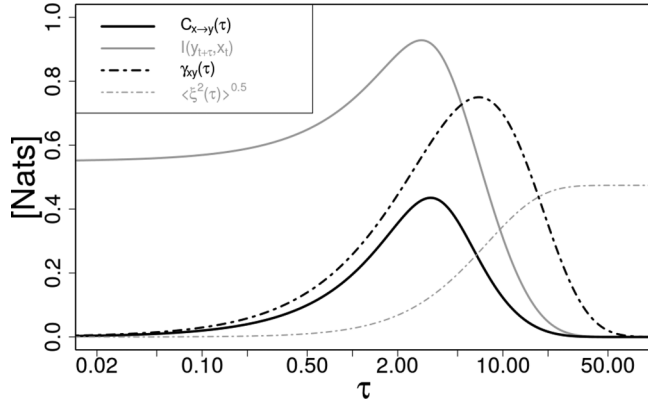


FIG. 8. The coefficient γ_{xy} of the vector autoregressive model compared with the information measures. γ_{xy} and $\sqrt{\langle \xi^2(\tau) \rangle}$ are adimensional. The parameters are $\beta = 0.2$, $t_{\text{rel}} = 10$, $\alpha = 0.3$, $D = 0.03$.

causal influence is not coincident with the network of direct influences because we also have to consider as causal all the indirect influences. Let us define the parents P_x of a node x as the set of all nodes in the network of direct influences that are able to reach x with directed paths. We expect all the parents P_x to have a causal influence on x , in general with different intensities and time scales. Similarly, we define the common parents P_{xy} of two nodes x and y as the set of all nodes in the network of direct influences that are able to reach both nodes x and y with directed paths.

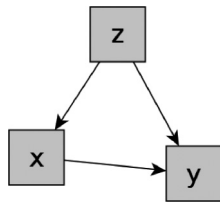
Then we generalize the definition of causal influence to the multidimensional case, adding the condition of the knowledge of the state of the common parents $P_{xy}(t)$ at time t to all the probability measures:

$$C_{x \rightarrow y}(\tau) = I(x(t), y(t + \tau) | P_{xy}(t)) - R(x(t), y(t); y(t + \tau) | P_{xy}(t)), \quad (17)$$

where $R(x(t), y(t); y(t + \tau) | P_{xy}(t))$ is defined as in Eq. (2), but with all the information measures conditioned to the knowledge of the state of the common parents $P_{xy}(t)$ at time t .

For simplicity we consider a network of three nodes without feedback, the so-called feed-forward loop:

$$\begin{aligned} \frac{dz}{dt} &= -\frac{z}{t_{\text{rel}}} + \sqrt{D_z} \Gamma_z(t), \\ \frac{dx}{dt} &= \alpha_x z - \beta_x x + \sqrt{D_x} \Gamma_x(t), \\ \frac{dy}{dt} &= \alpha_y z - \beta_y y + \gamma x + \sqrt{D_y} \Gamma_y(t). \end{aligned} \quad (18)$$



When the $x \rightarrow y$ interaction parameter is 0, $\gamma = 0$, the variable x is not a parent of y and therefore it should have no causal influence on it. Still, x and y can be highly correlated

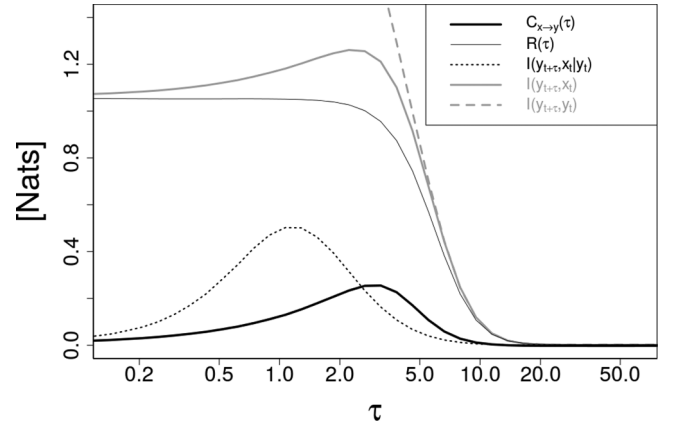


FIG. 9. Feed-forward loop, the three-dimensional general case. Causal influence $x \rightarrow y$ (numerical simulation). The parameters are $t_{\text{rel}} = 10$, $\gamma = \alpha_x = \alpha_y = 1$, $\beta_x = \beta_y = 0.2$, $D_z = 10$, $D_x = D_y = 0.1$.

due to the common parent z . Applying the above definition we analytically calculated the causal influence of x on y and it proved to be 0 (see the Appendix), $C_{x \rightarrow y}(\tau) = I(x(t), y(t + \tau) | z(t)) - R(x(t), y(t); y(t + \tau) | z(t)) = 0$. When $\gamma \neq 0$ our causal influence measure $C_{x \rightarrow y}(\tau)$ would detect the presence of the $x \rightarrow y$ influence (numerical results in Fig. 9). The shape of $C_{x \rightarrow y}(\tau)$ is qualitatively the same as in the BLRM. We verified numerically that the causal influence is correctly 0 for the cases $C_{y \rightarrow x} = C_{x \rightarrow z} = C_{y \rightarrow z} = 0$. The transfer entropy $T_{x \rightarrow y}(\tau) = I(y(t + \tau), x(t) | y(t), z(t))$ goes to 0 for $\tau \rightarrow 0$ because the white-noise term $\sqrt{D_y} \Gamma_y$ dominates the dynamics for short intervals.

Importantly, the fact of having a very small (negligible) direct interaction $z \rightarrow x$, i.e., $\alpha_x \ll \alpha_y$, implies that the probability distributions in the calculation of the causal influence $C_{x \rightarrow y}$ have to be conditioned on the common parent $z(t)$. On the contrary, without a direct interaction $z \rightarrow x$, i.e., $\alpha_x = 0$, z is not a common parent and therefore there is no conditioning on $z(t)$. However, as it should be, the conditioning makes no difference in the limit $\alpha_x \rightarrow 0$: we numerically verified that $\lim_{\alpha_x \rightarrow 0} C_{x \rightarrow y} = C_{x \rightarrow y}(\alpha_x = 0)$.

We note that even without a direct interaction $z \rightarrow y$, that is, $\alpha_y = 0$, the causal influence $C_{z \rightarrow y}$ can be positive due to the indirect influence $z \rightarrow x \rightarrow y$. The larger is the number of indirect passages between the considered nodes, the longer is the time period τ after which the peak of the causal influence is seen.

Here the conditioning on the common parents $P_{xy}(t)$ can be seen as a negative feature since it introduces again in the multidimensional case the synergistic effects, which are properties of the transfer entropy and of any conditioning [19]. However, there are no other possibilities since the construction of a PID lattice such as the one proposed by Williams and Beer [6] would require a generalization of our definition of redundancy $R(\tau)$ to more than two sources. For this we would need to have an expression for the mutual information I_{xy} between more than two variables, which is not defined in information theory [7].

III. DISCUSSION

Within the study of information flows in linear response models, our aim was to quantify how the effects of the asymmetric interactions are seen over time. In this work we propose a measure of causal influence $C_{x \rightarrow y}$. This is defined as the unique information [8] on the evolution of the response $y(t + \tau)$ given by the signal $x(t)$, i.e., the difference between the mutual information $I(x(t), y(t + \tau))$ and the redundancy $R(x(t), y(t); y(t + \tau))$. This is based on the measure of redundant information R , which we define as a composition of the mutual information between the two sources $I(x(t), y(t))$ and the total information that they give together on the target $I[(x(t), y(t)), y(t + \tau)]$. We did not derive the formula for this composition [Eq. (2)] from first principles; it is a choice inspired by an analogy with Markov chains. We stress that the causal influence is a function of the mutual information, of the time-lagged mutual information, and of the transfer entropy. In linear response models the causal influence is a peak function of time starting from 0 at $\tau = 0$, meaning that the effects of the asymmetric interactions are seen gradually over time, reflecting our view of the effects being visible only after the causes. Importantly, when there is no influence of a variable x on the dynamics of another variable y , i.e., in the network of direct influences there is no directed path starting from x and arriving in y , the causal influence is correctly 0 even if the two variables are highly correlated. The main difference in our information decomposition from the previously defined ones is that our redundancy R is explicitly dependent on the information shared between the two sources giving the redundant information on the target and is always less than or equal to that. Taking as redundancy just the minimum mutual information leads to threshold effects that do not seem to be appropriate for describing linear response models.

Finally, we discuss the strong limitation of the present work: there is currently no way of extending the definition of causal influence to any system with a general feedback structure. First, it is not clear whether the concept of causal influence still makes sense in the presence of feedback. When the variable x is influencing the dynamics of the variable y , and vice versa, forming a feedback loop, we can no longer define a signal and a response. The $x(t)$ at time t is influencing the evolution of the response $y(t + \tau)$ at time $t + \tau$ in many ways: directly and also indirectly through the loop $x(t) \rightarrow y(t + t') \rightarrow x(t + t'') \rightarrow y(t + \tau)$ with $\tau > t'' > t' > 0$, but also through the loops $x(t) \rightarrow y(t + t') \rightarrow x(t + t'') \rightarrow y(t + t''') \rightarrow x(t + t''') \dots \rightarrow y(t + \tau)$, and so on. These “successive” influences are in opposing directions for negative feedback loops and this implies that the information measures oscillate over time and our measure of causal influence oscillates as well and also assumes negative values. Since the mutual information $I(x(t), y(t + \tau))$ periodically assumes 0, we would have an oscillating causal influence with

any definition of R and we may conclude that the point-to-point communication scheme $(t, t + \tau)$ is not appropriate for a definition of causal influence in the presence of feedback.

Another option could be to take the original definition of transfer entropy (it considers the entire history of the studied processes) for discrete time sequences [16] and generalize it for continuous signals: $T_{x \rightarrow y}^{\text{seq}}(\tau) = I[y(t + \tau), (x(t - t'))_{t' \geq 0} | (y(t - t'))_{t' \geq 0}]$, where $(x(t - t'))_{t' \geq 0}$ and $(y(t - t'))_{t' \geq 0}$ are the semi-infinite whole history of the signal and response, respectively. For any set of stochastic differential equations without delays (also in the presence of feedback), if we already have the knowledge of the present values of the signal $x(t)$ and of the response $y(t)$, then the past of the signal $(x(t - t'))_{t' \geq 0}$ gives no additional information on the future of the response $y(t + \tau)$ and $T_{x \rightarrow y}^{\text{seq}}(\tau) = I[y(t + \tau), x(t) | (y(t - t'))_{t' \geq 0}] = I[y(t + \tau), x(t) | y(t), P[x(t) | (y(t - t'))_{t' \geq 0}]]$. This function is difficult to estimate even in linear response models. However, in the BLRM, since the knowledge of $(y(t - t'))_{t' \geq 0}$ gives an infinitely large amount of information on $x(t)$, the generalized transfer entropy will be 0, $T_{x \rightarrow y}^{\text{seq}}(\tau) = 0$, and this will be the case also for any other bidimensional model (also in the presence of feedback) with no intrinsic noise in the dynamics of the response. The BLRM is the prototype in whose frame we quantify the evident influence that the signal has on the response and we would not be satisfied to identify the generalized transfer entropy $T_{x \rightarrow y}^{\text{seq}}(\tau)$, which is here always 0, as a measure of the causal influence.

An additional problem in defining a measure of causal influence could be the partial information decomposition itself. What is the meaning of decomposing the mutual information? The mutual overlap between the two probability distributions $P(y(t + \tau) | y(t))$ and $P(y(t + \tau) | x(t))$ could be defined as $\langle \int_{-\infty}^{\infty} P(y(t + \tau) | y(t)) P(y(t + \tau) | x(t)) dy(t + \tau) \rangle_{x(t), y(t)}$ but then it is not easy to say whether this quantity can be used to define a measure of overlap in the space of Shannon entropies, i.e., the common information (redundancy) of $I(y(t + \tau), y(t))$ and $I(y(t + \tau), x(t))$.

In general, the causal influence quantifies the effective strength of asymmetric causal interactions and the time scale over which the effects are seen. Our measure is a good description of the dynamics of influences in linear response models. However, a generalization to nonlinear and non-Gaussian dynamics and for systems that have feedback is needed to approach the data analysis in complex systems.

ACKNOWLEDGMENTS

We thank A. De Martino, who gave seminal suggestions about this work, and M. Scazzocchio for data management and graph contributions. Work at Humboldt-Universität zu Berlin was supported by the German Research Council, grant RTG 1772 for Computational Systems Biology.

APPENDIX: THE CORRELATION DUE TO A COMMON PARENT RESULTS IN ZERO CAUSAL INFLUENCE

We study the particular case of the system of Eq. (18) without influence of the x on the y , i.e., with $\gamma = 0$. We calculate here all the information measures needed to show that the causal influence $C_{x \rightarrow y}(\tau)$ is 0. We start from those quantities which are

found also in the BLRM. The conditional expectation values and the standard deviations for the couples zx and zy are symmetric so we write them just once:

$$\begin{aligned} \langle z(t - \tau + t')z(t - \tau + t' + t'')|z(t) \rangle &= \int_{-\infty}^{+\infty} P(z(t - \tau + t' + t'') = \xi|z(t)) \xi \langle z(t - \tau + t')|z(t - \tau + t' + t'') = \xi \rangle d\xi \\ &= e^{-\frac{t''}{t_{\text{rel}}}} \left(z^2(t) e^{-\frac{2(\tau-t'-t'')}{t_{\text{rel}}}} + \sigma_z^2 \left(1 - e^{-\frac{2(\tau-t'-t'')}{t_{\text{rel}}}} \right) \right), \end{aligned} \quad (\text{A1})$$

$$\begin{aligned} \langle z(t - \tau + t')x(t - \tau + t')|z(t) \rangle &= \int_{-\infty}^{+\infty} P(z(t - \tau + t') = \xi|z(t)) \xi \langle x(t - \tau + t')|z(t - \tau + t') = \xi \rangle d\xi \\ &= \frac{\alpha_x t_{\text{rel}}}{\beta_x t_{\text{rel}} + 1} \left(z^2(t) e^{-\frac{2(\tau-t')}{t_{\text{rel}}}} + \sigma_z^2 \left(1 - e^{-\frac{2(\tau-t')}{t_{\text{rel}}}} \right) \right), \end{aligned} \quad (\text{A2})$$

$$\begin{aligned} \langle z(t - \tau + t')x(t)|z(t) \rangle &= \langle z(t - \tau + t')x(t - \tau + t')|z(t) \rangle e^{-\beta_x(\tau-t')} \\ &\quad + \alpha_x \int_0^{\tau-t'} \langle z(t - \tau + t')z(t - \tau + t' + t'')|z(t) \rangle e^{-\beta_x(\tau-t'-t'')} dt'' \\ &= \sigma_z^2 \frac{2\alpha_x t_{\text{rel}}}{\beta_x^2 t_{\text{rel}}^2 - 1} \left(e^{-\frac{\tau-t'}{t_{\text{rel}}}} - e^{-\beta_x(\tau-t')} \right) + z^2(t) \frac{\alpha_x t_{\text{rel}}}{\beta_x t_{\text{rel}} + 1} e^{-\frac{\tau-t'}{t_{\text{rel}}}}, \end{aligned} \quad (\text{A3})$$

$$\sigma_y^2 = \sigma_z^2 \frac{\alpha_y^2 t_{\text{rel}}}{\beta_y(1 + \beta_y t_{\text{rel}})} + \frac{D_y}{2\beta_y}, \quad (\text{A4})$$

$$\sigma_{y(t+\tau)|z(t)}^2 = \sigma_y^2 - \left(\frac{\sigma_z \alpha_y t_{\text{rel}}}{\beta_y t_{\text{rel}} - 1} \right)^2 \left(e^{-\frac{\tau}{t_{\text{rel}}}} - \frac{2e^{-\beta_y \tau}}{\beta_y t_{\text{rel}} + 1} \right)^2, \quad (\text{A5})$$

$$\begin{aligned} \sigma_{y(t+\tau)|x(t), y(t), z(t)}^2 &= \sigma_{y(t+\tau)|y(t), z(t)}^2 = \frac{\sigma_z^2 \alpha_y^2}{\beta_y(\beta_y + 1/t_{\text{rel}})(\beta_y - 1/t_{\text{rel}})^2} \left[(1 - \beta_y t_{\text{rel}})^2 - e^{-2\beta_y \tau} (1 + \beta_y t_{\text{rel}}) \right. \\ &\quad \left. + e^{-(\beta_y + \frac{1}{t_{\text{rel}}})\tau} 4\beta_y t_{\text{rel}} - e^{-\frac{2\tau}{t_{\text{rel}}}} \beta_y t_{\text{rel}} (1 + \beta_y t_{\text{rel}}) \right] + \frac{D_y}{2\beta_y} (1 - e^{-2\beta_y \tau}), \end{aligned} \quad (\text{A6})$$

$$\langle y(t - \tau)x(t)|z(t) \rangle = \langle y(t)x(t)|z(t) \rangle e^{\beta_y \tau} - \alpha_y \int_0^\tau \langle z(t - \tau + t')x(t)|z(t) \rangle e^{\beta_y t'} dt'. \quad (\text{A7})$$

$\langle y(t - \tau)x(t)|z(t) \rangle \rightarrow 0$ for $\tau \rightarrow \infty$ because the knowledge of $x(t)$ gives asymptotically no information on the distant past of y [even with the condition $z(t)$]; then

$$\langle y(t)x(t)|z(t) \rangle = \frac{\alpha_x \alpha_y t_{\text{rel}}^2}{(\beta_x t_{\text{rel}} + 1)(\beta_y t_{\text{rel}} + 1)} \left(z^2(t) + \frac{2\sigma_z^2}{t_{\text{rel}}(\beta_x + \beta_y)} \right). \quad (\text{A8})$$

Since $\langle z(t + t')x(t)|z(t) \rangle = \langle z(t + t')|z(t) \rangle \langle x(t)|z(t) \rangle$, whose quantities we know from Eqs. (6) and (7), we can now calculate $\langle y(t + \tau)x(t)|z(t) \rangle = \langle y(t)x(t)|z(t) \rangle e^{-\beta_y \tau} + \alpha_y \int_0^\tau \langle z(t + t')x(t)|z(t) \rangle e^{-\beta_y(\tau-t')} dt'$ and then

$$\langle y(t + \tau)x(t)|z(t) \rangle - \langle y(t + \tau)|z(t) \rangle \langle x(t)|z(t) \rangle = \sigma_z^2 \frac{2\alpha_x \alpha_y t_{\text{rel}} e^{-\beta_y \tau}}{(\beta_x t_{\text{rel}} + 1)(\beta_y t_{\text{rel}} + 1)(\beta_x + \beta_y)}, \quad (\text{A9})$$

which is independent of the condition $z(t)$, as is typically the case for linear systems. The information measures are easily calculated in the Gaussian case:

$$I_{\text{tot}} = \ln \left(\frac{\sigma_{y(t+\tau)|z(t)}^2}{\sigma_{y(t+\tau)|x(t), y(t), z(t)}^2} \right), \quad (\text{A10})$$

$$C(x(t), y(t + \tau)|z(t)) = \frac{\langle y(t + \tau)x(t)|z(t) \rangle - \langle y(t + \tau)|z(t) \rangle \langle x(t)|z(t) \rangle}{\sigma_{y(t+\tau)|z(t)} \sigma_{x(t)|z(t)}}, \quad (\text{A11})$$

$$I(x(t), y(t + \tau)|z(t)) = -\frac{1}{2} \ln[1 - C^2(x(t), y(t + \tau)|z(t))]. \quad (\text{A12})$$

Using the definition of redundancy [Eq. (2)], $R(\tau) = \frac{1}{2} \ln \left(\frac{e^{2(I_{xy} + I_{\text{tot}})}}{e^{2I_{xy}} + e^{2I_{\text{tot}}}} \right)$, with $I_{xy} = -\frac{1}{2} \ln(1 - C^2(x(t), y(t)|z(t)))$, we obtain the expected result:

$$C_{x \rightarrow y}(\tau) = I(x(t), y(t + \tau)|z(t)) - R(x(t), y(t); y(t + \tau)|z(t)) = 0. \quad (\text{A13})$$

- [1] A. Schopenhauer, *The World as Will and Representation* (Courier Corp., North Chelmsford, MA, 2012).
- [2] W. Bialek and S. Setayeshgar, *Proc. Natl. Acad. Sci. USA* **102**, 10040 (2005).
- [3] G. Tkačik, C. G. Callan, and W. Bialek, *Proc. Natl. Acad. Sci. USA* **105**, 12265 (2008).
- [4] B. N. Kholodenko, *Nat. Rev. Mol. Cell Biol.* **7**, 165 (2006).
- [5] C. Jarzynski, *Annu. Rev. Condens. Matter Phys.* **2**, 329 (2011).
- [6] P. L. Williams and R. D. Beer, [arXiv:1004.2515](https://arxiv.org/abs/1004.2515).
- [7] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (John Wiley & Sons, New York, 2012).
- [8] A. B. Barrett, *Phys. Rev. E* **91**, 052802 (2015).
- [9] G. E. Uhlenbeck and L. S. Ornstein, *Phys. Rev.* **36**, 823 (1930).
- [10] D. T. Gillespie, *Phys. Rev. E* **54**, 2084 (1996).
- [11] R Core Team, *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria, 2014).
- [12] H. Risken, *The Fokker-Planck Equation*, 2nd ed. (Springer, Berlin, 1989), pp. 153–158.
- [13] U. M. B. Marconi, A. Puglisi, L. Rondoni, and A. Vulpiani, *Phys. Rep.* **461**, 111 (2008).
- [14] I. Nemenman, *Phys. Biol.* **9**, 026003 (2012).
- [15] K. Christoph, T. Marc, and B. Demian, *Nature Commun.* **7**, 11061 (2016).
- [16] T. Schreiber, *Phys. Rev. Lett.* **85**, 461 (2000).
- [17] C. W. Granger, *Econometrica: J. Econometr. Soc.* **37**, 424 (1969).
- [18] L. Barnett, A. B. Barrett, and A. K. Seth, *Phys. Rev. Lett.* **103**, 238701 (2009).
- [19] R. G. James, N. Barnett, and J. P. Crutchfield, *Phys. Rev. Lett.* **116**, 238701 (2016).
- [20] E. Schneidman, W. Bialek, and M. J. Berry, *J. Neurosci.* **23**, 11539 (2003).