


Generative formalism of causality quantifiers for processesDmitry A. Smirnov **Saratov Branch, Kotelnikov Institute of Radio Engineering and Electronics of the Russian Academy of Sciences, 38 Zelyonaya St., Saratov 410019, Russia* (Received 8 August 2020; revised 4 December 2020; accepted 1 March 2022; published 25 March 2022)

The concept of dynamical causal effect (DCE) is generalized and equipped with a formalism which allows one to formulate in a unified manner and interrelate a variety of causality quantifiers used in time series analysis. An elementary DCE from a subsystem Y to a subsystem X is defined within the stochastic dynamical systems framework as a *response* of a future X state to an appropriate *variation* of an initial (X, Y) -state distribution or a certain parameter of Y or of the coupling element $Y \rightarrow X$; this response is quantified in a probabilistic sense via a certain *distinction functional*; elementary DCEs are assembled over a set of *initial variations* via an *assemblage functional*. To include all those aspects, a “triple brackets formula” for the general DCE is suggested and serves as a first principle to produce specific causality quantifiers as realizations of the general DCE. As an application, transfer entropy and Liang-Kleeman information flow are related surprisingly as opposite limit cases in a family of DCEs; it is shown that their “nats per time unit” may differ drastically. The suggested DCE viewpoint links any formal causality quantifier to “intervention-effect” experiments, i.e., future responses to initial variations, and so provides its dynamical interpretation, opening a way to its further physical interpretations in studies of physical systems.

DOI: [10.1103/PhysRevE.105.034209](https://doi.org/10.1103/PhysRevE.105.034209)**I. INTRODUCTION**

“Does a temporally evolving system (a process) influence another one?” is a central question in many studies [1–41]. If “yes,” one further asks whether such an influence (also called causal or directional coupling) is strong in some sense. Otherwise, this “yes” is not so informative. The former question is that of “coupling detection” [3] or “causal discovery” [40]. The latter question is that of “quantitative characterization of directional coupling” or “estimation of causality quantifiers.” Well-known monographs [42,43] tell us how to define and quantify causal couplings for random variables, but the situation is more problematic for processes. In the latter case, a causal coupling *exists* if a variation in one system *at a given time instant* produces a nonzero *future response* of another system. However, existence of a coupling is yet a *qualitative* statement. Therefore, numerous causality *quantifiers* have been developed and are still being suggested. It provokes multiple discussions concerning their relevance and meaning [23,27,29,32,35,44–49]. For example, the transfer entropy (TE) [1] is a celebrated concept [50] which generalizes the famous Wiener-Granger causality [51,52] and is said to express “information flow” [50], “information transfer” [1,13,17,53], “information transport” [1], “directed statistical coherence” [53], etc. Still, a critical remark [32] has shown that the TE interpretation is not always clear. Hot debates have recently occurred around spectral causalities [45–49], and this series of examples may readily be continued.

Various causality quantifiers for processes are used everywhere, from nuclear reactors [54], communication [55], and galactic cosmic rays [56] to ecology [20], neuroscience [10,12,45,57–62], and climate science [28,40,41,63–72]. Any of those quantifiers is often considered or newly introduced as a separate measure, independent of the others and valuable *per se*. The resulting controversy is that any causal coupling in a complex system may be stronger than couplings in other directions according to one quantifier and weaker according to another quantifier [29]. For example, spectral causalities have been estimated from nuclear reactor data [54] with an interesting conclusion about possible causes of an observed anomaly inferred from larger values of such causalities for certain pairs of signals. But what if another causality quantifier gives a different conclusion?

Spectral causalities related to the information flow in the TE sense [47,61] are widely applied to neuroimaging studies. Their review [61] presents them as an important tool in line with fMRI, EEG, and MEG techniques. However, another line of research [5,73–78] with solid basis in mathematical physics and many applications to climate science (e.g., [67,71,72,76]) develops the Liang-Kleeman information flow (LKIF) as a “rigorous notion *ab initio*” [78]. The LKIF is measured in “nats per time unit” like the TE rate, and both are often used as (nonlinear) causality quantifiers. Which of the two information flows is more appropriate and where? Not much attention has been paid to resolving such controversies and revealing whether a coupling is quantified in an appropriate sense.

Moreover, one often argues that the causal language and causal interpretations in time series analysis may often be improper, and so it may be preferable to refuse the very term “causality” in that field. The authors of Ref. [50] claim that

*smirnovda@yandex.ru

the debate on the concept of causality “has generated rather more heat than light” (p. 83) and suggest just to use a certain approach (e.g., the Granger causality) as “*a* (as opposed to *the*) notion of causality” (p. 83). If so, several questions remain. How are many possible approaches interrelated? Is each of them equally valid? Which of them is a better tool to reveal and quantify causal couplings in a concrete study?

To address all the above issues systematically, one would need a concrete formalism to derive various causality quantifiers from a well-grounded general concept of causality as from a “first principle.” As such a basis, one can readily take the interventionist approach of Pearl [42], who argues convincingly that the success of this approach is due to the fact that “causality has been mathematized” (Ref. [42], p. xiii). That approach has been applied to a stochastic dynamical system (SDS) in Ref. [29] where the concept of dynamical causal effect (DCE) has been introduced with further developments in Refs. [49,79–82]. In particular, the TE has been shown to be approximately equal to a certain short-term DCE and related to several long-term DCEs under some conditions [81]. Still, the existing DCE concept is insufficient to formulate the LKIF as a DCE, to relate the TE to a certain DCE *exactly*, and to interrelate the two quantifiers within the DCE framework. Here the DCE concept is generalized and equipped with more detailed and exact formalism and terminology. Namely, the DCE is expressed via the “triple brackets formula” which produces specific causality quantifiers (including the LKIF and the TE) as its realizations. The enriched terminology and notations allow one to formulate relations between causality quantifiers in a short and precise manner.

Note that the basic problem here is *not an inverse problem* of inferring couplings from data, *but a direct problem* of quantifying couplings for a given system. After solving the direct problem, one can return to the practically important inverse problem with new tools. “A given system” here is an SDS consisting of two subsystems X and Y characterized at time t with vectors x_t and y_t of arbitrary finite dimensions which constitute together a state vector (x_t, y_t) of the full SDS. The SDS is understood in the sense of *Markovian* random dynamical system [83], i.e., its state vector (x_t, y_t) uniquely determines probability density functions (PDFs) of all future states. It is a close generalization of the concept of deterministic dynamical system (e.g., [84–86]) and so represents a basic paradigm in physics (e.g., [87–89]). Hence, the DCE framework developed here is as general.

It is implied here that evolution of an SDS may also depend on a parameter vector a remaining constant through time. So, for a given SDS, a researcher can specify any initial state (x_0, y_0) and a parameter a (if any) and observe future states in arbitrarily many independent experiments (trials) to compare ensembles of time realizations under different conditions. This setting is encountered in cases of, e.g., (1) fully known evolution equations of an SDS, (2) a “black box” algorithm implementing evolution of an SDS where a researcher can provide initial states and parameters as input and get future states as output, and (3) a real-world system which can be manipulated and whose dynamics is argued to possess the SDS properties (approximately). The entire approach is based neither on explicitly given evolution equations nor on passively observed time series, but relies on the general definition

of the SDS and can be applied to both those situations. Due to such a generality, it appears to be capable of producing numerous causality quantifiers as specific DCEs and revealing their common and distinctive features.

Section II provides a typical example of a controversial situation with the TE and the LKIF as causality quantifiers and suggests the generalized concept of DCE with a more developed terminology (a mini-language). Its application to interpreting and interrelating the two information flows is given in Sec. III with unexpected results for the LKIF and an exact derivation of the TE as a DCE. The formalism readily extends to more than two subsystems and to partly observed states as discussed in Sec. IV together with other perspectives. Conclusions are given in Sec. V. Details are left to the Appendixes and Supplemental Material [90].

II. CAUSALITY QUANTIFIERS FOR PROCESSES

Section II A presents an illustrative example of coupled overdamped oscillators where the TE and the LKIF values drastically differ from each other, so the same directional coupling is qualified *simultaneously* as very strong according to the TE and arbitrarily weak or even zero according to the LKIF. Section II B introduces the concept of DCE in a more general way and with more details than was done previously [29].

A. Example of controversy

The SDS which describes two overdamped oscillators X and Y (e.g., [49,79,81]) and is widely used as a simple model of irregular processes (e.g., [91,92]) reads

$$\begin{aligned}\dot{x} &= -a_x x + a_{xy} y + \xi_{x,t}, \\ \dot{y} &= -a_y y + a_{yx} x + \xi_{y,t},\end{aligned}\quad (1)$$

where (ξ_x, ξ_y) is a bivariate zero-mean uncorrelated Gaussian white noise with intensities $(\Gamma_{xx}, \Gamma_{yy})$, i.e., $\langle \xi_{x,t} \xi_{x,t'} \rangle = \Gamma_{xx} \delta(t - t')$, $\langle \xi_{y,t} \xi_{y,t'} \rangle = \Gamma_{yy} \delta(t - t')$, and $\langle \xi_{x,t} \xi_{y,t'} \rangle = 0$, where angle brackets denote expectation. Relaxation times of the uncoupled processes X_t and Y_t (i.e., for $a_{xy} = a_{yx} = 0$) are $t_x = 1/a_x$ and $t_y = 1/a_y$. Their ratio is $m_{xy} = a_y/a_x = 1/m_{yx}$. If $m_{xy} > 1$, the coupling $Y \rightarrow X$ is “from a fast source Y to a slow recipient X ” [81]. The product $a_{xy} a_{yx} > 0$ corresponds to positive feedback, while $a_{xy} a_{yx} < 0$ to negative feedback. This product is zero for unidirectional coupling (i.e., no feedback).

Figure 1(a) shows a single time realization of the uncoupled processes X_t and Y_t whose stationary variances equal $\sigma_{X,0}^2 = \Gamma_{xx}/(2a_x)$ and $\sigma_{Y,0}^2 = \Gamma_{yy}/(2a_y)$. Let us use relative coupling parameters $\beta_{xy} = a_{xy} \sigma_{Y,0}/(a_x \sigma_{X,0})$ and $\beta_{yx} = a_{yx} \sigma_{X,0}/(a_y \sigma_{Y,0})$ [81] and consider a bidirectional coupling with $\beta_{xy}/\beta_{yx} = -m_{xy}$ which corresponds to “relatively equivalent couplings with negative feedback” [81]. Figure 1(b) shows a time realization of the process (X_t, Y_t) for such coupling and equal relaxation times, i.e., for $a_x = a_y$ and $\beta_{xy} = -\beta_{yx}$, whose oscillatory character essentially differs from the uncoupled case. For the same bidirectional coupling, Figure 1(c) compares two ensembles of realizations of X_t for the same value of x_0 and two different values of y_0 (three realizations from each ensemble). The two sets of realizations

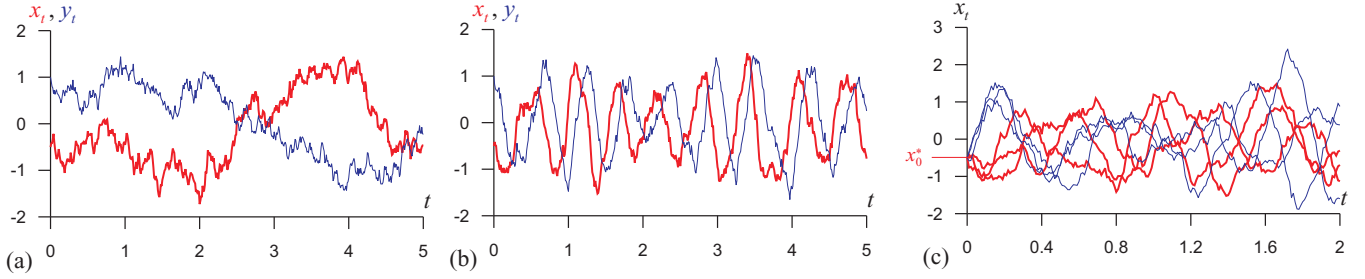


FIG. 1. Time realizations of the SDS (1) with $a_x = a_y = 1$ and $\Gamma_{xx} = \Gamma_{yy} = 1$: (a) a realization of X_t (red thick line) and Y_t (blue thin line) for $a_{xy} = a_{yx} = 0$ starting from $x_0^* = -0.5, y_0^* = 1$; (b) a realization of X_t (red thick line) and Y_t (blue thin line) for $a_{xy} = -a_{yx} = -10$ starting from $x_0^* = -0.5, y_0^* = 1$; (c) realizations of X_t for $a_{xy} = -a_{yx} = -10$ starting from $x_0^* = -0.5, y_0^* = 1$ (red thick lines) and $x_0^* = -0.5, y_0^{**} = -1$ (blue thin lines). The two ensembles differ from each other considerably for $t < 0.3$ and are statistically indistinguishable for $t > 1.2$ [see also Fig. 2(a)].

essentially differ from each other for the near future ($t < 0.3$) which is another manifestation of the influence $Y \rightarrow X$.

For the bidirectional coupling under consideration, let us compare numerical values of the TE rate and the LKIF as two widely used “information flows” characterizing causal couplings. Their definitions are recapitulated in Sec. III along with specific formulas. Here just their values are reported to highlight a controversial situation. Let us denote $\tau_{Y \rightarrow X}$ the TE rate, i.e., the derivative of the TE with respect to its temporal horizon (Sec. III A). Let us denote $l_{Y \rightarrow X}$ the LKIF for the stationary initial PDF as usually defined (Sec. III B). The TE rate for $\beta_{xy}/\beta_{yx} = -m_{xy}$ reads

$$\tau_{Y \rightarrow X} = a_x \beta_{xy}^2 / 4. \quad (2)$$

The value of $|\beta_{xy}|$ may rise unboundedly retaining stationarity of the process and leading to greater “oscillation period” in Fig. 1(b). Hence, the TE rate can become arbitrarily large too. The LKIF for the SDS (1) with any parameters reads [76]

$$l_{Y \rightarrow X} = a_{xy} \sigma_{XY} / \sigma_X^2, \quad (3)$$

where σ_X^2 is the stationary variance of X_t and σ_{XY} is the stationary cross-covariance $\langle X_t Y_t \rangle$. It can be shown that $\sigma_{XY} = 0$ and, hence, $l_{Y \rightarrow X} = 0$ for the bidirectional coupling with $\beta_{xy}/\beta_{yx} = -m_{xy}$. Thus, the coupling $Y \rightarrow X$ is arbitrarily strong when quantified with the TE rate and zero when quantified with the LKIF. A moderate variation of parameters in some vicinity of the above values keeps an arbitrarily large value of the TE rate and an arbitrarily small value of the LKIF, i.e. the controversial characterization of the “coupling strength” as simultaneously small and large is robust. Everything is the same for the coupling in the direction $X \rightarrow Y$.

As shown in Figs. 1(b) and 1(c), effects of the coupling $Y \rightarrow X$ on the dynamics of X are large. Hence, the large TE rate seems to characterize the coupling adequately, while the zero LKIF does not. Does it mean that the LKIF may not generally be called a “causality quantifier”? Then what about the solid argumentation [78] of the LKIF as a rigorous notion *ab initio*? On the other hand, in a wide range of situations both the TE rate and the LKIF take on similar values which are proportional to each other with a moderate factor. For example, for a unidirectional coupling $Y \rightarrow X$ (i.e., $a_{yx} = 0$) from a slow source (i.e., $a_x \gg a_y$), one derives $l_{Y \rightarrow X} \approx 4\tau_{Y \rightarrow X}$ (Sec. III D). Moreover, both $l_{Y \rightarrow X}$ and $\tau_{Y \rightarrow X}$ are zero as soon as the coupling $Y \rightarrow X$ is absent (i.e., $a_{xy} = 0$). So the LKIF

may well be a reasonable causality quantifier. Is the LKIF a relevant quantifier only sometimes? Is the TE generally better than the LKIF? To address these issues, the DCE formalism and terminology are suggested below on the basis of the interventional causality concept (Appendix B 1) combined with the dynamical systems viewpoint (Appendix B 2). A wide interdisciplinary context is given in Appendix A. This formalism will allow us to justify both the TE rate and the LKIF as possible causality quantifiers through explicating the interventional meaning in which they characterize “coupling strength” (Secs. III A and III B) and precisely formulate their common and distinctive features (Secs. III C and III D). More importantly, it applies to many other quantifiers serving as a general first principle as discussed in Sec. IV and Supplemental Material [90].

B. Generative formalism

In usual language, causality means that a *cause* as something independent produces an *effect*. For a simple quantitative expression, consider a random variable U whose PDF depends on another variable V . One denotes a conditional PDF of U for an *imposed* value of $V = v$ as $p(u|\text{do}(v))$, where the special notation $\text{do}(v)$ highlights the active imposition which is called *intervention* in the famous do-calculus [42]. If the interventional PDFs $p(u|\text{do}(v^*))$ and $p(u|\text{do}(v^{**}))$ for some pair of different values v^* and v^{**} differ from each other, one says that the causal coupling $V \rightarrow U$ exists. Its *causal effect* [42] is any measure of difference between the two PDFs. For example, one often uses the average causal effect (ACE) [42,93] as the difference of the conditional expectations $E(U|\text{do}(v^{**})) - E(U|\text{do}(v^*))$. Further details are given in Appendix B 1.

Consider any SDS which consists of two subsystems X and Y and denote it \mathcal{S} ; see Appendix C 4 for rigorous definitions including Markovianity of \mathcal{S} . Its full state vector (X_t, Y_t) is a combination of the two random vectors X_t and Y_t of arbitrary finite dimensions. Let $\xi_{(0,t)}$ be a random event which is a realization of noises which influence both subsystems over the time interval $[0, t]$. If the value (x_0, y_0) of the full initial state is given, no matter actively imposed or passively observed, the PDF of the future state (X_t, Y_t) at any $t > 0$ is uniquely determined. Then an evolution operator Φ_t maps the initial state (x_0, y_0) to the future random state (X_t, Y_t) . The latter is a random variable whose value in a single trial depends on

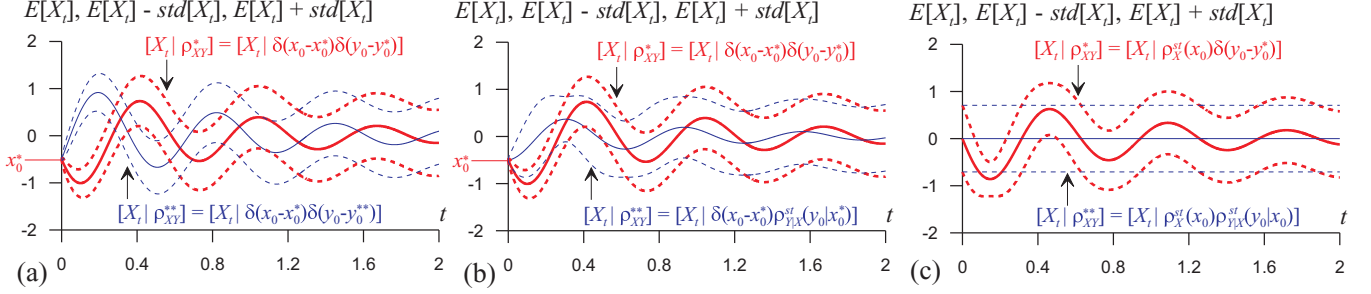


FIG. 2. Functionally conditional expectations of X_t (solid lines) \pm one standard deviation (dashed lines) under the reference (red thick lines) and alternative (blue thin lines) initial conditions for the SDS (1) with $a_x = a_y = 1$, $a_{xy} = -a_{yx} = -10$, $\Gamma_{xx} = \Gamma_{yy} = 1$, and $x_0^* = -0.5$, $y_0^* = 1$, $y_0^{**} = -1$: (a) the reference $\rho_{XY}^* = \delta(x - x_0^*)\delta(y - y_0^*)$ and the alternative $\rho_{XY}^{**} = \delta(x - x_0^*)\delta(y - y_0^{**})$ [see also Fig. 1(c)]; (b) the reference $\rho_{XY}^* = \delta(x - x_0^*)\delta(y - y_0^*)$ and the alternative $\rho_{XY}^{**} = \delta(x_0 - x_0^*)\rho_{Y|X}(y_0|y_0^*)$ as used in the TE definition; (c) the marginal PDF $\rho_X^*(x_0)$ with the reference conditional PDF $\delta(y_0 - y_0^*)$ and the alternative conditional PDF $\rho_{Y|X}(y_0|x_0)$ as used in the LKIF definition. In each panel, the two futures differ from each other considerably at small $t < 0.3$ and also at some greater t , e.g., $t \approx 0.8$. All plots are obtained by solving Eqs. (E6) and (E12).

$\xi_{(0,t)}$ realized in that trial. Such a mapping reads $(X_t, Y_t) = \Phi_t(x_0, y_0)$ with the two projections of the full operator Φ_t given by $X_t = \Phi_t^X(x_0, y_0)$ and $Y_t = \Phi_t^Y(x_0, y_0)$.

The causal coupling $Y \rightarrow X$ between the subsystems X and Y means that a change of an *initial* value y_0 produces a change of a *future* random variable X_t given an initial value $x_0 = x_0^*$. A general justification of such an understanding of causality is presented in Appendixes B 1 and B 2. Figure 1(c) presents an example of such a Y change performed and the corresponding X change produced. Let us define here that the produced change of X_t can be quantified in any way, e.g., as any change of its PDF [29] or a change of its particular value x_t given a particular noise realization $\xi_{(0,t)}^*$. Then the produced change can be quantified as any difference between the random vectors $X_t^* = \Phi_t^X(x_0^*, y_0^*)$ and $X_t^{**} = \Phi_t^X(x_0^*, y_0^{**})$. Let us specifically call this difference here *distinction* and define it as a certain functional, not compulsorily metrics or distance (see Appendix C 3). Let us denote such a *distinction functional* with the figure brackets and a double vertical delimiter as $\{X_t^* || X_t^{**}\}$.

A change of y_0 from y_0^* to y_0^{**} has been called *initial intervention* in Ref. [29]. However, due to Markovianity of the SDS \mathcal{S} , the PDF $p(x_t|x_0, y_0)$ for passive observations coincides with the interventional PDF $p(x_t|\text{do}(x_0), \text{do}(y_0))$ where the values (x_0, y_0) are imposed by intervention [42] (see Appendix B 2). Instead of interventional experiments, one can equivalently compare ensembles of time realizations *passively observed* after the states (x_0^*, y_0^*) and (x_0^*, y_0^{**}) each of which is encountered at many different time instants in a long observed time series. So the term “intervention” is not compulsory for an SDS, because real interventions are not necessary. One can also say “initial change,” “initial perturbation,” or “initial variation” from y_0^* to y_0^{**} . The last term is used here as more neutral.

To generalize the above concept of initial variation [29], let us recall that one often asks how an SDS evolves from an *ensemble of initial states* (e.g., [5,38,75,88]) rather than from a single state (x_0^*, y_0^*) . Let such an ensemble be described with a PDF $\rho_{XY}(x_0, y_0) = \rho_X(x_0)\rho_{Y|X}(y_0|x_0)$. A single initial state (x_0^*, y_0^*) represents an ensemble with a specific PDF $\rho_{XY}(x_0, y_0) = \delta(x_0 - x_0^*)\delta(y_0 - y_0^*)$; its behavior for the

SDS (1) is exemplified by the thick red lines in Fig. 2(b) and both thick red and thin blue lines in Fig. 2(a). Behavior of ensembles with initial PDFs which are not Dirac δ 's initial PDFs is exemplified with the thin blue lines in Fig. 2(b) and both thick red and thin blue lines in Fig. 2(c). Let us call the PDF ρ_{XY} a (*functional*) *initial condition* contrary to a single initial *state*. Then the future state (X_t, Y_t) is a random vector which depends on two independent random events: a particular initial state (x_0, y_0) drawn from ρ_{XY} and a particular noise realization $\xi_{(0,t)}$. The future X_t is determined by the function $X_t(x_0, y_0, \xi_{(0,t)})$ and an initial PDF $\rho_{XY}(x_0, y_0)$. In order to characterize the causal coupling $Y \rightarrow X$, one should compare the futures X_t^* and X_t^{**} for two different functional initial conditions ρ_{XY}^* and ρ_{XY}^{**} with the same marginal PDF $\rho_X(x_0)$ (see examples in Fig. 2), i.e., $\rho_{XY}^* = \rho_X(x_0)\rho_{Y|X}^*(y_0|x_0)$ and $\rho_{XY}^{**} = \rho_X(x_0)\rho_{Y|X}^{**}(y_0|x_0)$. This is because an uncoupled subsystem X exhibits the same future random state X_t for the same marginal initial PDF ρ_X independently of $\rho_{Y|X}$, so one gets $X_t^* = X_t^{**}$ and $\{X_t^* || X_t^{**}\} = 0$ in any sense. If the coupling $Y \rightarrow X$ exists, the futures X_t^* and X_t^{**} differ (i.e., $\{X_t^* || X_t^{**}\} \neq 0$) in some sense for some t and some $\rho_{Y|X}^* \neq \rho_{Y|X}^{**}$. Let us call the ordered pair $(\rho_{XY}^*, \rho_{XY}^{**})$ the *initial condition variation* which consists of the *reference initial condition* ρ_{XY}^* and the *alternative initial condition* ρ_{XY}^{**} .

Let us call the variable X_t , its PDF, and its statistical characteristics obtained under the initial condition ρ_{XY} *functionally conditional* and introduce shorthand notations. The notation $p(x_t|\cdot)$ is not appropriate since it stands for the ordinary conditioning on a single state. For the functionally conditional variable X_t itself, let us use another kind of brackets $[X_t|\rho_{XY}]$ and denote its PDF similarly as $p_X^{(t)}[x_t|\rho_{XY}]$. Note that the ordinary conditional PDF is $p_X^{(t)}(x_t|x_0^*, y_0^*) = p_X^{(t)}[x_t|\delta(x_0 - x_0^*)\delta(y_0 - y_0^*)]$ and so $p_X^{(t)}[x_t|\rho_{XY}] = \int p_X^{(t)}(x_t|x_0, y_0)\rho_{XY}(x_0, y_0)dx_0 dy_0$. Let us denote the functionally conditional variance $\text{var}[X_t|\rho_{XY}]$, the functionally conditional Shannon entropy $H[X_t|\rho_{XY}]$, etc. In particular, $H[X_t|\rho_{XY}] = -\int p_X^{(t)}[x_t|\rho_{XY}] \ln p_X^{(t)}[x_t|\rho_{XY}] dx_t$. Throughout the paper, all integrals are taken over the entire range of X_t .

Denote $\tilde{\Phi}_t^X$ an operator which maps an initial condition ρ_{XY} to a future random state $[X_t|\rho_{XY}]$, i.e., $[X_t|\rho_{XY}] =$

$\tilde{\Phi}_t^X(\rho_{XY})$. Everything is similar for $\tilde{\Phi}_t^Y$ and the full operator $\tilde{\Phi}_t$. So the square brackets $[X_t, Y_t | \rho_{XY}]$ with different $t > 0$ encode the entire functioning of an SDS (see Appendix C 4). Let us call the value of the distinction functional $\{[X_t | \rho_{XY}^*] || [X_t | \rho_{XY}^{**}]\}$ the *dynamical causal effect* (DCE) of the initial condition variation $(\rho_{XY}^*, \rho_{XY}^{**})$. This is an *elementary* DCE since it quantifies the response to a *single* initial variation. Note that various versions of the distinction functional are possible (Appendix C 4), “ranging” from the difference of expectations to the Kullback-Leibler divergence and so on.

The evolution of an SDS is often described with an operator $\Phi_t(x_0, y_0, a)$ which depends on a parameter vector a whose value remains constant through time. The vector a is often divided into four components $(a_x, a_y, a_{xy}, a_{yx})$ which describe individual subsystems (a_x for X and a_y for Y) and their couplings (a_{xy} for $Y \rightarrow X$ and a_{yx} for $X \rightarrow Y$). For example, if X and Y are damped oscillators, then a_x and a_y may include individual damping coefficients and natural frequencies, while a_{xy} and a_{yx} may include coefficients of a resistive or any other coupling. Formally, $a_{xy} = 0$ means that X evolves independently of Y , i.e., X_t does not depend on y_0, a_y and a_{yx} , given x_0 and a_x . Causality $Y \rightarrow X$ may then be characterized via a response of X_t to a change of a , i.e., either of a_{xy} or a_y . Such a change is also an initial variation, but a *parameter variation* in contrast to an *initial condition variation*. Let us call a combination of the functional initial condition ρ_{XY} and the parameter value a *generalized initial condition* and denote it $\theta = \{\rho_{XY}, a\}$. Then an initial variation in general is given by the ordered pair $(\theta^*, \theta^{**}) = (\{\rho_{XY}^*, a^*\}, \{\rho_{XY}^{**}, a^{**}\})$ which may include only an initial condition variation (if $a^* = a^{**}$), or only a parameter variation (if $\rho_{XY}^* = \rho_{XY}^{**}$), or both.

One can be interested in characterizing effects of many different initial condition variations $(\rho_{XY}^*, \rho_{XY}^{**})$. Their initial conditions may often be specified with a parameter vector λ as $(\rho_{XY,\lambda}^*, \rho_{XY,\lambda}^{**})$. For example, both initial conditions can be Dirac δ 's where λ includes coordinates of their locations $\lambda = (x^*, y^*, y^{**})$: $\rho_{XY,\lambda}^* = \delta(x_0 - x_0^*)\delta(y_0 - y_0^*)$ and $\rho_{XY,\lambda}^{**} = \delta(x_0 - x_0^{**})\delta(y_0 - y_0^{**})$. Comparing evolutions from different initial states is of interest since numerous states are encountered even in a single stationary time series. In general, one may vary some components of a and include such components of a^* and a^{**} into the vector λ , then a parameterized set of initial variations $(\theta_\lambda^*, \theta_\lambda^{**})$ describes both initial condition and parameter variations. Let us call *assemblage* any procedure for obtaining a single value quantifying the causal effect $Y \rightarrow X$ from a set of elementary DCEs $Y \rightarrow X$ determined for various $\lambda \in \Lambda$. Such an “assembled” value may be, e.g., a weighted average of an elementary DCE over Λ (i.e., an “aggregate” value) or a maximal value of an elementary DCE over Λ , etc. Concerning parameter variations, one is often interested in a single variation [29,49,79,80] to compare the dynamical regimes established for $t \rightarrow \infty$ at two different parameter values a^* and a^{**} , e.g., to assess thereby an “overall role” of the coupling $Y \rightarrow X$ by comparing the dynamics of X in a free regime (i.e., at $a_{xy}^* = 0$) and at a nonzero $a_{xy}^{**} \neq 0$. Below, if one considers a single initial variation, the assemblage is called *trivial*. In general, an assemblage is implemented via any relevant functional of an elementary DCE defined over a set Λ . Let us call it an *assemblage functional* and denote it $\langle \cdot \rangle_\Lambda$ even though it is not compulsorily a certain average.

Note that the distinction functional may also be defined in general so to depend on the parameters which are components of the above vector λ or just some additional quantities. Let us include any such parameters into the vector λ too and denote such a parameter-dependent distinction explicitly as $\{\cdot | \cdot\}_\lambda$. An assemblage is to be performed over the parameters of the distinction too.

Let us now define the general finite-time DCE via the above three kinds of brackets as “triple brackets”

$$\mathcal{C}_{Y \rightarrow X}^{(t)} = \langle \{[X_t | \theta_\lambda^*] || [X_t | \theta_\lambda^{**}]\}_\lambda \rangle_\Lambda, \quad (4)$$

where the distinction subscript λ may be omitted if the distinction functional does not depend on parameters. If either the initial conditions, the distinction, or the assemblage do not depend on certain component of λ , such a component may be omitted in the respective subscript of Eq. (4) while the other components are retained. If the temporal horizon t is not too large as compared to any characteristic time scale of the SDS \mathcal{S} , the DCE (4) can be called short-term or transient [29,79,80]. If time is continuous and $t \rightarrow 0$, then the DCE rate $c_{Y \rightarrow X} = d\mathcal{C}_{Y \rightarrow X}^{(t)}/dt|_{t=0}$ represents the “very short-term” DCE. Conversely, if $t \rightarrow \infty$, the DCE can be called asymptotic (stationary, equilibrium) or long-term [29,79,80]. Since the latter DCE can reflect long-term changes in dynamics under a parameter variation, e.g., under switching the coupling $Y \rightarrow X$ on or off, it is often of interest in practice [79,80,94]. Finally, a *distributed* temporal horizon can be defined as a vector $t = (t_1, \dots, t_N)$ with $X_t = (X_{t_1}, \dots, X_{t_N})$ which implies a comparison of future *segments*, e.g., via power spectral densities [49], and is briefly commented on only in Sec. IV.

All the above definitions are given for the causal coupling $Y \rightarrow X$. Obviously, everything is the same for the DCE in the direction $X \rightarrow Y$. Further mathematical details of the definitions are given in Appendix C 4.

To summarize the entire DCE definition, (1) one performs an initial variation $(\theta_\lambda^*, \theta_\lambda^{**})$, (2) the SDS evolution operator $[\cdot | \theta_\lambda]$ produces an X response on a temporal horizon t , (3) the distinction functional $\{\cdot | \cdot\}_\lambda$ provides an elementary DCE for the single initial variation, and (4) having results for different initial variations, the assemblage functional $\langle \cdot \rangle_\Lambda$ gives a particular DCE. Any reasonable choice of all these elements produces a concrete DCE as a meaningful causality quantifier by construction. So the general DCE (4) includes a variety of particular DCEs which are its *realizations*. This infinite-dimensional realm of DCEs contains whole families of quantifiers obtained through varying the above elements. Some DCEs can be estimated directly from a passively observed time series, while obtaining the others requires additional efforts and assumptions (see Sec. IV). Below, different particular quantifiers are denoted with different letters in place of \mathcal{C} , such as $T_{Y \rightarrow X}^{(t)}$, $L_{Y \rightarrow X}^{(t)}$, etc., and similarly for the DCE rates $\tau_{Y \rightarrow X}$, $l_{Y \rightarrow X}$, etc.

Many causality quantifiers have been suggested based on different ideas during the last two or three decades (e.g., [1,5]) and many others are currently (say, during the last two or three years) being suggested (e.g., [38]). An underlying relation between each newly invented quantifier and the DCE concept usually remains implicit. However, due to the general character of the “variation-response” formalism, originating

from the well-grounded concepts of interventional causality and stochastic dynamical system, and a flavor of necessity in its construction, I suppose that any quantity relevant as a *causality quantifier for processes* can be expressed in the form (4). If it is not yet expressed in that form, one should just recognize the corresponding initial variations, distinction, and assemblage. To apply this principle, the two information flows are “dissected” and interrelated below.

III. APPLICATION

The TE and the LKIF are shown below to be *exactly* certain DCEs. They are also related to each other and some other DCEs. This case study confirms the potential role of the DCE formalism as a general principle behind various causality quantifiers. The main results are formulated as theorems just to highlight their exact character.

A. Transfer entropy

The TE is an extremely popular quantifier studied and applied in numerous works, e.g., in a recent monograph [50]. Both the original and the infinite-history versions of the TE are investigated quantitatively and related to some DCEs in Ref. [81]. The original TE was introduced from the perspective of stationary Markov process prediction [1] as the average reduction of uncertainty in X_t if y_0 becomes known, given x_0 . So the TE $T_{Y \rightarrow X}^{(t)}$ is the average difference of the Shannon entropies of X_t conditioned on x_0^* and on (x_0^*, y_0^*) :

$$T_{Y \rightarrow X}^{(t)} = \iint [H(X_t|x_0^*) - H(X_t|x_0^*, y_0^*)] p_{XY}^{st}(x_0^*, y_0^*) dx_0^* dy_0^*, \quad (5)$$

where $p_{XY}^{st}(x, y) = p_X^{st}(x)p_{Y|X}^{st}(y|x)$ is a stationary PDF and $H(X_t|x_0^*, y_0^*) = -\int p_X^{(t)}(x|x_0^*, y_0^*) \ln p_X^{(t)}(x|x_0^*, y_0^*) dx$ is the Shannon entropy at time t of the ensemble which starts from the reference initial condition $\rho_{XY,\lambda}^* = \delta(x_0 - x_0^*)\delta(y_0 - y_0^*)$, i.e., the functionally conditional Shannon entropy $H[X_t|\rho_{XY,\lambda}^*]$. Similarly, $H(X_t|x_0^*) = -\int p_X^{(t)}(x|x_0^*) \ln p_X^{(t)}(x|x_0^*) dx$ is the Shannon entropy of the ensemble which starts from the alternative initial condition $\rho_{XY,\lambda}^{**} = \delta(x_0 - x_0^*)p_{Y|X}^{st}(y_0|x_0^*)$ where y_0 is freed to vary according to the conditional PDF $p_{Y|X}^{st}(y_0|x_0^*)$, i.e., this is $H[X_t|\rho_{XY,\lambda}^{**}]$. The assemblage parameter is $\lambda = (x_0^*, y_0^*)$. So the TE definition (5) is recognized immediately as a finite-time DCE (4) with the above initial condition variations, the distinction $H[X_t|\rho_{XY,\lambda}^{**}] - H[X_t|\rho_{XY,\lambda}^*]$, and the assemblage as the weighted average with $p_{XY}^{st}(x_0^*, y_0^*)$. It proves the following theorem.

Theorem 1. For any SDS \mathcal{S} , the transfer entropy (5) is a DCE (4) of the initial condition variations given by $\rho_{XY,\lambda}^* = \delta(x_0 - x_0^*)\delta(y_0 - y_0^*)$ and $\rho_{XY,\lambda}^{**} = \delta(x_0 - x_0^*)p_{Y|X}^{st}(y_0|x_0^*)$ with $\lambda = (x_0^*, y_0^*)$ on a finite temporal horizon t with the distinction $H[X_t|\rho_{XY,\lambda}^{**}] - H[X_t|\rho_{XY,\lambda}^*]$ and the assemblage $\langle \cdot \rangle_\lambda = \int \int (\cdot) p_{XY}^{st}(x_0^*, y_0^*) dx_0^* dy_0^*$.

Note that the TE is formulated as a DCE *exactly*, in contrast to the previous version of the formalism [29,81] where the TE was shown to be only *approximately* equal to a certain

DCE of initial *state* variations. This is achieved due to the generalized definition of the initial variation which includes functional initial conditions instead of initial states.

The TE is often called “information flow” [50], but sometimes “information transfer” [17] in contrast to the information flow of Ay and Polani [95]. Let us denote the latter $A_{Y \rightarrow X}^{(t)}$. For a first-order Markov process, $A_{Y \rightarrow X}^{(t)}$ corresponds to the alternative initial condition with y_0 freed to vary according to its *marginal* stationary PDF $p_Y^{st}(y_0)$, i.e., $\rho_{XY,\lambda}^* = \delta(x_0 - x_0^*)p_Y^{st}(y_0)$, instead of the *conditional* PDF $p_{Y|X}^{st}(y_0|x_0^*)$ used in the TE definition. If a stationary regime is close to synchrony, e.g., to the identical synchronization $x_t \approx y_t$ [96], the conditional PDF $p_{Y|X}^{st}(y_0|x_0^*)$ is close to a Dirac δ . Then variability of y_0 according to $p_Y^{st}(y_0)$ is much stronger than that according to $p_{Y|X}^{st}(y_0|x_0^*)$. Another difference of $A_{Y \rightarrow X}^{(t)}$ is that the assemblage is a weighted average with $p_X^{st}(x_0^*)p_Y^{st}(y_0^*)$, i.e., over mutually independent x_0^* and y_0^* , instead of the joint PDF $p_{XY}^{st}(x_0^*, y_0^*)$ used in the TE definition. Let us call any PDF of the kind $f(x_0^*)g(y_0^*)$ “randomized,” since it corresponds to randomization of X_0 and Y_0 in statistical experiments, i.e., to drawing the values of X_0 and Y_0 in each trial independently of each other (see, e.g., [42]).

For regimes close to synchrony, $A_{Y \rightarrow X}^{(t)}$ may strongly differ from $T_{Y \rightarrow X}^{(t)}$ and be more sensitive to coupling changes and more powerful for causal discovery [27,95]. However, as *quantifiers*, these two DCEs have just different meanings as effects of different initial variations. Indeed, a region of nonzero values of $p_{XY}^{st}(x_t, y_t)$ for the identical synchronization is localized at $x_t = y_t$. Then the TE is zero since it does not involve any nonzero variations of y_0 relatively to x_0 . In contrast, $A_{Y \rightarrow X}^{(t)}$ can be quite large because nonzero variations of y_0 are involved and the coupling $Y \rightarrow X$ is nonzero and may be even strong enough to provide stability of a synchronous regime. This large $A_{Y \rightarrow X}^{(t)}$ quantifies a short-term effect of initial variations of y_0 about x_0 performed independently of x_0 , while this zero $T_{Y \rightarrow X}^{(t)}$ shows a short-term effect of the (zero) variations of y_0 about x_0 which occur in the synchronous regime. If one is interested in studying transients from various initial states, $A_{Y \rightarrow X}^{(t)}$ is a more relevant quantifier since it characterizes a wider region in the state space. If one is interested only in what happens in an established regime (e.g., what electric currents flow through the wires connecting two electric circuits X and Y and so what power dissipates in those wires in an established regime), then $A_{Y \rightarrow X}^{(t)}$ is irrelevant, while $T_{Y \rightarrow X}^{(t)}$ provides necessary information. So the DCE viewpoint just explicates to what “variation-response” question a causality quantifier answers. Its purpose does not reduce to finding a better quantifier under some conditions. Any solution to the latter problem is not generally applicable since it inevitably relies on the choice of an external *ad hoc* criterion for what is a good quantifier.

Note that both $T_{Y \rightarrow X}^{(t)}$ and $A_{Y \rightarrow X}^{(t)}$ are special cases of a quantity $I_{Y \rightarrow X}^{(t)}$ which is the TE in whose definition p_{XY}^{st} is replaced with an arbitrary PDF ρ_{XY} , i.e., the alternative initial condition is $\rho_{XY,\lambda}^* = \delta(x_0 - x_0^*)\rho_{Y|X}(y_0|x_0^*)$ and the assemblage is a weighted average with $\rho_{XY}(x_0^*, y_0^*)$. Let us call $I_{Y \rightarrow X}^{(t)}$ “extended TE” [Fig. 3(a)]. It is always nonnegative. Its advantageous feature is that its nonzero value for some ρ_{XY} and $t > 0$ is a necessary and sufficient sign of the existence of the

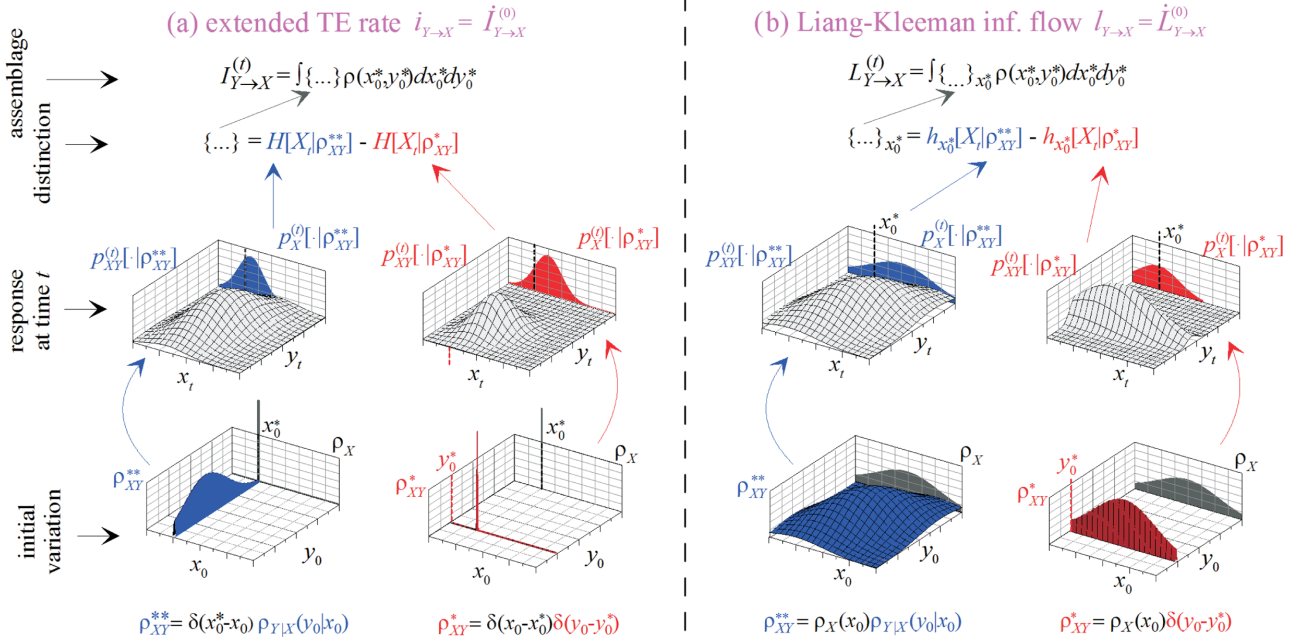


FIG. 3. Qualitative illustrations of the two DCEs which are information flows: (a) the extended TE rate, (b) the LKIF. The first (bottom) row shows the pairs of initial conditions (i.e., the initial variations) with their marginal PDFs, the second row shows the responses to those initial variations on a finite temporal horizon t , the third row shows the distinctions, and the fourth row shows the assemblages. The evolution of PDFs in all columns qualitatively corresponds to the SDS (1) with both positive coupling coefficients.

coupling $Y \rightarrow X$. The extended TE is well defined even for an SDS without any stationary PDF, while $T_{Y \rightarrow X}^{(t)}$ and $A_{Y \rightarrow X}^{(t)}$ are defined only if a stationary PDF exists. Very short-term effects are characterized by the TE rate $\tau_{Y \rightarrow X} = \frac{dT_{Y \rightarrow X}^{(t)}}{dt} \Big|_{t=0}$ and the extended TE rate $i_{Y \rightarrow X} = \frac{dI_{Y \rightarrow X}^{(t)}}{dt} \Big|_{t=0}$.

B. Liang-Kleeman information flow

As compared to the TE, the LKIF is a more theoretically motivated, formalism-driven notion. It relies on the ideas from mathematical physics and atmospheric science [5] and is well grounded on a firm mathematical (Liouville equation) and physical (hydrodynamics) basis. The LKIF has been systematically developed as a “rigorous notion *ab initio*” in Ref. [78]. Novel research often refers to this notion and conclusions as well-established knowledge (e.g., [71,72]).

In continuous time, the LKIF relies on the Liouville equation for a noise-free system [5] or the Fokker-Planck equation for a noisy SDS [75] given by stochastic differential equations widely used in physics (e.g., [88]). The latter equations in Itô’s sense read

$$\begin{aligned} dx &= f_x(x, y) dt + g_{xx}(x, y) dw_x, \\ dy &= f_y(x, y) dt + g_{yy}(x, y) dw_y, \end{aligned} \quad (6)$$

where W_x and W_y are mutually independent standard Wiener processes. Each Wiener process enters only one equation, so the noises in X and Y are mutually independent that corresponds to zero terms $g_{xy} = g_{yx} = 0$ in Ref. [75]. This is done here only for convenience, all derivations would be the same for cross-correlated noises. Further, x and y are one-dimensional as in Refs. [5,75], but arbitrary finite dimensions can be considered in the same way. The PDF

$p_{XY}^{(t)}(x_t, y_t | x_0, y_0)$ is readily obtained via solving the Fokker-Planck equation [97,98].

The basic idea of Ref. [5] is to compare the Shannon entropy rate $\dot{H}(X_t)$ at $t = 0$ for any given $\rho_{XY}(x_0, y_0)$ to that rate under an additional condition of “ y frozen” [73,75]. As the former rate, the authors take indeed the Shannon entropy rate for the initial ensemble with some PDF ρ_{XY} , i.e., the functionally conditional $\dot{H}[X_t | \rho_{XY}]$ according to the terminology suggested here. However, as the latter rate, the quantity \dot{H}_X^* is defined in Ref. [5] on the basis of *formal similarity* to the Shannon entropy rate of the whole system $\dot{H}(X_t, Y_t)$. For the deterministic case $g_{xx} = g_{yy} = 0$, the rate $\dot{H}(X_t, Y_t)$ reads

$$\dot{H}(X_t, Y_t) = \iint \left(\frac{\partial f_x}{\partial x} + \frac{\partial f_y}{\partial y} \right) \rho_{XY} dx dy. \quad (7)$$

The entropy rate of X “with y frozen” is defined [5] as

$$\dot{H}_X^* = \iint \frac{\partial f_x}{\partial x} \rho_{XY} dx dy. \quad (8)$$

The difference $\dot{H}(X_t) - \dot{H}_X^*$ is called the information flow (LKIF) which then equals [5]

$$l_{Y \rightarrow X} = \iint \left(-\frac{\partial(\rho_X f_x)}{\partial x} \right) \rho_{Y|X} dx dy. \quad (9)$$

The notation $l_{Y \rightarrow X}$ and others differ from the notations in Ref. [5] because of a separate set of notations here.

With nonzero noises, Liang’s formula [75] reads

$$l_{Y \rightarrow X} = \iint \left(-\frac{\partial(\rho_X f_x)}{\partial x} + \frac{1}{2} \frac{\partial^2(\rho_X g_{xx}^2)}{\partial x^2} \right) \rho_{Y|X} dx dy. \quad (10)$$

This more general form of the quantifier is still often called the LKIF in the literature. Note that the LKIF can be negative

as distinct from the TE. Its further peculiarity (and often a disadvantage) is that $l_{Y \rightarrow X} = 0$ for any randomized initial PDF $\rho_{XY}(x_0, y_0) = \rho_X(x_0)\rho_Y(y_0)$ [76], even if the coupling $Y \rightarrow X$ is arbitrarily strong in other respects, e.g., in terms of the TE (see Sec. III D).

The LKIF is fruitfully used in the whole line of causality research [5,67,71–78]. However, its formal reasoning [75] does not show in what kind of experiments the LKIF equals the difference of $\dot{H}(X_t)$ for any two ensembles of time realizations of the SDS (6). The meaning of $l_{Y \rightarrow X}$ has remained enigmatic in this sense, despite some efforts [73,74,78] to provide its clear interpretation beyond the formal beauty and some reasonable properties such as $l_{Y \rightarrow X} = 0$ for the absent coupling $Y \rightarrow X$.

To apply the general DCE principle of interpreting any existing causality quantifier, one must find a particular DCE which is equivalent to $l_{Y \rightarrow X}$ or, in other words, to express $l_{Y \rightarrow X}$ in the form (4). This task can be solved, first, by noting that the reasoning of Eq. (10) in Ref. [75] is based on the evolution equation for the quantity $h_{x_0^*}(X_t) = -\ln p_X^{(t)}(x_0^*)$. The latter can be called *local entropy* since the Shannon entropy of a random variable X_t with a PDF $p_X^{(t)}$ reads $H(X_t) = \int p_X^{(t)}(x_0^*) h_{x_0^*}(X_t) dx_0^*$, i.e., equals the local entropy $h_{x_0^*}$ averaged with the weight function $p_X^{(t)}(x_0^*)$. Then, after some derivations (Appendix D 2), one can show that the LKIF is the difference of *local entropy* rates $\dot{h}_{x_0^*}[X_t | \rho_{XY}^{**}] - \dot{h}_{x_0^*}[X_t | \rho_{XY}^*]$ for the reference initial condition $\rho_{XY, y_0^*}^*(x_0, y_0) = \rho_X(x_0)\delta(y_0 - y_0^*)$ and the alternative $\rho_{XY}^{**}(x_0, y_0) = \rho_X(x_0)\rho_{Y|X}(y_0|x_0)$ averaged over (x_0^*, y_0^*) with $\rho_{XY}(x_0^*, y_0^*)$. The difference of the local entropies turns out to be a specific distinction functional depending on x_0^* as a parameter. So the LKIF is the rate of the corresponding DCE $L_{Y \rightarrow X}^{(t)}$ illustrated in Fig. 3(b). This leads to the next theorem whose detailed proof is given in Appendix D 2.

Theorem 2. For the SDS (6), the LKIF (10) is the rate of the DCE (4) of the initial condition variations given by $\rho_{XY, y_0^*}^*(x_0, y_0) = \rho_X(x_0)\delta(y_0 - y_0^*)$ and $\rho_{XY}^{**}(x_0, y_0) = \rho_{XY}(x_0, y_0)$ with the distinction $\langle \cdot | \cdot \rangle_{x_0^*} = h_{x_0^*}[X_t | \rho_{XY}^{**}] - h_{x_0^*}[X_t | \rho_{XY, y_0^*}^*]$ and the assemblage $\langle \cdot \rangle_{\Delta} = \int \int (\cdot) \rho_{XY}(x_0^*, y_0^*) dx_0^* dy_0^*$.

Thus, the DCE formalism *deciphers* that the LKIF quantifies how strongly the functionally conditional *local entropy* rate $\dot{h}_{x_0^*}(X_t)$ changes on average if y_0 is freed to vary according to $\rho_{Y|X}(y_0|x_0)$ as compared to $\delta(y_0 - y_0^*)$, given $\rho_X(x_0)$. This DCE can *not* be reduced to the difference of the *Shannon entropy* rates for any two ensembles, since for each x_0^* the quantity $\dot{h}_{x_0^*}[X_t | \rho_{XY, y_0^*}^*]$ is averaged over y_0^* with a separate weighting function $\rho_{Y|X}(y_0^*|x_0^*)$ and so there is no averaging over x_0^* in the form $\int p(x_0^*) \ln p(x_0^*) dx_0^*$. This circumstance has not been revealed previously, since the authors focused on deriving formulas for the specific class of SDS (6) instead of taking the general SDS and DCE viewpoint.

It can be shown that Theorem 2 is valid for a discrete-time SDS of Refs. [73,74]: One should just replace the DCE rate with the finite-time DCE on the temporal horizon $t = 1$ and the Fokker-Planck equation with the Frobenius-Perron operator. It is straightforward to confirm the validity of the Theorem 2 for vector-valued variables X_t and Y_t of arbitrary dimensions. Moreover, if the formulation of the LKIF as a DCE is taken as

the *definition* of the LKIF, this quantifier readily applies to any SDS including discrete-state Markov processes and Markov chains where no explicit functions (f_x, g_{xx}, f_y, g_{yy}) are defined, but only transition probabilities are given. The previous works [5,75] have not studied such systems since they rely on the explicit SDS equations (6). Such broadening of the LKIF applicability confirms the conceptual and practical usefulness of the DCE viewpoint.

C. Interrelation

Let us compare the extended TE rate $i_{Y \rightarrow X}$ and the LKIF $l_{Y \rightarrow X}$ for the same arbitrary initial PDF $\rho_{XY} = \rho_X \rho_{Y|X}$. It is more convenient to compare their finite-time counterparts—the extended TE $I_{Y \rightarrow X}^{(t)}$ and the finite-time LKIF $L_{Y \rightarrow X}^{(t)}$. Note that the TE involves the localized marginal PDF $\rho_{X, x_0^*}^* = \rho_{X, x_0^*}^{**} = \delta(x_0 - x_0^*)$, while the LKIF involves the full PDF $\rho_X^* = \rho_X^{**} = \rho_X(x_0)$. Conversely, the distinction functional for the TE involves the full Shannon entropies $H(X_t)$, i.e., $h_x(X_t)$ weighted with $p_X^{(t)}(x)$ over the entire real axis, while the distinction functional for the LKIF involves only the local entropies, i.e., $h_x(X_t)$ “weighted” with $\delta(x - x_0^*)$ (Fig. 3). The conditional PDFs in the reference and alternative initial conditions are defined in the same way for both quantifiers: $\rho_{Y|X, y_0^*}^*(y_0|x_0) = \delta(y_0 - y_0^*)$ and $\rho_{Y|X}^{**}(y_0|x_0) = \rho_{Y|X}(y_0|x_0)$. The assemblage functionals are also the same average over (x_0^*, y_0^*) with $\rho_{XY}(x_0^*, y_0^*)$. So the two quantifiers in the direction $Y \rightarrow X$ differ only by their distinction functionals and by the marginal PDFs of X_0 in their initial condition variations.

To find a closer link, let us relate the two quantifiers smoothly within a family of DCEs (4). Define a DCE $Q_{Y \rightarrow X}^{(t)}$ involving the marginal PDF $\rho_X(x_0) = \tilde{\rho}_X(x_0)$ which is nonzero only in an interval $(x_0^* - \Delta_i, x_0^* + \Delta_i)$ as

$$\tilde{\rho}_X(x_0) = \frac{\rho_X(x_0)}{\int_{x_0^* - \Delta_i}^{x_0^* + \Delta_i} \rho_X(x) dx}$$

and $\tilde{\rho}_X(x_0) = 0$ outside that interval. Let the tilde imply that dependence of ρ_X on x_0^* and Δ_i , so $\tilde{\rho}_X$ is just a shorter notation for the parameter-dependent $\rho_{X, x_0^*, \Delta_i}$. The reference initial condition reads

$$\rho_{XY, x_0^*, \Delta_i, y_0^*}^* \equiv \tilde{\rho}_{XY}^* = \tilde{\rho}_X(x_0)\delta(y_0 - y_0^*),$$

and the alternative initial condition is

$$\rho_{XY, x_0^*, \Delta_i}^{**} \equiv \tilde{\rho}_{XY}^{**} = \tilde{\rho}_X(x_0)\rho_{Y|X}(y_0|x_0).$$

Define the distinction functional as the difference of the “window” entropies $\tilde{H}(X_t)$ where the weighting function is nonzero only within an interval $(x_0^* - \Delta_d, x_0^* + \Delta_d)$, i.e., for a random variable U with a PDF p_U the window entropy in an interval $(x_0^* - \Delta_d, x_0^* + \Delta_d)$ reads

$$\tilde{H}(U) = \frac{-\int_{x_0^* - \Delta_d}^{x_0^* + \Delta_d} p_U(u) \ln p_U(u) du}{\int_{x_0^* - \Delta_d}^{x_0^* + \Delta_d} p_U(u) du}.$$

The tilde over H implies the dependence on parameters x_0^* and Δ_d , so \tilde{H} is a shorter notation for $H_{x_0^*, \Delta_d}$. So the distinction

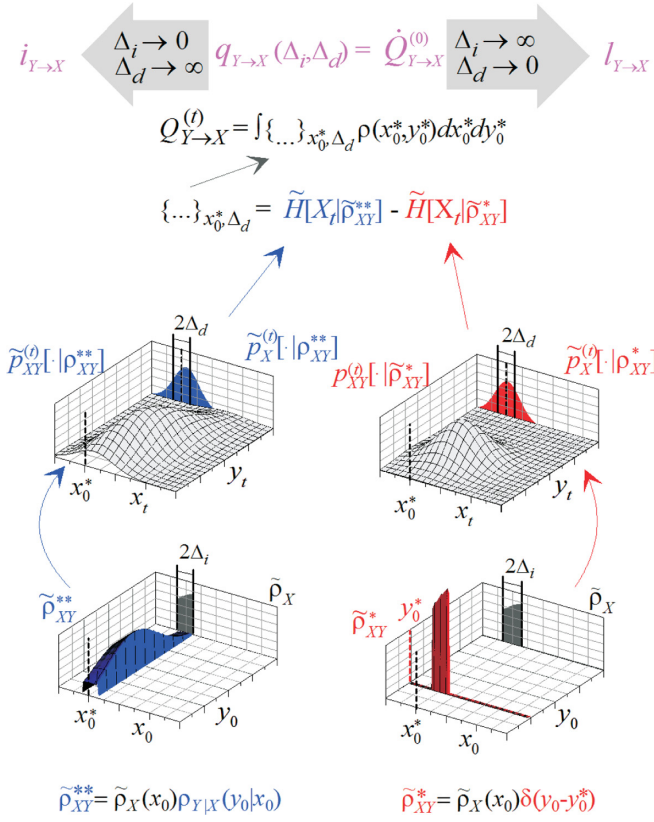


FIG. 4. Qualitative illustrations to Theorem 3, i.e., to the parameterized DCE linking the TE and the LKIF (see Fig. 3). Meanings of the rows are the same as in Fig. 3.

functional reads

$$\{[X_t | \tilde{\rho}_{XY}^*] || [X_t | \tilde{\rho}_{XY}^{**}]\}_{x_0^*, \Delta_d} = \tilde{H}[X_t | \tilde{\rho}_{XY}^{**}] - \tilde{H}[X_t | \tilde{\rho}_{XY}^*].$$

As the assemblage, define the average over y_0^* (which is a parameter of the reference initial condition only) and x_0^* (which is a parameter of the distinction and both initial conditions) with $\rho_{XY}(x_0^*, y_0^*)$. Thereby, one arrives at the finite-time DCE $Q_{Y \rightarrow X}^{(t)}(\Delta_i, \Delta_d)$ (see Fig. 4) given by

$$Q_{Y \rightarrow X}^{(t)}(\Delta_i, \Delta_d) = \int_{-\infty}^{\infty} (\tilde{H}[X_t | \tilde{\rho}_{XY}^{**}] - \tilde{H}[X_t | \tilde{\rho}_{XY}^*]) \rho_{XY}(x_0^*, y_0^*) dx_0^* dy_0^*. \quad (11)$$

Define its rate $q_{Y \rightarrow X}(\Delta_i, \Delta_d)$ as

$$q_{Y \rightarrow X}(\Delta_i, \Delta_d) = \left. \frac{dQ_{Y \rightarrow X}^{(t)}(\Delta_i, \Delta_d)}{dt} \right|_{t=0}. \quad (12)$$

For $\Delta_i \rightarrow 0$, one gets

$$\begin{aligned} \tilde{\rho}_X &\rightarrow \delta(x_0 - x_0^*), \\ \tilde{\rho}_{XY}^* &\rightarrow \delta(x_0 - x_0^*) \delta(y_0 - y_0^*), \\ \tilde{\rho}_{XY}^{**} &\rightarrow \delta(x_0 - x_0^*) \rho_{Y|X}(y_0|x_0). \end{aligned}$$

For $\Delta_d \rightarrow \infty$, the distinction functional becomes

$$\{U^* || U^{**}\}_{x_0^*, \Delta_d} \rightarrow H(U^{**}) - H(U^*),$$

i.e., the difference of the Shannon entropies, while x_0^* is a parameter of the initial conditions, not of the distinction. So

one gets the extended TE rate $i_{Y \rightarrow X} = q_{Y \rightarrow X}(0, \infty)$ which equals $\tau_{Y \rightarrow X} = q_{Y \rightarrow X}(0, \infty)$ if $\rho_{XY} = \rho_{XY}^{st}$.

For $\Delta_i \rightarrow \infty$ one gets

$$\begin{aligned} \tilde{\rho}_X &\rightarrow \rho_X(x_0), \\ \tilde{\rho}_{XY}^* &\rightarrow \rho_X(x_0) \delta(y_0 - y_0^*), \\ \tilde{\rho}_{XY}^{**} &\rightarrow \rho_X(x_0) \rho_{Y|X}(y_0|x_0). \end{aligned}$$

For $\Delta_d \rightarrow 0$, the distinction functional becomes

$$\{U^* || U^{**}\}_{x_0^*, \Delta_d} \rightarrow h_{x_0^*}(U^{**}) - h_{x_0^*}(U^*),$$

i.e., the difference of the local entropies, while x_0^* is a parameter of the distinction, but not of the initial conditions. So one gets the LKIF $l_{Y \rightarrow X} = q_{Y \rightarrow X}(\infty, 0)$. These considerations prove the following theorem.

Theorem 3. For an SDS \mathcal{S} , a two-parameter family of DCE rates $q_{Y \rightarrow X}(\Delta_i, \Delta_d)$ defined by Eqs. (11) and (12) contains the extended TE rate and the LKIF as its mutually opposite limit cases: $q_{Y \rightarrow X}(0, \infty) = i_{Y \rightarrow X}$ and $q_{Y \rightarrow X}(\infty, 0) = l_{Y \rightarrow X}$.

So the TE rate and the LKIF in the direction $Y \rightarrow X$ differ not as “a data-driven” and “a rigorous *ab initio*” quantifier, which circumstance depends on a researcher’s view. Their objective difference is the opposite character of their distinction functionals and marginal PDFs of X_0 in their initial variations under the same assemblage and the same conditional PDFs of Y_0 in their respective initial conditions. Theorem 3 applies to X_t and Y_t of arbitrary dimensions and to a discrete-time SDS.

This seemingly *qualitative* difference of the two quantifiers as the *opposite* limit cases does not always lead to a strong *quantitative* difference, since the entire family $q_{Y \rightarrow X}(\Delta_i, \Delta_d)$ may exhibit the same or very close values for some systems or at some values of the parameter a . For other systems, the values of the two quantifiers may differ arbitrarily strongly. A numerical example below serves to explicate and illustrate those typical cases.

D. Numerical example

Consider again the SDS (1) which is just the SDS (6) with constant noise intensities $g_{xx}^2 = \Gamma_{xx}$, $g_{yy}^2 = \Gamma_{yy}$, and linear drift terms $f_x = -a_x x$ and $f_y = -a_y y$. A nonzero coupling coefficient a_{xy} provides the stationary variance different from the “free” variance $\sigma_x^2 \neq \sigma_{x,0}^2$. The respective relative change of variance under “switching the coupling $Y \rightarrow X$ on” is an *asymptotic* DCE $S_{Y \rightarrow X} = (\sigma_x^2 - \sigma_{x,0}^2) / \sigma_{x,0}^2$ [79–81]. It is often of interest in practice (e.g., [94]) and used here for a fuller understanding of the TE-LKIF relation. This DCE involves the coupling parameter variation $(a_{xy}^*, a_{xy}^{**}) = (0, a_{xy})$, the infinite horizon $t \rightarrow \infty$, the distinction as the relative difference of variances $\{U^* || U^{**}\} = (\sigma_{U^{**}}^2 - \sigma_{U^*}^2) / \sigma_{U^*}^2$, and the trivial assemblage. The initial conditions ρ_{XY}^* and ρ_{XY}^{**} are arbitrary since the stationary PDF reached at $t \rightarrow \infty$ is unique. By solving algebraic equations for the second statistical moments (see [79,81] and Appendix E), one gets

$$S_{Y \rightarrow X} = \frac{\beta_{xy}^2 + m_{xy} \beta_{xy} \beta_{yx}}{(1 + m_{xy})(1 - \beta_{xy} \beta_{yx})}. \quad (13)$$

Both the TE rate and the LKIF for the initial PDF $\rho_{XY}(x_0, y_0) = \rho_{XY}^{st}(x_0, y_0)$ can be derived analytically from the DCE definition (4) after finding explicitly the stationary

PDF p_{XY}^{st} and the transition PDFs which are all Gaussian (see [29] and Appendix E).

Taking into account that the LKIF for the SDS (1) is given by Eq. (3), some further algebra within the DCE framework (see Appendix E) leads to a surprisingly simple exact relationship

$$l_{Y \rightarrow X} t_x = \frac{S_{Y \rightarrow X}}{1 + S_{Y \rightarrow X}} = \frac{\sigma_X^2 - \sigma_{X,0}^2}{\sigma_X^2}, \quad (14)$$

which holds true for any parameter values which provide stationarity of the process (1), i.e., for $\beta_{xy}\beta_{yx} < 1$. So the LKIF $l_{Y \rightarrow X} t_x$ measured in “nats per recipient relaxation time” simply relates to the asymptotic DCE $S_{Y \rightarrow X}$ and equals the “relative contribution” of the coupling $Y \rightarrow X$ to the recipient variance σ_X^2 . This relationship has not yet been known since the LKIF was not considered from the DCE viewpoint. It deserves to be fixed as a final theorem.

Theorem 4. For the SDS (1) with a stationary PDF p_{XY}^{st} , the LKIF (10) defined with the PDF $\rho_{XY} = p_{XY}^{\text{st}}$ and multiplied by the recipient relaxation time t_x relates to the asymptotic DCE on variance $S_{Y \rightarrow X}$ through Eq. (14) and so equals the relative contribution of the coupling $Y \rightarrow X$ to the recipient variance, i.e., $l_{Y \rightarrow X} t_x = \sigma_X^2 / \sigma_{X,0}^2 - 1$.

The TE rate reads (see [81] and Appendix E)

$$\tau_{Y \rightarrow X} = a_x \beta_{xy}^2 (1 - r_{st}^2) (1 + S_{X \rightarrow Y}) / 4, \quad (15)$$

where r_{st} is the stationary zero-lag cross-correlation $r_{st} = \sigma_{XY} / (\sigma_X \sigma_Y)$. Under the conditions of $|\beta_{xy} / \beta_{yx}| \gg m_{xy}$ (called the relatively predominant coupling $Y \rightarrow X$ [81]) and weak couplings $\beta_{xy}^2 \ll 1$, $\beta_{yx}^2 \ll 1$, it holds true that

$$\tau_{Y \rightarrow X} t_{\min} \approx S_{Y \rightarrow X} / 4, \quad (16)$$

where $t_{\min} = \min\{t_x, t_y\}$ is the minimum of the two relaxation times of the system (1). So $\tau_{Y \rightarrow X} t_{\min}$ is the TE rate measured in “nats per minimal relaxation time” which simply relates to the long-term DCE. Equations (14) and (16) show that “nats per time unit” for $\tau_{Y \rightarrow X}$ generally differ from those for $l_{Y \rightarrow X}$. Under the above conditions for Eq. (16), the two quantifiers are related as

$$\tau_{Y \rightarrow X} / l_{Y \rightarrow X} \approx (1 + a_y / a_x) / 4. \quad (17)$$

In particular, those conditions are met for a unidirectional coupling $Y \rightarrow X$ which is weak enough ($\beta_{xy}^2 \ll 1$). Let us consider this case and note that numerical values (i.e., “nats per time unit”) of the two quantifiers may be either strongly different as $\tau_{Y \rightarrow X} \gg l_{Y \rightarrow X}$ if the coupling source is much faster ($a_y \gg a_x$; see Fig. 5 at $m_{xy} \gg 1$), or quite similar as $l_{Y \rightarrow X} = 4\tau_{Y \rightarrow X}$ if the coupling source is much slower ($a_y \ll a_x$; see Fig. 5 at $m_{xy} \ll 1$). The difference between the TE rate and the LKIF increases with $S_{Y \rightarrow X}$ [cf. Figs. 5(a) and 5(b)]: $l_{Y \rightarrow X}$ saturates under the increase of $S_{Y \rightarrow X}$ while $\tau_{Y \rightarrow X}$ increases unboundedly.

Why is $l_{Y \rightarrow X} \ll \tau_{Y \rightarrow X}$ for a unidirectional coupling from the fast source, but not for that from the slow source? In the former case, the cross-correlation coefficient is small $r_{st} \leq 1 / \sqrt{m_{xy}} \ll 1$ for an arbitrarily strong coupling [81], while in the latter case r_{st} gets close to unity for a strong coupling. Further, r_{st} enters linearly the expression for the LKIF (3), not the expression for the TE rate (15). To see the role of the small

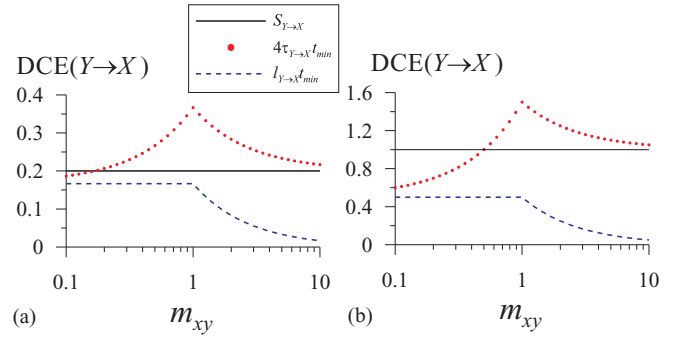


FIG. 5. Three DCEs for the SDS (1) with unidirectional coupling ($a_{yx} = 0$), β_{xy}^2 is computed from Eq. (13) for each m_{xy} to provide a given $S_{Y \rightarrow X}$ and the respective $\tau_{Y \rightarrow X}$ and $l_{Y \rightarrow X}$: (a) $S_{Y \rightarrow X} = 0.2$; (b) $S_{Y \rightarrow X} = 1$. Solid lines show $S_{Y \rightarrow X}$, dotted lines $4\tau_{Y \rightarrow X} t_{\min}$, and dashed lines $l_{Y \rightarrow X} t_{\min}$.

cross-correlation more clearly, consider the extended TE rate $i_{Y \rightarrow X}$ and the LKIF for a randomized initial PDF $\rho_{XY} = \rho_X \rho_Y$, where the cross-correlation r of X_0 and Y_0 is necessarily $r = 0$. For definiteness, take $\rho_{XY} = p_{XY}^{\text{st}}$. Then $i_{Y \rightarrow X}$ equals the Ay-Polani information flow and reads $i_{Y \rightarrow X} = \tau_{Y \rightarrow X} / (1 - r_{st}^2)$, i.e., $i_{Y \rightarrow X} \geq \tau_{Y \rightarrow X}$. For a unidirectional coupling $Y \rightarrow X$, it takes an especially simple form $i_{Y \rightarrow X} = a_x \beta_{xy}^2 / 4$ and gets arbitrarily large for $\beta_{xy}^2 \gg 1$. In contrast, $l_{Y \rightarrow X}$ is exactly zero for any randomized ρ_{XY} , independently of the coupling parameter β_{xy}^2 , of the extended TE rate, and of the asymptotic DCE $S_{Y \rightarrow X}$ which may all be arbitrarily large.

The stationary PDF is itself a randomized PDF for $\beta_{xy} / \beta_{yx} = -m_{xy}$ and it is very close to a randomized PDF in a vicinity of such parameter values. So a robust situation is that the TE rate is arbitrarily large, while the LKIF is simultaneously arbitrarily small. Such a drastic difference between the two quantifiers is met in the case of negative feedback; see negative a_{xy} in Figs. 6 and 7. Then $S_{Y \rightarrow X} > 0$ if $|\beta_{xy} / \beta_{yx}| > m_{xy}$ (i.e., the coupling $Y \rightarrow X$ is relatively predominant [81]) and simultaneously $S_{X \rightarrow Y} < 0$ since $|\beta_{yx} / \beta_{xy}| < m_{yx}$ (the coupling $X \rightarrow Y$ is relatively deficient [81]). For the boundary situation of relatively equivalent couplings $|\beta_{xy} / \beta_{yx}| = m_{xy}$ (see the red arrows in Fig. 6 and the value of $a_{xy} = -10$

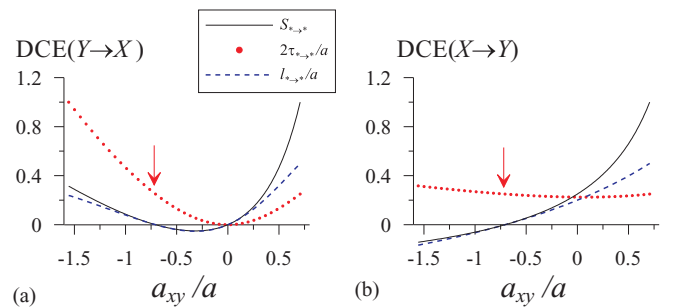


FIG. 6. Three DCEs for the SDS (1) with bidirectional coupling for $a_x = a_y = a$ and $a_{yx} = a / \sqrt{2}$ with red arrow indicating the “point of drastic difference” between $\tau_{* \rightarrow *}$ and $l_{* \rightarrow *}$: (a) for the direction $Y \rightarrow X$; (b) for the direction $X \rightarrow Y$. The relaxation times are $t_x = t_y = t_{\min} = 1/a$. Solid lines show $S_{* \rightarrow *}$, dotted lines $2\tau_{* \rightarrow *} / a$, and dashed lines $l_{* \rightarrow *} / a$.

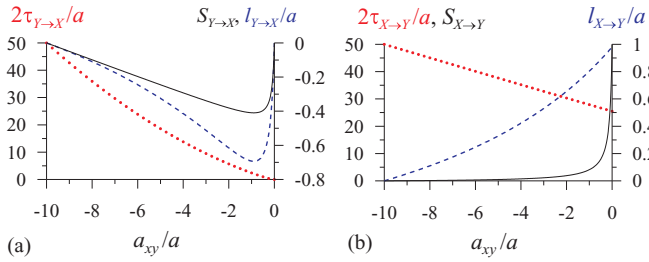


FIG. 7. Three DCEs for the SDS (1) with bidirectional coupling for $a_x = a_y = a$ and $a_{yx} = 10a$: (a) for the direction $Y \rightarrow X$; (b) for the direction $X \rightarrow Y$. The relaxation times are $t_x = t_y = t_{\min} = 1/a$. Solid lines show $S_{* \rightarrow *}$, dotted lines $2\tau_{* \rightarrow *}/a$, and dashed lines $l_{* \rightarrow *}/a$.

in Fig. 7), the stationary PDF is randomized and so $r_{st} = 0$ and $S_{Y \rightarrow X} = S_{X \rightarrow Y} = 0$, even though the coupling parameters β_{xy}^2 and β_{yx}^2 may be arbitrarily large. Speaking more physically, even large couplings of this type do not change the stationary variances of X_t and Y_t as compared to the uncoupled processes, i.e., such couplings do not change the integral powers of the signals x_t and y_t since the variance equals the integral of the power spectral density. However, they change the power spectral densities, e.g., a strong peak arises at the frequency $\omega \approx \sqrt{|a_{xy}a_{yx}|}$ for $|a_{xy}a_{yx}| \gg a_x a_y$ [Fig. 8(b), solid line] instead of the “free” red noise spectra [Fig. 8(b), dashed line] [49]. One can also easily see an oscillatory character of the coupled processes as compared to a more irregular character of the uncoupled ones from their time realizations in Figs. 1(a) and 1(b). In this spectral sense, an “information flow” due to the nonzero coupling $Y \rightarrow X$ is great. The corresponding arbitrarily large TE rate $\tau_{Y \rightarrow X} = a_x \beta_{xy}^2/4$ adequately reflects this circumstance. In contrast, the LKIF $l_{Y \rightarrow X} = 0$ is not sensitive to this coupling which is often a disadvantage. So the LKIF is an information flow in a very restricted sense.

Figure 7 presents the three DCEs versus negative a_{xy} for a stronger coupling $a_{yx} = 10a$ to be compared to Fig. 6 for $a_{yx} = a/\sqrt{2}$. The value $a_{xy}/a = -10$ in Fig. 7(1) corresponds to a randomized stationary PDF. This strong coupling is well reflected by the relative TE rate whose value $\tau_{Y \rightarrow X}/a = 25$

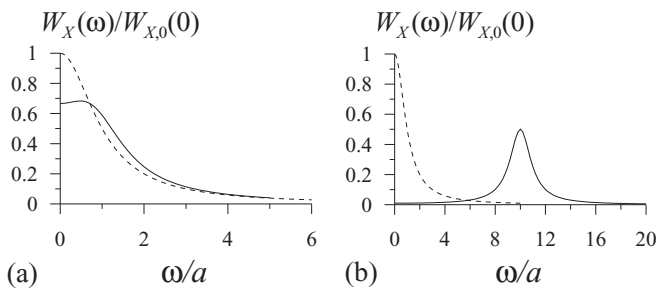


FIG. 8. Power spectral densities $W_X(\omega)$ of X for the SDS (1) in the uncoupled regime $a_{yx} = 0$ (dashed lines) and for a bidirectional coupling $a_{yx} = -a_{xy}$ (solid lines) which maintains the same integral power of X as for the uncoupled regime: (a) $a_{yx} = -a_{xy} = a/\sqrt{2}$; (b) $a_{yx} = -a_{xy} = 10a$. The densities are divided by the zero-frequency value of the power spectral density of the uncoupled process $W_{X,0}(0) = \Gamma_{xx}/(2\pi a^2)$.

nats should be regarded as very large, since $\tau_{X \rightarrow Y}/a \approx 12$ nats for the unidirectional coupling $X \rightarrow Y$ [$a_{xy}/a = 0$ in Fig. 7(b)] corresponds to a huge $S_{X \rightarrow Y} = 50$, i.e., the variance of Y is multiplied by 51 due to switching the coupling on. The respective large short-term future responses to different initial condition variations are shown in Fig. 2: for single initial states [Fig. 2(a)], for the initial variation used in the TE [Fig. 2(b)], and for the initial variation used in the LKIF [Fig. 2(c)]. The LKIF is zero for $a_{xy}/a = -10$. However, this seemingly strange zero “causal strength” of the strong coupling can be further understood as an advantageous feature of the LKIF from the DCE viewpoint. Namely, since the variances σ_X^2 and σ_Y^2 do not differ here from their free values $\sigma_{X,0}^2$ and $\sigma_{Y,0}^2$, the Shannon entropies of the marginal stationary PDFs of X and Y do not differ from their free values too. So $l_{Y \rightarrow X}$ and $l_{X \rightarrow Y}$ correctly claim that the Shannon entropy of the stationary PDF does not change, i.e., there is no “information flow” in that sense. Note from Eq. (14) that the sign of the LKIF is the same as the sign of $S_{Y \rightarrow X}$, and so it shows whether the variance σ_X^2 increases or decreases due to switching the coupling $Y \rightarrow X$ on (see also Figs. 6 and 7). This characterization is provided by the LKIF, but not by the TE which is nonnegative by definition.

For a further discussion, note that Ref. [76] considers the value of $l_{Y \rightarrow X} = 0$ for mutually independent X_0 and Y_0 as a deep fact concerning causality and claims that “causation implies correlation ... resolving the continuing debate” ([76], p. 9) or “obviously, two uncorrelated events ($r = 0$) must be noncausal” ([76], p. 3). This conclusion is often repeated (e.g., [71]). However, a nonzero correlation of the simultaneous states of X and Y is not at all necessary for the existence of a causal coupling as the above example clearly shows. Rather, this feature of the LKIF is often its *weakness*, since the LKIF turns out to be *insensitive* even to strong couplings of a certain type (or with certain initial variations), while the TE readily reveals them. On the other hand, $l_{Y \rightarrow X}$ can show that a coupling does not change the variance σ_X^2 , and a coupling effect is absent in this sense, which is not reflected by the TE. *If* one is interested only in the integral power and its changes, the TE appears to be *inappropriate* despite its other advantages, while $l_{Y \rightarrow X}$ is then quite relevant. So each of the two quantifiers gives a meaningful description of causal effects, and none of them is necessarily suitable in any situation. Since a coupling can manifest itself in diverse properties of dynamics, multiple DCEs (4) are needed to capture this diversity. Which quantifier is most appropriate in practice depends on the question of interest.

The DCE framework should help a researcher to consider any causality quantifier in a critical way, formulate concrete questions about its meaning, explicate its properties, and either purposefully apply it to the problem at hand or replace it with another, more appropriate quantifier. Even the above comparison of the TE and the LKIF may help a researcher to decide whether to prefer a quantifier based on the Shannon entropies and sensitive to any coupling (TE) or a quantifier based on local entropies, possibly zero for some kind of strong couplings, and reflecting the sign of the asymptotic effect on variance (LKIF). This is already a more informed decision than an abstract consideration of the TE as a “data-driven” quantifier and the LKIF as a “rigorous *ab initio*” one. A

further study can provide more concrete recommendations on their advantages for some classes of systems. There can be no universal “algorithm” to select a causality quantifier suitable for any practical problem and question as it is clear from the whole realm (4) of quantifiers capturing different causal effects. A researcher should explicate which particular question about a system under study a causality quantifier must answer. It inevitably depends on the problem at hand, and so a critical attitude to any preselected quantifier can be useful. In this essential sense, the DCE viewpoint should help one to navigate among numerous quantifiers and obtain their dynamical and physical interpretations, such as the LKIF and the TE as indicators of an integral power change versus any power spectral density change in the above example of the overdamped oscillators.

IV. PERSPECTIVES

The previous classification of DCEs [29] was based on the division of the initial variations into the initial state variations and the parameter variations and the division of the temporal horizons into the short and long (infinite) ones. The DCE formalism suggested here is more flexible. It allows us to arrange causality quantifiers into a complex structure resembling a “tree” whose root is the general DCE (4) and “branches” represent different degrees of its concretization.

It is worth noting a useful kind of DCEs which involve a *distributed* temporal horizon, i.e., where t is a vector (t_1, \dots, t_N) and $X_t = (X_{t_1}, \dots, X_{t_N})$. If all $t_i \rightarrow \infty$ and $N \rightarrow \infty$ with diminishing $t_{i+1} - t_i$, the horizon becomes effectively continuous, so one has a long time realization as the future X_t . If the distinction functional compares power spectral densities, one gets spectral DCEs as introduced in Ref. [49] to continue the discussion of Refs. [45–48]. In particular, the widely used Granger-Geweke spectrum [61,99] appears to be a DCE $Y \rightarrow X$ of a parameter variation (switching the noise Γ_{yy} on) [49] with the trivial assemblage and the distinction equal to the relative difference of power spectral densities. However, the Granger-Geweke spectrum is no longer such a DCE in the case of more than two interacting subsystems [49,100]. The spectral DCEs and about two dozen other causality quantifiers are presented in a single table in the Supplemental Material [90] in order to illustrate how Eq. (4) “generates” various quantifiers and to confirm the generality of the DCE concept.

If a quantifier cannot be exactly reduced to any DCE, it may be numerically close to a DCE under some conditions. The famous Wiener-Granger causality [51,52] defined via mean-squared prediction improvement based on the infinite past is not exactly a DCE, but it is often close to a DCE [79] of initial condition variations on a finite temporal horizon with the distinction equal to the relative difference of variances and the assemblage equal to the average with a stationary PDF. Conditioning on the infinite past in the Wiener-Granger causality serves as a substitute for the conditioning on the (often unobserved) initial states in a DCE. Such approaches as convergent cross-mapping [20] and similar techniques [16,33] efficiently *detect* unidirectional couplings between deterministic dynamical systems and are often used in practice (e.g., [56]). However, they are not formulated as

responses to any variations (interventions) and so are not easy to interpret as any “causality strength.” Relating them to some DCEs seems to be of interest, in particular, to understand better their dynamical (and sometimes even physical) causal meaning beyond being a sign of the coupling existence.

More than two subsystems are often present and must be taken into account in a practical situation. Then the other processes interacting with X and Y may be represented as a third subsystem Z . The DCE formalism readily extends to such a situation via including initial states and parameters of Z into the generalized initial condition $\theta = \{\rho(x, y, z), a\}$ where the parameter vector includes more components to describe three individual dynamics and different pairwise couplings as $a = (a_x, a_y, a_z, a_{xy}, a_{yx}, a_{xz}, a_{zx}, a_{yz}, a_{zy})$. All definitions are exactly the same as in the case of two subsystems with the states of Z entering the set of conditioning variables; see Appendix F 1 for a concrete example.

Concerning an inverse problem, a DCE can be estimated from a time series directly (in a nonparametric way) if it is a DCE of initial condition variations and full state vectors of an SDS are observed. However, the problems of confounders and unvisited states generally become relevant. Then one should either perform appropriate interventions (e.g., [6,19,57]) or identify a parameterized model SDS (e.g., [101,102]) and apply Eq. (4) to the obtained model (see Appendix F 2).

It should be noted that estimation of any causality quantifier from a passively observed time series is not *per se* a tool for *causal discovery*. Such a discovery can be performed only in combination with another assumption. Namely, causal relations are *assumed in an underlying SDS*, and the DCE viewpoint *makes this assumption explicit*, exactly as structural causal models in Ref. [42] do. Such an SDS may be specified in a less detailed way, e.g., the assumption that certain vectors x and y constitute a full state of the entire SDS [i.e., there are no confounders of (x_0, y_0) and (x_t, y_t) at $t > 0$] also specifies that SDS. For a causal analysis, it is always desirable to formulate such assumptions explicitly, in agreement with Pearl’s effort and advice “to explicate slippery concepts” [42]. Still, one often prefers seemingly more universal model-free characteristics of interdependence between observed signals x_t and y_t , e.g., the convergent cross-mapping, or the Wiener-Granger causality, or the Granger-Geweke spectra, or the infinite-history TE. However, to interpret any such quantity unambiguously as a *causality quantifier*, one should anyway imply some variations and responses to agree with well-developed causality formalism of Refs. [42,43]. The SDS is just a universal tool to express such relations for processes: its Markovianity provides a “closure” of the problem (to define DCEs nonambiguously despite the noise influence) and can be relaxed under further assumptions (Appendix F 2). So the SDS and DCE concepts just represent closely the “flow of causation from past into future” [103]. Therefore, I expect that the DCE concept will be suitable to formulate a variety of causality quantifiers in the same manner and arrange them into a single picture. As arguments in favor of such a role, note the relation between the TE and the LKIF revealed here, relations between the two TE versions and several long-term DCEs in Ref. [81], relations between the phase-dynamic causalities in Ref. [82], interpretations of the spectral causalities in Ref. [49] contributing to their previous discussion [45,47], and many

quantifiers expressed as DCEs in Tables I and II of the Supplemental Material [90]. Explicit usage of the DCE formalism can make a coupling analysis in practice more reliable, e.g., via including additional quantifiers to check important conclusions about most influential causal couplings such as those for nuclear reactor systems [54].

V. CONCLUSIONS

Dozens of causality quantifiers are used to study irregular processes of different origin. Some of them were suggested a long time ago (from the 1950s to the 1990s) in the fields of applied mathematics and econometrics, e.g., the Wiener-Granger causality [51,52] and the Granger-Geweke spectrum [99]. Others have arisen in the last two decades (the TE [1], the LKIF [5], etc.) and new ones are still being actively suggested (e.g., [38,39]) in the fields of nonlinear dynamics and information theory. Some of them are similar to each other and differ by their normalization like various spectral causalities (e.g., [60]) or by conditioning variables like different versions of the TE (e.g., [22,23,81]). Others are obtained from apparently completely different ideas, e.g., the phase-dynamics modeling [3,4,104–108] and the convergent cross-mapping [20]. To navigate in this variety of quantifiers and perform a purposeful choice, a researcher needs a unifying viewpoint and formalism which would allow one to arrange numerous quantifiers into a single picture and formulate exactly their common features and essential differences. The formalism of dynamical causal effects developed here on the basis of the previous studies [29,49,79–81] is a promising tool to address those issues. The main results of this work can be summarized as follows.

It is shown that many causality quantifiers for processes can be derived as realizations of the general DCE (4), i.e., the latter can serve as a first principle. The DCE formalism provides a flexible language to describe such quantifiers in a precise and unified manner using the notions of initial variation (an ordered pair of the reference and alternative initial conditions), single-point or distributed temporal horizon, distinction functional, and assemblage functional. The initial variation is generalized here to include the functional initial conditions which are the distributions of the initial states rather than the single initial states. Due to this generalization, the transfer entropy is shown to be *exactly* a DCE with specific functional initial conditions, rather than an *approximation* to a DCE with single-state variations [29,81]. The Liang-Kleeman information flow is also shown to be exactly a DCE. Moreover, it is shown that the LKIF compares local entropies of the ensembles with specific functional initial conditions, rather than the Shannon entropies of any ensembles as it has been always thought previously. The two information flows are related here as opposite limit cases in a specific DCE family, their “nats per time unit” are shown to differ from each other essentially and to describe different manifestations of a directional coupling in dynamics. Thus, for a two-dimensional linear stochastic system, the TE is sensitive to any nonzero coupling, including such a coupling which changes power spectral density of the coupling recipient without changing its integral power, while the LKIF is not sensitive to the latter coupling.

Being defined precisely through “intervention-effect” experiments with a stochastic dynamical system, the DCE concept provides an unambiguous dynamical interpretation of a causality quantifier. From a physical viewpoint, this is an advantageous feature as compared to a widespread arbitrary understanding of any formal causality quantifier in use as a quantity whose nonzero value characterizes a causal coupling in the way most interesting to a researcher in a concrete field. For example, one might be ready to think that a greater spectral causality represents a coupling which is most responsible for an anomaly in nuclear reactor systems [54]. However, such conclusions are not supported by solid arguments and a quantifier in use may not describe the quantitative aspect of interest. In contrast, the well-defined dynamical causal meaning of the DCE gives us an opportunity to reach its further physical interpretation for a concrete physical system through relating state variables and parameters of a model SDS to physical quantities. In studies of temporally evolving systems, the dynamical interpretation of a causality quantifier provided by the DCE viewpoint seems to be a necessary step to revealing its physical meaning and avoiding arbitrary interpretations.

This study seems to provide enough arguments to expect that, under further development, the DCE formalism can readily become the core of a rich and general *theory of causality quantifiers for processes*. Then apart from serving for the theoretical underpinning of different causality quantifiers, it should be also helpful for their practical purposeful choice for concrete problems.

ACKNOWLEDGMENTS

I am grateful to six anonymous referees for useful comments and suggestions. The work is supported by the Russian Science Foundation (Grant No. 19-12-00201).

APPENDIX A: INTERDISCIPLINARY CONTEXT

On the one hand, the present work shows how the *theory of oscillations* can systematically look at the entire field of *causality quantification for processes*. The theory of oscillations is understood here as a general *physical theory* (e.g., [84,109–111]) whose mathematical form is given by the *dynamical systems theory* (e.g., [83,85,86]). On the other hand, this work shows how *structural causal modeling* (e.g., [42]) can be fully realized to *quantify* causal couplings for processes within dynamical systems framework. This Appendix presents the author’s view on such an interface of disciplines and an interdisciplinary character of the present work.

1. Statistics and causality

Detection and quantification of relations between observed variables have long been important tasks in mathematical statistics with correlation and partial correlation, regression and multiple regression as the most popular tools (e.g., [112,113]). Yet researchers often tried to achieve a more ambitious goal of inferring *causal couplings* on the basis of statistical analysis. Numerous attempts to interpret correlation-like (associative) quantities as measures of

causal couplings (the famous question of “correlation versus causation”) led to frequent controversies and debates as summarized, in particular, in Pearl’s monograph [42] used here as one of the bases.

Results of those efforts were such fruitful approaches as *structural equation modeling* (SEM) [114,115] and *potential-outcome model* [116,117], the former being much more widely known and “adopted by economists and social scientists” [42] (p. 134). SEM also used regression equations ($u = \beta v + \varepsilon$ as a simple one-dimensional example), but distinguished between causal variables (v) and effects (u) via placing them on the corresponding side in each equation. The right-hand side was often used for causes and the left-hand side for effects. Then such an equation represented an “autonomous causal mechanism.” The name *structural equation* was coined because positions of the variables on one or another side reflected the causal structure of the entire system under study.

As stressed in Ref. [42], such a causal meaning was not fixed in any formal notation or an explicit term. So, in the course of subsequent applications, many practitioners tended to consider a structural equation as a usual regression equation forgetting that a causal meaning is assumed, not inferred from passive observations. “Econometric textbooks invariably devote most of their analysis to estimating structural parameters, but they rarely discuss the role of these parameters in policy evaluation” [42] (p. 136). In other words, many authors focused on “*how to estimate*” and made “*what to estimate*” implicit and sometimes misunderstood. As a result, typical statements of leading scientists in that field in 1980s and 1990s became “It would be very healthy if more researchers abandoned thinking of and using terms such as cause and effect” or “The only meaning I have ever determined for such an equation is that it is a shorthand way of describing the conditional distribution of y given x ” as cited in Ref. [42] (p. 137). Many SEM practitioners forgot that “assumptions needed for drawing causal conclusions from parameters are communicated to us by the scientist who declared the equation *structural*; they are already encoded in the *syntax* of the equation” [42] (p. 137). Pearl summarizes why the causal content of SEM escaped from the consciousness of practitioners: “1. SEM practitioners have sought to gain respectability for SEM by keeping causal assumptions implicit, since statisticians, the arbiters of respectability, abhor assumptions that are not directly testable. 2. The algebraic language that has dominated SEM lacks the notational facility needed to make causal assumptions, as distinct from statistical assumptions, explicit. By failing to equip causal relations with precise mathematical notation, the founding fathers in fact committed the causal foundations of SEM to oblivion” [42] (p. 138). He stresses: “Ironically, we are witnessing one of the most bizarre circles in the history of science: causality in search of language and, simultaneously, the language of causality in search of its meaning” [42] (p. 135).

2. Structural causal modeling

As an important clarifying and correcting step, Pearl has introduced explicitly the concept of *interventions and effects* and the respective formalism (do-calculus) [42] which

distinguishes between passively observed (i.e., ordinary) conditional distribution (PDF) and interventional conditional PDF. The latter is a PDF of a variable U under the condition that a variable V is *actively imposed* to take on a value v . It is denoted $p(u|\text{do}(v))$. This formalism removes uncertainties underlying many previous controversies and errors and gives rise to the modern development of the field of structural causal modeling (SCM). This name is taken from the previous SEM with the clarification that the equality sign is not an algebraic equality, but works more like “an assignment symbol in programming languages” [42] (p. 138). The result is that “causality has undergone a major transformation: from a concept shrouded in mystery into a mathematical object with well-defined semantics and well-founded logic. Paradoxes and controversies have been resolved, slippery concepts have been explicated, and practical problems relying on causal information that long were regarded as either metaphysical or unmanageable can now be solved using elementary mathematics. Put simply, causality has been mathematized” [42] (p. xiii). SCM is fruitfully used in sociology, epidemiology, and other fields (e.g., a monograph [118] relies on [42] in a detailed study of the mediation phenomenon and some others), but remains almost unknown to physicists.

SCM is most interested in *causal discovery*, i.e., *detection* of direct and indirect causal couplings, in a possibly large set of variables. If causalities are reliably detected, i.e. causal structure is revealed, estimation of their quantifiers is not considered as too problematic and such measures as the average causal effect (ACE) and similar ones often appear sufficient [42,118]. Some works in the field of SCM consider causal couplings between processes, but with the same focus: either causal discovery (e.g., [119–122]) or finding a single widely applicable quantifier somewhat generalizing the ACE (e.g., [27]). So *causality quantification* has not arisen there as a separate field, as a rich and complicated problem requiring its own concepts (language) and theory for a fuller realization of the interventional causality ideas.

3. Stochastic processes and time series analysis

Analysis of couplings between processes is to a significant extent a separate field which have been considered within mathematical statistics as an inverse problem of the stochastic processes theory, e.g., [123–125]. The most well-developed formalisms are cross-correlations and cross-spectra with multiple applications which constitute a good deal of time series analysis techniques, e.g., [126,127]. The problem of revealing *causal* couplings between processes has been raised in the works of Wiener [51] and Granger [52] (Wiener-Granger causality) and continued to Granger-Geweke causality spectra [99,100], directed information transfer [128], partial directed coherence [129], and multiple elaborations on these approaches, e.g., advanced methods suggested in Refs. [10,60]. All such characteristics are estimated from passive observations. They are associative, no causal assumptions are most often expressed in their definitions, but in many cases (with toy models) such quantifiers are shown to provide reasonable characterization of causal influences. As a result, there are many attempts to interpret them in practice as causality quantifiers without further justifications.

The reasons for multiple spurious conclusions are analyzed, e.g., in Ref. [130].

A recent example of fierce debates concerns the Granger-Geweke causality spectra which represent “information flow” in the sense of Wiener-Granger causality (or of the infinite-history transfer entropy [15]). Their applications constitute already a large field of research in neuroscience; e.g., Ref. [61] provides a review. A recent PNAS paper [45] has still criticized them for disagreement with “intuitive notion of causality” and suggested switching to the system identification perspective. The replies [46,47] have defended those quantifiers and clarified the estimation issues. Concerning causal interpretations, Ref. [47] indicates that Granger-Geweke causality is “data-driven and ‘data-agnostic’ (it makes few assumptions about the generative process ...) ... and as such is well-suited to exploratory analyses. It delivers an information-theoretic interpretation ... which also stands as an *effect size* for directed information flow between components of the system.” Then these quantifiers are in fact understood *only* as asymmetric associative characteristics expressed with information-theoretic tools. Attempts of causal interpretation in the “intervention-effect” sense are not respectable. This is cautious and accurate. But if we just stop at this point, doesn’t it get similar to the above situation in statistics described as “causality in search of language and, simultaneously, the language of causality in search of its meaning”?

4. Dynamical systems and theory of oscillations

The theory of oscillations can be understood as a physical interpretation of the “pure” theory of dynamical systems [83,85,86]. This is how it is understood by the “Russian school” (called thus, e.g., in Ref. [85]) in nonlinear oscillations theory [84,109]. Initially, this discipline relied on the concept of deterministic dynamical system [84–86] specified, e.g., by ordinary differential equations (ODEs) $\dot{x}_t = f_x(x_t, y_t)$, $\dot{y}_t = f_y(x_t, y_t)$, which was then generalized to the concept of stochastic dynamical system (SDS) specified, e.g., by stochastic differential equations (SDEs) [87,88,131] $\dot{x}_t = f_x(x_t, y_t) + \xi_{x,t}$, $\dot{y}_t = f_y(x_t, y_t) + \xi_{y,t}$. With ODEs, the basic problem setting is the initial-value (Cauchy) problem, where one specifies an initial state (x_0, y_0) and obtains a unique future time realization as a particular solution. With SDEs, one similarly specifies an ensemble of initial states (i.e., a PDF, Dirac δ is a special case) as an initial condition for the Fokker-Planck equation (e.g., [97,98]) and finds the PDF of future time realizations. So an evolution equation specifies causal relations between an initial condition and a future (e.g., [103]). Physicists naturally deal with initial-value problems and such causal relations, so they do not need to discuss the term “cause” since they have no difficulties with it: all assumptions are explicit. Then evolution equations of an SDS encode a causal meaning in their syntax like structural equations in SCM: a dependence of the left-hand side (which is a state temporal derivative in differential equations or a future state in difference equations) on the right-hand side (which is a function of an initial state and noise) is often derived from certain physical laws and considerations.

Relations between an initial state (x_0, y_0) and a future state (x_t, y_t) at a concrete t are not *per se* of interest to the theory of oscillations which studies an oscillatory (or any complex) behavior *as a whole* contrary to “previous dynamics” interested in finding concrete values at concrete times as stressed in the foundational monographs [84,109] (see also Ref. [111]). Therefore, a coupling role within the theory of oscillations is adequately characterized by studying how a dynamical regime changes when a nonzero coupling between subsystems (say, $Y \rightarrow X$) is introduced, i.e., whether synchronization is established or not, etc. Therefore, in time series analysis, the “dynamical systems community” readily used cross-correlations and cross-spectra as tools to characterize dynamics as a whole, but was long not interested in such details as prediction improvement for the near future (as given by the Wiener-Granger causality) or detection of a dependence of a given variable at a concrete time $t > t_0$ on different variables at initial time t_0 (as studied in SCM).

Starting from the works [1,2] (from the side of information theory and nonlinear dynamical systems) and [3] (from the side of the theory of oscillations itself), the problem of revealing causal (also called there directional) couplings between oscillators from time series (i.e., an inverse problem) has attracted considerable attention in that community as well. Due to richness of possible dynamical characterization, different groups have suggested numerous techniques to detect causal couplings and developed various *quantifiers*. They have been published along with their applications in many hundreds of papers, selected works are cited here as Refs. [1–41]. So *causality quantification for processes* has become even larger and more diverse field than previously. However, a large number of various approaches has still been lacking a unifying oscillation-theoretic viewpoint to derive different causality quantifiers from a single principle and, therefore, does not make a united discipline. Many authors develop their ideas implying an underlying SDS, but they often try to make the respective time series estimation techniques universally applicable, i.e., free of any model assumptions. Thereby, a natural underpinning for causality quantifiers coming from the dynamical systems perspective often remains *implicit* and so tends to disappear from subsequent applied works.

For example, the transfer entropy has been suggested [1] with the premise of a Markov process and conditioning on an initial state. It is related in essence to the assumption of an SDS and initial condition variations explicitly formulated here. However, in many subsequent works (e.g., [15,26,27]) it has been understood from a stochastic process viewpoint as allowing an infinite-history conditioning (discussed, e.g., in a monograph [50]). Currently, it is often claimed to mean only “information flow” in the sense of associative characteristic, e.g., one says that the TE does not measure “causal mechanism,” but only “causal effect” [132] which can then be understood as an effect of taking the data from one process into account when predicting the future of another. Many other useful quantifiers are also considered independently of each other as being related only to specific time series analysis ideas. Doesn’t it remind one of a possible “escape of a causal content” from the currently used causality (or directional coupling) quantifiers for processes? Then, rephrasing Pearl’s formulation, we need to transform those causality quanti-

fiers into a mathematical object with well-defined semantics and well-founded logic, to resolve controversies, to explicate slippery concepts, etc. For that, the firm basis of the explicit interventionist viewpoint and the paradigm of a stochastic dynamical system can naturally be used.

5. Contribution of this work

This work originates from the theory of oscillations and is inspired by the SCM's "explication spirit" in combination with the (stochastic) dynamical systems perspective. It develops a formalism which allows us to derive various causality quantifiers for processes from a single principle making the entire field intrinsically united and providing a basis for the development of a *formal theory of causality quantifiers for processes*. It seems that the problem is posed in such a generality for the first time.

This work focusses on the basic setting of two subsystems X and Y constituting an SDS. The entire SDS and so the causal structure are taken here to be fully known, i.e., causal inference is not needed. So the problem of *causality quantification* is a *direct* problem, *not an inverse* one. This work takes into account quantifiers used in SCM and aims at a fuller realization of the SCM ideas in the field of dynamical systems which has almost not been explored by SCM explicitly.

Concerning the connection of this work to previous works on causality quantifiers for processes, it aims at arranging those quantifiers into a united system. It allows us to reveal interrelations between existing quantifiers and shows how novel quantifiers can be produced. This work *does not* suggest any quantifier *instead of or in addition to* previously known quantifiers: It shows their deep common roots and tries to create a united discipline from a large set of previously disjoint, independent approaches.

6. Level of mathematical rigor

The theory of oscillations is a *physical* theory which abstracts from a concrete physical origin of a system under study as formulated in [110]. It is not a (purely) mathematical theory since it deals with such physical (though abstract) notions as initial state and phase orbit of an oscillatory system, constant parameter of a system, amplitude of oscillations, resonance, synchronization, changes of dynamics under parameter variations, and others, which readily get filled with a concrete physical content (mechanical, electrical, etc.). Thus, a state x may be a vector containing the coordinate and velocity of a mechanical oscillator X , an individual parameter a_y may be the natural frequency of a system Y , a coupling parameter a_{xy} may represent a physical coupling $Y \rightarrow X$ as a spring constant for mechanical oscillators, etc.

The theory of oscillations studies such dynamical (oscillatory) phenomena as, e.g., coherent or stochastic resonance, synchronization as mutual adjustment of oscillation rhythms, etc. It uses mathematical formalism of the theory of dynamical systems. However, the pure theory of dynamical systems (either deterministic [86] or random [83] ones) as a mathematical discipline does not often refer to resonance, synchronization, etc., but deals with rigorous conditions for existence and uniqueness of solutions, attractors, etc. The theory of oscillations is interested in such rigorous questions in the second

turn, as soon as they relate more closely to some physical situations.

In the same sense, the present work provides a "physical" formalism. All the notions involved (initial states, future response, long-term changes of dynamics under parameter variations, etc.) can readily be applied to describe such oscillatory phenomena as resonance, synchronization, etc. Even though some rigorous details are given here, such mathematical issues as exact conditions for the existence of certain limits and integrals are beyond the scope of this work. However, all notions are introduced in a completely operational way and can be easily proven to exist at least in the examples studied here and in a wide range of basic SDSs often used in modeling of various physical processes, e.g., for stochastic linear (and low-dimensional nonlinear) damped oscillators, stationary linear SDS of arbitrary dimension, and finite-state Markov chains.

APPENDIX B: PRELIMINARIES

This Appendix recapitulates well-grounded concepts of causality and dynamical systems which form the basis of the DCE formalism. This is the concept of interventional causality (Appendix B 1) combined with the stochastic dynamical systems viewpoint (Appendix B 2).

1. Causality and interventions

Let us return to the example of a random variable U affected by another variable V mentioned in Sec. II B, where one compares the interventional PDFs $p(u|\text{do}(v^*))$ and $p(u|\text{do}(v^{**}))$ to characterize a causal effect $V \rightarrow U$, e.g., as the ACE $E(U|\text{do}(v^{**})) - E(U|\text{do}(v^*))$. Obviously, a conditional PDF $p(u|v)$ in passive observations may well differ from the interventional PDF $p(u|\text{do}(v))$. Statistical dependence between U and V , i.e., different PDFs $p(u|v^*)$ and $p(u|v^{**})$ for some v^* and v^{**} , may arise also due to the influence $U \rightarrow V$ or due to the influences of a hidden third factor W called common driver or confounder: $W \rightarrow U$ and $W \rightarrow V$. If both these situations are excluded, then the observational PDF $p(u|v)$ reflects the causal coupling $V \rightarrow U$. If, moreover, an intervention $\text{do}(v)$ does not change "the mechanism" underlying the coupling $V \rightarrow U$, then $p(u|v) = p(u|\text{do}(v))$. Under these assumptions, the PDF $p(u|\text{do}(v))$ and any effect $V \rightarrow U$ can be recovered from passive observations of (U, V) .

As an explicit functional form for such causal relations, one often uses the *structural causal model* (e.g., [42]) which states that the value of the variable U in each trial is *generated* via a certain function $u = \Phi(v, \xi)$ where ξ is the value of a random variable independent of V (an exogenous variable). In contrast, imposing $U = u$ does not affect V : the inverse (with respect to u) function $v = \Phi_u^{-1}(u, \xi)$ has no causal meaning. The function Φ represents "the mechanism" of the influence $V \rightarrow U$. This is what should be preserved during an intervention for the observational and interventional conditional PDFs $p(u|v)$ and $p(u|\text{do}(v))$ to coincide.

If a third factor W is present and its value can also be imposed, then an effect $V \rightarrow U$ not mediated by W is defined via comparison of the PDFs $p(u|\text{do}(w^*), \text{do}(v^*))$

and $p(u|\text{do}(w^*), \text{do}(v^{**}))$. If W contains all potential confounders, the discovery of the coupling $V \rightarrow U$ from passive observations of (U, V, W) is possible, while it may be impossible from passive observations of (U, V) . The large field of *causal inference* often based on structural causal modeling is strongly interested in *causal discovery* in large sets of variables [42,43,118]. In particular, the mediation phenomena are quite important both conceptually and practically (e.g., [66,118]). Estimation of “causal strengths” is also a part of causal inference, but quantifiers in use remain mostly simple, e.g., the ACE often suffices to characterize relations between variables (see Appendix A 2). Even when one deals with time-varying quantities, e.g., [27,40,57,95,119–122,133–135], a purposeful choice of a causality quantifier from a multitude of characteristics and their interrelations are usually not the subject of any attention.

2. Dynamical systems

The above concept of interventional causality is ubiquitous also in the fields of mathematical modeling and dynamical systems, but that word is rarely used there. Consider an autonomous deterministic dynamical system characterized with a state vector $x_t \in R^n$ at time t . Its evolution from any initial state x_0 to a future state x_t ($t > 0$) is specified with a deterministic evolution operator Φ_t^{det} as $x_t = \Phi_t^{\text{det}}(x_0)$. The latter may result from integration of differential equations over an interval $(0, t)$ starting from a given initial state, which is the usual initial value problem in mathematical physics. Then, instead of saying that the value x_0 is *given*, one can equivalently say that it is *imposed*. The conditional PDF $p(x_t|\text{do}(x_0))$ is given by the Dirac $\delta p(x_t|\text{do}(x_0)) = \delta[x_t - \Phi_t^{\text{det}}(x_0)]$. The pair (X_0, X_t) corresponds to the above pair (V, U) , and there is a causal coupling $X_0 \rightarrow X_t$ in accordance with the above terminology: any influence $X_t \rightarrow X_0$ is excluded by the meaning of a dynamical system as a model of physical processes with the arrow of time, while any third factor is absent by definition. If some value $x_{t'} = x^*$ is encountered in a passively observed time series at any time instant t' , the future is exactly the same as after imposing the value x^* at $t = 0$. So the “interventional” PDF $p(x_t|\text{do}(x_0))$ coincides with the observational PDF $p(x_t|x_0)$. Physicists rarely use the term “causality” since under the dynamical systems setting there are no problems with understanding causality and distinguishing it from the ordinary correlation. Still, they sometimes do so; e.g., such famous researchers as Kalman *et al.* [103] found it relevant to claim that the evolution equations of a dynamical system determine “the flow of causation from past into future” [103] (p. 5).

An important generalization of the deterministic dynamical system is the *stochastic dynamical system* (SDS), where x_t at $t > 0$ depends also on a random event $\xi_{(0,t)}$ which acts on X over an interval $(0, t)$ and is independent of x_0 and any previous random events $\xi_{(t',0)}$ with $t' < 0$, e.g., [97,125]. Widely known examples of the SDS and $\xi_{(0,t)}$ are (1) stochastic difference equations or autoregressive processes (e.g., [136]) where $\xi_{(0,t)}$ is a finite sequence of values of independent identically distributed (i.i.d.) random variables; (2) stochastic differential equations (driven by white noise; e.g., [88]) where $\xi_{(0,t)}$ can be approximated with a finite i.i.d. sequence;

and (3) Markov chains where $\xi_{(0,t)}$ is an abstract random event. A full evolution operator of an SDS can be written as $x_t = \Phi_t^{\text{sto}}(x_0, \xi_{(0,t)})$ with $\Phi_t^{\text{sto}}: R^n \times \Xi_{(0,t)} \rightarrow R^n$ where $\xi_{(0,t)} \in \Xi_{(0,t)}$ in the respective probability space (see Appendix C 1). This is a particular case of the structural causal model where the causal meaning takes place due to the arrow of time. Given x_0 , the random variable X_t is a specific function of a random event $\xi_{(0,t)}$ and so is characterized with a respective PDF $p(x_t|\text{do}(x_0))$. This PDF is also called transition PDF. So an SDS produces a future random variable X_t for a given initial state x_0 , i.e., performs a mapping $X_t = \Phi_t(x_0)$ with $\Phi_t: R^n \rightarrow V_n(\Xi_{(0,t)})$ where $V_n(\Xi_{(0,t)})$ denotes the space of n -dimensional random variables $X(\xi_{(0,t)})$ which are functions of $\xi_{(0,t)} \in \Xi_{(0,t)}$. The PDF $p(x_t|\text{do}(x_0))$ is no longer a Dirac δ in general, but is still uniquely determined by the imposed initial state x_0 and can be obtained, e.g., via solving an initial value problem for the Fokker-Planck equation [97,98] in the case of stochastic differential equations. An SDS’s evolution is a (first-order) *Markovian* stochastic process X_t because $\xi_{(t,t+\tau)}$ is independent of X_t and any previous $\xi_{(t',t)}$ with $t' < t$ (there are no confounders of X_0 and X_t). Therefore, the transition PDF $p(x_t|x_0)$ for passive observations coincides with the interventional PDF $p(x_t|\text{do}(x_0))$.

The SDS is a particular case of random dynamical system (RDS) which is called *Markovian RDS*; see Ref. [83], pp. 105–107, and Appendix C 2 below. A general RDS may involve, e.g., nonwhite noises and so the interventional and passively observed transition PDFs may not coincide. That case is not considered in this work, but may be addressed similarly to SDS and requires additional assumptions when one turns to estimation of causality quantifiers from a time series (see Appendix F 2).

APPENDIX C: TERMS AND NOTATIONS

This Appendix provides technical details for the DCE formalism: the suggested novel terms and notations are justified in Appendixes C 1, C 2, and C 3, while full mathematical definitions of all elements of the general DCE are given in Appendix C 4.

1. Functional conditioning

Consider a random vector-valued variable consisting of two components (U, V) (each is a finite-dimensional vector) which is function of a random event ω defined on some *probability space*; e.g., [137–140]. The latter is a triple which includes a space Ω of elementary events ω , a σ -algebra B of its subsets called *events*, and a probability measure P on $\{\Omega, B\}$. One sometimes calls $\{\Omega, B\}$ *field of events* [140]. Denote PDF of (U, V) as $p_{UV}(u, v)$ which can be a generalized function with Dirac δ components. Denote functionals of a random variable U with parentheses: expectation $E(U)$, variance $\text{var}(U)$ (sometimes σ_U^2), and (differential) Shannon entropy $H(U)$.

Conditional PDF of U under the condition $V = v$ is $p_{U|V}(u|v) = p_{UV}(u, v)/p_V(v)$ where $p_V(v)$ is marginal PDF of V . In statistical sense, the joint PDF $p_{UV}(u, v)$ describes an infinite ensemble of realizations (u, v) called an independent sample. Then $p_{U|V}(u|v^*)$ describes an infinite ensemble

of trials with such realizations (u, v) that $v = v^*$, i.e., with marginal PDF of V equal to $\delta(v - v^*)$, which can be selected from the above original ensemble. By construction, the random vector (U, V) with $v = v^*$ is defined on the same field of events $\{\Omega, B\}$, but with a probability measure different (in general) from the original measure P .

Let us create another infinite ensemble via selecting realizations (u, v) from the original ensemble of all realizations so that the relative number of realizations with the value of V within an interval $(v, v + dv)$ equals $\rho_V(v)dv$. Then the function $\rho_V(v)$ is a new marginal PDF of V . The resulting vector (U, V) is defined on the same field of events $\{\Omega, B\}$, but with yet another probability measure. The resulting variable U is called here *functionally conditional* because the condition is given by the probability density function of V in contrast to a single value $V = v^*$ used in the ordinary conditioning. Let us denote this variable itself and its PDF with the square brackets as $[U|\rho_V]$ and $p_U[u|\rho_V]$, notice $p_U[u|\delta(v - v^*)] = p_{U|V}(u|v^*)$. The functionally conditional PDF equals $p_U[u|\rho_V] = \int p_{U|V}(u|v)\rho_V(v)dv$. Let us denote its functionals as $\text{var}[U|\rho_V]$ and $H[U|\rho_V]$. Note that the notation $H(U)$ implies some probability measure for the full vector (U, V) but does not show it explicitly, so the meaning of $H(U)$ depends on the context in this respect.

In a practical setting, a dependence between U and V described with $p_{U|V}(u|v)$ may well be determined not only by the influence $V \rightarrow U$, but also by the influence $U \rightarrow V$ or the influence of a hidden third variable (confounder) on U and V . To reveal the influence (causal coupling) $V \rightarrow U$, it is important to distinguish between (passive) selection of trials with $V = v$ from the original ensemble and performing trials by *imposing* $V = \text{do}(v)$ which is a different condition [42]. In do-calculus [42], an intervention $\text{do}(v)$ means that “causal mechanisms” responsible for any influences on V are blocked, while all other “causal mechanisms” remain unchanged. These mechanisms are not given explicitly in the definition of the original probability space $\{\Omega, B, P\}$. Hence, they are defined additionally when one introduces the concept of intervention $V = \text{do}(v)$. In other words, even the field of events corresponding to a variable with *imposed* values $V = \text{do}(v)$ differs from the original field of events $\{\Omega, B\}$ since the latter corresponds to a (passive) selection of trials from the original ensemble with $p_V(v)$. Denote the interventional conditional PDF $p_{U|V}(u|\text{do}(v))$. The PDF $p_{U|V}(u|\text{do}(v))$ may well differ from the original PDF $p_{U|V}(u|v)$ [42]. Despite the same notation $p_{U|V}$, the difference of the two functions is encoded in the notation $\text{do}(v)$. The PDF of imposed values $\rho_V(\text{do}(v))$ can be selected arbitrarily. Then the interventional functionally conditional PDF reads $p[u|\rho_V(\text{do}(\cdot))] = \int p_{U|V}(u|\text{do}(v))\rho_V(\text{do}(v))dv$.

If $p_{U|V}(u|v) = p_{U|V}(u|\text{do}(v))$, then imposition and (passive) selection of $V = v$ are equivalent in terms of the resulting conditional PDF of U . As an example, take a random variable U with realizations generated by a function $u = U(v, \omega)$ whose domain is Cartesian product of two probability spaces with elementary events $V = v$ and ω , where the events ω are described with a probability measure P and independent of V . If the imposition of $V = \text{do}(v)$ maintains the function $U(v, \omega)$ and the measure P unchanged, then the ordinary conditional PDF $p_{U|V}(u|v) = p_{U|V}(u|\text{do}(v))$ and so

describes “an autonomous causal mechanism $V \rightarrow U$.” Then also $p[u|\rho_V] = p[u|\rho_V(\text{do}(\cdot))]$. Such U may be a future state of an SDS at $t > 0$, V is its present state at $t = 0$, and ω is a random event (noise) acting during an interval $(0, t)$. In that case, neither U nor ω can affect V , so there are no confounders of U and V . Imposing an initial state (and looking at the future) and passively selecting it from an observed time series (and tracing the continuation) provide the same conditional PDF of the future. Then the field of events explicitly includes what is “an autonomous causal mechanism $V \rightarrow U$ retained under $V = \text{do}(v)$ ”: this is an evolution operator of the SDS.

2. Stochastic dynamical system

An SDS is understood here as a system whose initial *state* uniquely determines all future PDFs, i.e., whose evolution is a (first-order) Markov process Z_t : given an initial state z_0 , a PDF of any future Z_t ($t > 0$) does not depend on the past states z_t (with $t < 0$); e.g., [97]. Concerning the term, the foundational monograph [83] uses the name “random dynamical system” (RDS) which is, roughly speaking, a dynamical system under the influence of any stationary noise. Evolution of its variables may not be a Markov process. In continuous time, such an RDS can be specified, e.g., with a “random differential equation” (RDE; see [83], pp. 57–58), i.e., an ordinary differential equation driven by noise whose realizations can be integrated in a usual sense. “Stochastic differential equation” (SDE) is the name for a differential equation driven by white noise ([83], pp. 68–71). Its solution is a Markov process, and the respective RDS is a particular case of Markovian RDS ([83], pp. 105–106). Since such systems are especially relevant here, the term SDS is used for convenience to denote any Markovian RDS, including SDEs, discrete-time and discrete-state systems, etc. Many works understand the term SDS in this sense. In the same spirit, the monograph [141] calls a dynamical system generated with an SDE “stochastic differential system.” The term SDS is also often used in a more general sense close to the RDS, e.g., in Refs. [88, 89, 131, 142, 143], but even there an SDS is more often related to Markov processes. A similar notion widely used in time series analysis is “state space model”; e.g., [101].

3. Distinction functional

To quantify the difference between any two vectors of a metric space, in particular, between random variables U^* and U^{**} , one can use such notions as metrics, distance, and divergence; see, e.g., a comment in Ref. [38]. Any of these characteristics is zero if $U^* = U^{**}$ in a relevant probabilistic sense (e.g., almost surely or in the mean square sense). “Metrics” as the most strict notion [38] is nonnegative and symmetric, and implies the triangle inequality. The mean-squared difference is an example of metrics. “Distance” is often used as a synonym of metrics, but sometimes [38] with possible violation of the triangle inequality. “Divergence” is nonnegative, but not necessarily symmetric. In particular, the widely used Kullback-Leibler divergence (KLD) is a divergence of distributions given as $D_{KL}(p_{U^*}||p_{U^{**}}) = \int p_{U^*}(u) \ln[p_{U^*}(u)/p_{U^{**}}(u)]du$. One can use quantifiers of difference which are not necessarily nonnegative, e.g., the

difference of expectations $E(U^{**}) - E(U^*)$ defines the average causal effect (ACE) [42]. Similarly, the difference of the logarithms of the generalized variances (i.e., the determinants of covariance matrix) of certain vectors $\ln |\Sigma(U^{**})| - \ln |\Sigma(U^*)|$ defines the Wiener-Granger causality [15,144], and the difference of the Shannon entropies $H(U^{**}) - H(U^*)$ of certain vectors defines the transfer entropy.

To include all these possibilities, *distinction functional* $\{U^*||U^{**}\}$ or just *distinction* is defined here as any continuous functional of the pair (U^*, U^{**}) which is equal to zero if $U^* = U^{**}$ in a relevant probabilistic sense. The two vertical lines in the notation remind the KLD which is asymmetric in respect of its arguments. Combination of this delimiter with the figure brackets shows that it is not generally the KLD, but can be any functional, since different kinds of brackets often denote different functionals not even restricted to take on positive values only. Thus, the suggested notation $\{U^*||U^{**}\}$ vividly reminds three aspects of the distinction: a functional, possibly asymmetric, possibly taking on negative values.

4. Definitions

Consider an SDS \mathcal{S} consisting of two subsystems X and Y with a full state vector $(x_t, y_t) \in R^{n+m}$ where $x_t \in R^n$ and $y_t \in R^m$. Let its operator $\Phi_t: R^{n+m} \rightarrow V_{n+m}(\Xi_{(0,t)})$ be well defined for any $t > 0$ producing a random future state $(X_t, Y_t) = \Phi_t(x_0, y_0)$ where the space $V_n(\Xi_{(0,t)})$ is defined in Appendix B 2. Its X projection is given by the operator $\Phi_t^X: R^{n+m} \rightarrow V_n(\Xi_{(0,t)})$ which maps an initial state to a random variable $X_t = \Phi_t^X(x_0, y_0)$. Similarly, the other projection operator is $\Phi_t^Y: R^{n+m} \rightarrow V_m(\Xi_{(0,t)})$. Both the time t and the state vector can be either continuous or discrete.

The operator Φ_t is well defined for a wide range of physically interesting SDSs, e.g., for any Markov chains and difference equations (discrete-time systems) or stochastic differential equations (6) with sufficiently smooth (satisfying Lipschitz's condition; e.g., [145], p. 181) functions on their right-hand side. Below the future state is assumed to possess a sufficiently regular PDF for a distinction functional of interest (e.g., the difference of the Shannon entropies) to exist. For some DCEs to be well defined, a stationary PDF must also exist (e.g., for the TE) and the respective conditions are easily formulated for Markov chains and linear systems. In this work, all such mathematical conditions are implied to be met when a particular DCE is defined. Their more exact and rigorous formulations are not the subject of this work and relate to the properties of a concrete SDS, not specifically to a DCE. At the physical level of rigor, it suffices that all quantities under study exist for a wide range of paradigmatic SDS (linear stochastic differential equations, some low-dimensional nonlinear oscillators, Markov chains, etc.) and are defined in a completely constructive manner to be numerically estimated from ensembles of realizations for any SDS whose evolution can be observed from any initial state and for any parameter value of interest.

Let us specify an ensemble of initial states with a PDF $\rho_{XY}(x_0, y_0) = \rho_X(x_0)\rho_{Y|X}(y_0|x_0)$ and call it a (functional) initial condition. Let us denote $[X_t|\rho_{XY}] = \tilde{\Phi}_t^X(\rho_{XY})$ the functionally conditional random variable X_t obtained under the condition ρ_{XY} through the operator $\tilde{\Phi}_t^X: L_\rho \rightarrow$

$V_n(X_0, Y_0, \Xi_{(0,t)})$, where L_ρ is the space of PDFs (nonnegative with unit integral) and $V_n(X_0, Y_0, \Xi_{(0,t)})$ is the space of random variables $X(x_0, y_0, \xi_{(0,t)})$ defined over the respective probabilistic space. Let us denote the PDF of $[X_t|\rho_{XY}]$ as $p_X^{(t)}[x_t|\rho_{XY}]$.

The operators $\tilde{\Phi}_t$, $\tilde{\Phi}_t^X$, and $\tilde{\Phi}_t^Y$ are defined in a completely operational way. If an SDS' evolution is well defined over an interval $(0, t)$, one can compute or observe a future value x_t given an initial state (x_0, y_0) and a particular event $\xi_{(0,t)}$. This future is given through some operator $x_t = \Phi_t^{st0}(x_0, y_0, \xi_{(0,t)})$. Such an x_t is a realization of the functionally conditional variable $[X_t|\rho_{XY}]$ where one randomly draws (x_0, y_0) according to the PDF $\rho_{XY}(x_0, y_0)$ and independently draws $\xi_{(0,t)}$ according to its own probability measure. The PDF $p_X^{(t)}[x_t|\rho_{XY}]$ can be estimated if one performs many independent trials and obtains a set of values of x_t .

Definition 1. Call any ordered pair of functional initial conditions $(\rho_{XY}^*, \rho_{XY}^{**})$ an *initial condition variation* in an SDS \mathcal{S} . Call ρ_{XY}^* a *reference* initial condition and ρ_{XY}^{**} an *alternative* initial condition.

To quantify a coupling $Y \rightarrow X$, consider $(\rho_{XY}^*, \rho_{XY}^{**})$ with the same marginal PDF $\rho_X^*(x_0) = \rho_X^{**}(x_0)$ and generally different $\rho_{Y|X}^*(y_0|x_0)$ and $\rho_{Y|X}^{**}(y_0|x_0)$.

Definition 2. For an SDS \mathcal{S} , call the ordered pair $([X_t|\rho_{XY}^*], [X_t|\rho_{XY}^{**}])$ produced by the operator $\tilde{\Phi}_t^X$ the *future response* of X on the *temporal horizon* $t > 0$ to the initial condition variation $(\rho_{XY}^*, \rho_{XY}^{**})$.

To quantify the "strength" of this response with a scalar, one can select any continuous functional of the two futures which should take on zero value if these futures are equal to each other for any random event $(x_0, y_0, \xi_{0,t})$, but may well be nonzero for a less strict coincidence. Indeed, realizations of $[X_t|\rho_{XY}^*]$ and $[X_t|\rho_{XY}^{**}]$ can be generated jointly, i.e., in the same trial: for example, one can generate the two futures in each single trial with the same noise realization for both of them. If, moreover, both initial conditions are Dirac δ 's, then one compares just two particular time realizations of an SDS starting from different initial states and driven with the same noise realization. This is similar, e.g., to the definition of the conditional Lyapunov exponents. Under any kind of mutually dependent joint generation, the joint PDF $p([X_t|\rho_{XY}^*], [X_t|\rho_{XY}^{**}])$ is not the product of the two marginal PDFs and its mixed momenta can be used to characterize the difference between $[X_t|\rho_{XY}^*]$ and $[X_t|\rho_{XY}^{**}]$. In case of the independent generation of $[X_t|\rho_{XY}^*]$ and $[X_t|\rho_{XY}^{**}]$ (i.e., independent evolutions of the two ensembles), one compares the marginal PDFs of $[X_t|\rho_{XY}^*]$ and $[X_t|\rho_{XY}^{**}]$ as is most often the case. Let us call the functional selected to compare the two futures *distinction functional* (or just *distinction*) and denote it with another kind of brackets with a delimiter $\{\cdot||\cdot\}$.

Definition 3. Call the *distinction functional* such a continuous functional of the future response $([X_t|\rho_{XY}^*], [X_t|\rho_{XY}^{**}])$ which is zero at least if $[X_t|\rho_{XY}^*] = [X_t|\rho_{XY}^{**}]$ for any realization and can be nonzero otherwise. Denote it $\{\{[X_t|\rho_{XY}^*]||[X_t|\rho_{XY}^{**}]\}$ where $\{\cdot||\cdot\}: V_n(X_0, Y_0, \Xi_{(0,t)}) \times V_n(X_0, Y_0, \Xi_{(0,t)}) \rightarrow R$.

Examples of distinction functionals are diverse. Thus, for an initial variation given by two Dirac δ initial conditions and joint generation of X_t with the same particular noise realization $\xi_{(0,t)}$, the two futures are just usual nonrandom vectors x_t^* and x_t^{**} and their distinction can be just the Euclidean

norm of their difference. Such a distinction functional makes use of its “access” to the detailed generation mechanism $\Phi_t^{sto}(x_0, y_0, \xi_{(0,t)})$ and evaluates X_t for each single noise realization. Further, one can be interested in taking expectation of such a noise-resolved distinction over various $\xi_{(0,t)}$. Then the distinction is a mean-squared difference $\langle ([X_t|\rho_{XY}^*] - [X_t|\rho_{XY}^{**}])^2 \rangle$ over the joint PDF of the future response; i.e., only this joint PDF should be known, not an evolution for each individual noise realization. In even less detail, one can quantify the difference between marginal PDFs $p_X^{(t)}[X_t|\rho_{XY}^*]$ and $p_X^{(t)}[X_t|\rho_{XY}^{**}]$ which is called comparison of random variables *in probability*. Only the latter version is used in Sec. III. The distinction is not necessarily a distance, may be asymmetric, and may take on negative values (see also Appendix C 3).

Definition 4. Call the value of the distinction functional $\{[X_t|\rho_{XY}^*] || [X_t|\rho_{XY}^{**}]\}$ for a single initial variation $(\rho_{XY}^*, \rho_{XY}^{**})$ an *elementary dynamical causal effect* (DCE) in the direction $Y \rightarrow X$.

To introduce a DCE for a set of initial variations, one should somehow *assemble* the elementary DCEs. Let us parametrize all initial conditions with a vector λ and denote them $(\rho_{XY,\lambda}^*, \rho_{XY,\lambda}^{**})$. In particular, for Dirac δ initial conditions located at (x_0^*, y_0^*) and (x_0^{**}, y_0^{**}) , the *assemblage parameter* reads $\lambda = (x_0^*, y_0^*, x_0^{**}, y_0^{**})$. Denote the elementary DCE as

$$C_{Y \rightarrow X,\lambda}^{(t)} = \{[X_t|\rho_{XY,\lambda}^*] || [X_t|\rho_{XY,\lambda}^{**}]\}. \quad (C1)$$

Assemblage may be the average over some *assemblage set* Λ with some weighting function $p_\Lambda(\lambda)$. It can be a maximal value of $C_{Y \rightarrow X,\lambda}^{(t)}$ over Λ or any other functional acting on $C_{Y \rightarrow X,\lambda}^{(t)}$ which is a function of $\lambda \in \Lambda$.

Definition 5. Call any continuous functional acting on the elementary DCE $C_{Y \rightarrow X,\lambda}^{(t)}$ (as a function of $\lambda \in \Lambda$) an *assemblage functional* and denote it $\langle C_{Y \rightarrow X,\lambda}^{(t)} \rangle_\Lambda$, where $\langle \cdot \rangle_\Lambda: L(\Lambda) \rightarrow R$ and $L(\Lambda)$ is the space of scalar-valued functions $f(\lambda)$ with the domain Λ .

The angle brackets are used because it is often some average (aggregation). The simplest example is the *trivial assemblage*, i.e., taking a single initial variation as is often done with parameter variations [29,49,79–81]. Another example is Dirac δ initial conditions with the average of the elementary DCEs over their locations $p_\Lambda(x_0^*, y_0^*, x_0^{**}, y_0^{**}) = \rho_X(x_0^*)\rho_{Y|X}(y_0^*|x_0^*)\rho_{Y|X}(y_0^{**}|x_0^{**})$ suggested for a specific causality quantifier in Ref. [29]. In Appendix D, let us also take the convention that in the case of averaging, the assemblage may be equivalently denoted as $\langle \cdot \rangle_\Lambda \equiv \langle \cdot \rangle_{p_\Lambda}$.

The distinction may in general also depend on its own parameters. For example, it can compare the two functionally conditional future PDFs at a given point x_0^* , rather than over the entire domain. The latter case arises in Sec. III B for the LKIF interpreted as a DCE. Let us include parameters of the distinction into the assemblage parameter λ and write the elementary DCE $\{[X_t|\rho_{XY,\lambda}^*] || [X_t|\rho_{XY,\lambda}^{**}]\}_\lambda$. The initial conditions and the distinction may depend on the same parameters and/or each of them on its own parameters. The assemblage may be performed over all parameters included into λ or over some of them.

Definition 6. Call the value of any assemblage functional $\langle C_{Y \rightarrow X,\lambda}^{(t)} \rangle_\Lambda$ the *dynamical causal effect* $Y \rightarrow X$ of *initial con-*

dition variations in an SDS \mathcal{S} :

$$C_{Y \rightarrow X}^{(t)} = \langle \{[X_t|\rho_{XY,\lambda}^*] || [X_t|\rho_{XY,\lambda}^{**}]\}_\lambda \rangle_\Lambda. \quad (C2)$$

Recall that Eq. (C2) is obtained for a fixed value of a parameter vector $a = (a_x, a_y, a_{xy}, a_{yx})$ in \mathcal{S} . A coupling parameter $a_{xy} = 0$ means that X evolves independently of Y and any DCE $Y \rightarrow X$ (C2) then equals zero. To describe initial variations of a and their effects, let us consider that the DCE $Y \rightarrow X$ can quantify also a change of the future X_t in response to a change of a (either a_{xy} or a_y) which is a parameter variation (a^*, a^{**}) .

Definition 7. Call the combination of the functional initial condition ρ_{XY} and the parameter value a the *generalized initial condition* $\theta_\lambda = \{\rho_{XY,\lambda}, a\}$.

Let the parameter λ include (a^*, a^{**}) . Define the reference generalized initial condition as $\theta_\lambda^* = \{\rho_{XY,\lambda}^*, a^*\}$ and the alternative as $\theta_\lambda^{**} = \{\rho_{XY,\lambda}^{**}, a^{**}\}$. The initial variation $(\theta_\lambda^*, \theta_\lambda^{**})$ may be just an initial condition variation (if $a^* = a^{**}$), or only a parameter variation (if $\rho_{XY,\lambda}^* = \rho_{XY,\lambda}^{**}$), or a mixed one. After the assemblage over such an extended λ , one gets the general DCE in the resulting form (4) which is just the final definition.

Definition 8. Call the value of any assemblage functional $\langle C_{Y \rightarrow X,\lambda}^{(t)} \rangle_\Lambda$ given by Eq. (4) the general dynamical causal effect $Y \rightarrow X$ in an SDS \mathcal{S} .

APPENDIX D: DERIVATIONS FOR LKIF

In the starting work [5], the authors claimed that their quantifier “is consistent with Schreiber’s transfer entropy. The transfer entropy is a Kullback entropy-like quantity. ... The essence of this philosophy is reflected in our formalism. ... However, our formalism differs quantitatively. ... The major difference lies in that A and B [two terms in the expression for TE, $D.S.$] are not strictly in a form of entropy increase, while entropy increase forms the building blocks for our formalism. This difference might lead to different results with the same problem.” This consideration is based mainly on the formal similarity of the two quantifiers. Even after the further works [73–75], it has remained unclear how strongly the values of the two quantifiers can differ and why.

Below, the formulas are first given for $l_{Y \rightarrow X}^{\text{dir}}$ which directly implements the condition “ Y is frozen” in the DCE formalism (Appendix D 1). Second, the LKIF $l_{Y \rightarrow X}$ is derived as a particular DCE to prove Theorem 2 (Appendix D 2). Third, the difference from the original derivation [75] of $l_{Y \rightarrow X}$ is explained (Appendix D 3).

1. Implementation of “ Y frozen”

In the general case of nonzero noises, the Fokker-Planck equation for the system (6) reads

$$\begin{aligned} \frac{\partial p_{XY}^{(t)}}{\partial t} = & - \frac{\partial (f_x p_{XY}^{(t)})}{\partial x} - \frac{\partial (f_y p_{XY}^{(t)})}{\partial y} \\ & + \frac{1}{2} \frac{\partial^2 (g_{xx}^2 p_{XY}^{(t)})}{\partial x^2} + \frac{1}{2} \frac{\partial^2 (g_{yy}^2 p_{XY}^{(t)})}{\partial y^2} \end{aligned} \quad (D1)$$

and describes the evolution of the PDF $p_{XY}^{(t)}(x, y)$ starting from some initial condition $p_{XY}^{(0)}(x_0, y_0) = \rho_{XY}(x_0, y_0)$. Hence, the PDF $p_{XY}^{(t)}(x, y)$ is functionally conditional on $\rho_{XY}(x_0, y_0)$; i.e.,

in our full notation it reads $p_{XY}^{(t)}(x, y) = p_{XY}^{(t)}[x, y|\rho_{XY}]$. The quantities derived below are also functionally conditional, but the notation $[\cdot|\rho_{XY}]$ is often omitted for compactness as in Eq. (D1).

Denote a all parameters which may enter Eq. (6), e.g., parameters of the functions f_x , g_{xx} , etc. Select an arbitrary initial condition $\rho_{XY}(x_0, y_0) = \rho_X(x_0)\rho_{Y|X}(y_0|x_0)$. In the DCE language, the corresponding marginal Shannon entropy rate $\dot{H}(X_t)$ at $t = 0$ describes the evolution from the initial condition $\rho_{XY}^{**} = \rho_{XY}$, i.e., from $\theta^{**} = \{\rho_{XY}, a\}$. Then the condition of “ Y frozen” must correspond to the initial PDF $\rho_{XY}^*(x_0, y_0) = \rho_X(x_0)\delta(y_0 - y_0^*)$ and to the parameter values specified so to “freeze” the dynamics of Y as $f_y = g_{yy} = 0$. To provide the latter, let us introduce an auxiliary parameter i_Y as a multiplier before the entire right-hand side of the equation for dy in Eq. (6): that right-hand side then equals $i_Y(f_y + g_{yy}dw_y)$. Setting $i_Y = 0$ makes Y_t equal to the reference initial state y_0^* at any t (i.e., frozen). The alternative initial condition θ^{**} includes $i_Y = 1$. So let us define a mixed initial variation with the reference $\theta^* = (\rho_X(x_0)\delta(y_0 - y_0^*), i_Y = 0)$ and the alternative $\theta^{**} = (\rho_X(x_0)\rho_{Y|X}(y_0|x_0), i_Y = 1)$, set $t \rightarrow 0$, the distinction as the difference of the Shannon entropies $\{U^*||U^{**}\} = [H(U^{**}) - H(U^*)]/t$, and the assemblage as the average over y_0^* with $\rho_Y(y_0^*)$. Let us denote the resulting DCE $I_{Y \rightarrow X}^{\text{dir}}$ and consider it below.

Let us follow Ref. [75] to derive the functionally conditional $\dot{H}(X_t)$ under the condition θ^{**} . First, integrate both sides of Eq. (D1) over y to get the evolution equation for (the functionally conditional) $p_X^{(t)}(x)$:

$$\frac{\partial p_X^{(t)}(x)}{\partial t} = - \int \frac{\partial(f_x p_{XY}^{(t)})}{\partial x} dy + \frac{1}{2} \int \frac{\partial^2(g_{xx}^2 p_{XY}^{(t)})}{\partial x^2} dy. \quad (\text{D2})$$

All integrals here and below are taken from $-\infty$ to ∞ . Two integrals of the derivatives $\partial(\cdot)/\partial y$ have diminished since they equal the differences of some terms at $\pm\infty$, which are zero since $p_{XY}^{(t)}$ and its derivatives are assumed to quickly decay to zero at infinities. Via multiplying both sides of (D2) by $-[1 + \ln p_X^{(t)}(x)]$, one gets

$$\begin{aligned} - \frac{\partial(p_X^{(t)} \ln p_X^{(t)})}{\partial t} &= \int (1 + \ln p_X^{(t)}) \frac{\partial(f_x p_{XY}^{(t)})}{\partial x} dy \\ &\quad - \frac{1}{2} \int (1 + \ln p_X^{(t)}) \frac{\partial^2(g_{xx}^2 p_{XY}^{(t)})}{\partial x^2} dy. \end{aligned} \quad (\text{D3})$$

Via integrating (D3) over x , one obtains

$$\begin{aligned} \dot{H}(X_t) &= \iint \ln p_X^{(t)} \frac{\partial(f_x p_{XY}^{(t)})}{\partial x} dx dy \\ &\quad - \frac{1}{2} \iint \ln p_X^{(t)} \frac{\partial^2(g_{xx}^2 p_{XY}^{(t)})}{\partial x^2} dx dy. \end{aligned} \quad (\text{D4})$$

Via integrating (D4) by parts, one gets equivalently

$$\begin{aligned} \dot{H}(X_t) &= - \iint f_x \frac{\partial \ln p_X^{(t)}}{\partial x} p_{XY}^{(t)} dx dy \\ &\quad - \frac{1}{2} \iint g_{xx}^2 \frac{\partial^2 \ln p_X^{(t)}}{\partial x^2} p_{XY}^{(t)} dx dy. \end{aligned} \quad (\text{D5})$$

For $t = 0$, it becomes

$$\begin{aligned} \dot{H}[X_t|\theta^{**}] &= - \iint f_x \frac{d \ln \rho_X}{dx_0} \rho_{XY} dx_0 dy_0 \\ &\quad - \frac{1}{2} \iint g_{xx}^2 \frac{d^2 \ln \rho_X}{dx_0^2} \rho_{XY} dx_0 dy_0. \end{aligned} \quad (\text{D6})$$

Here $f_x, g_{xx}, \rho_X, \rho_{XY}$ are functions of x_0 and y_0 as distinct from Eqs. (D1)–(D5) where they are functions of x and y . The arguments are the same as the integration variables and are not explicitly shown for compactness. Noticing that $\frac{1}{\rho_X} \frac{\partial(f_x \rho_X)}{\partial x_0} = f_x \frac{d \ln \rho_X}{dx_0} + \frac{\partial f_x}{\partial x_0}$, one can finally rewrite (D6) as

$$\begin{aligned} \dot{H}[X_t|\theta^{**}] &= - \iint \frac{\partial(f_x \rho_X)}{\partial x_0} \rho_{Y|X} dx_0 dy_0 \\ &\quad + \iint \left(\frac{\partial f_x}{\partial x_0} - \frac{g_{xx}^2}{2} \frac{d^2 \ln \rho_X}{dx_0^2} \right) \rho_{XY} dx_0 dy_0. \end{aligned} \quad (\text{D7})$$

Under the reference θ^* , let us substitute $\rho_{XY}^* = \rho_X \delta(y_0 - y_0^*)$ for ρ_{XY} into Eq. (D6) and get the y_0^* -dependent Shannon entropy rate at $t = 0$ as

$$\begin{aligned} \dot{H}[X_t|\theta^*] &= - \int f_x(x_0, y_0^*) \frac{d \ln \rho_X}{dx_0} \rho_X dx_0 \\ &\quad - \frac{1}{2} \int g_{xx}^2(x_0, y_0^*) \frac{d^2 \ln \rho_X}{dx_0^2} \rho_X dx_0. \end{aligned} \quad (\text{D8})$$

Note that in obtaining this equation through Eqs. (D2)–(D6), the two terms with $\partial(\cdot)/\partial y$ diminish due to $i_Y = 0$. But even if $i_Y \neq 0$, these two terms would diminish in Eq. (D2) just due to the decay of the PDF $p_{XY}^{(t)}$ and its derivatives at infinities. Therefore, due to $t \rightarrow 0$, the mixed initial variation (θ^*, θ^{**}) can be equivalently replaced here with the initial condition variation ($\rho_{XY}^*, \rho_{XY}^{**}$) and $i_Y = 1$ in both θ^* and θ^{**} . The entropy $H[X_t|\theta^*]$ for the initial condition with $i_Y = 0$ differs from that for the same initial condition with $i_Y = 1$ only if t is finite. As well, the higher-order temporal derivatives of $H[X_t|\theta^*]$ at $t = 0$ depend on i_Y , but not the first derivative under consideration. So the DCE $I_{Y \rightarrow X}^{\text{dir}}$ can be equivalently defined as the DCE involving just the initial condition variations ($\rho_{XY}^*, \rho_{XY}^{**}$).

The entropy rate in Eq. (D8) depends on the chosen y_0^* . Let us multiply both sides of Eq. (D8) by $\rho_Y(y_0^*)$ and integrate over y_0^* to get the average quantity

$$\begin{aligned} \langle \dot{H}[X_t|\rho_{XY}^*] \rangle_{\rho_Y(y_0^*)} &= - \iint f_x \frac{d \ln \rho_X}{dx_0} \rho_X \rho_Y dx_0 dy_0 \\ &\quad - \frac{1}{2} \iint g_{xx}^2 \frac{d^2 \ln \rho_X}{dx_0^2} \rho_X \rho_Y dx_0 dy_0, \end{aligned} \quad (\text{D9})$$

where the notation y_0^* on the right-hand side is changed to y_0 just for convenience of further comparisons. Similarly to Eq. (D7), let us rewrite Eq. (D9) as

$$\begin{aligned} \langle \dot{H}[X_t|\rho_{XY}^*] \rangle_{\rho_Y(y_0^*)} &= - \iint \frac{\partial(f_x \rho_X)}{\partial x_0} \rho_Y dx_0 dy_0 + \iint \left(\frac{\partial f_x}{\partial x_0} - \frac{g_{xx}^2}{2} \frac{d^2 \ln \rho_X}{dx_0^2} \right) \\ &\quad \times \rho_X \rho_Y dx_0 dy_0, \end{aligned} \quad (\text{D10})$$

where the first term on the right-hand side diminishes (just integrate first over x_0 to see that) that gives

$$\begin{aligned} & \langle \dot{H}[X_t | \rho_{XY}^*] \rangle_{\rho_Y(y_0^*)} \\ &= \iint \left(\frac{\partial f_x}{\partial x_0} - \frac{g_{xx}}{2} \frac{d^2 \ln \rho_X}{dx_0^2} \right) \rho_X \rho_Y dx_0 dy_0. \end{aligned} \quad (\text{D11})$$

The DCE $l_{Y \rightarrow X}^{\text{dir}}$ is the difference between (D7) and (D11) which equals

$$\begin{aligned} l_{Y \rightarrow X}^{\text{dir}} &= - \iint \frac{\partial(f_x \rho_X)}{\partial x_0} \rho_{Y|X} dx_0 dy_0 \\ &+ \iint \left(\frac{\partial f_x}{\partial x_0} - \frac{g_{xx}}{2} \frac{d^2 \ln \rho_X}{dx_0^2} \right) (\rho_{XY} - \rho_X \rho_Y) dx_0 dy_0. \end{aligned} \quad (\text{D12})$$

Again, all functions under the integrals in Eq. (D12) have the arguments x_0 and y_0 . If the PDF ρ_{XY} is randomized (i.e., $\rho_{XY} = \rho_X \rho_Y$) or the coupling $Y \rightarrow X$ is absent (i.e., both f_x and g_{xx} do not depend on y), then $l_{Y \rightarrow X}^{\text{dir}} = l_{Y \rightarrow X} = 0$. In general, $l_{Y \rightarrow X}^{\text{dir}}$ differs from the LKIF $l_{Y \rightarrow X}$ (10). Note that the Shannon entropy rate conditioned on ρ_{XY}^* (D11) and entering $l_{Y \rightarrow X}^{\text{dir}}$ (D12) is obtained via averaging with $\rho_X(x_0)$ in Eq. (D8) to provide the Shannon entropy of the ensemble, with the result independent of any x_0 or x_0^* . After that, averaging over y_0^* gives such a term in Eq. (D12) which is the average with $\rho_X \rho_Y$. Such an average is absent from the LKIF $l_{Y \rightarrow X}$ by its definition.

2. Proof of Theorem 2: LKIF is a DCE

To show that the LKIF $l_{Y \rightarrow X}$ (14) is a DCE, let us consider the evolution of $h_x(X_t) = -\ln p_X^{(t)}(x)$ as it is also used in Ref. [75]. The quantity $h_x(X_t)$ is the *local entropy* since the Shannon entropy is its weighted average $H(X_t) = \int p_X^{(t)}(x) h_x(X_t) dx$. Let us multiply both sides of Eq. (D2) by $-1/p_X^{(t)}(x)$ and get

$$\frac{\partial h_x(X_t)}{\partial t} = \int \frac{1}{p_X^{(t)}} \frac{\partial(f_x p_{XY}^{(t)})}{\partial x} dy - \frac{1}{2} \int \frac{1}{p_X^{(t)}} \frac{\partial^2(g_{xx}^2 p_{XY}^{(t)})}{\partial x^2} dy. \quad (\text{D13})$$

Recall the initial condition $\rho_{XY} = \rho_X \rho_{Y|X}$ and observe that $h_x(X_t)$ at any $x = x_0^*$ is the functionally conditional quantity $h_{x_0^*}[X_t | \rho_{XY}^*]$ which depends on the parameter x_0^* . Let us denote its rate $\dot{h}_{x_0^*}[X_t | \rho_{XY}^*]$ and rewrite Eq. (D13) for $t = 0$ and the initial condition $\rho_{XY}^*(x_0, y_0) = \rho_X(x_0) \rho_{Y|X}(y_0 | x_0)$ as

$$\begin{aligned} \dot{h}_{x_0^*}[X_t | \rho_{XY}^*] &= \int \frac{1}{\rho_X(x_0^*)} \frac{\partial[f_x(x_0^*, y_0) \rho_{XY}(x_0^*, y_0)]}{\partial x_0^*} dy_0 \\ &- \frac{1}{2} \int \frac{1}{\rho_X(x_0^*)} \frac{\partial^2[g_{xx}(x_0^*, y_0) \rho_{XY}(x_0^*, y_0)]}{\partial x_0^{*2}} dy_0. \end{aligned} \quad (\text{D14})$$

Under the reference initial condition $\rho_{XY, y_0^*}^* = \rho_X(x_0) \delta(y_0 - y_0^*)$, Eq. (D14) becomes

$$\begin{aligned} \dot{h}_{x_0^*}[X_t | \rho_{XY, y_0^*}^*] &= \frac{1}{\rho_X(x_0^*)} \frac{\partial[f_x(x_0^*, y_0^*) \rho_X(x_0^*)]}{\partial x_0^*} \\ &- \frac{1}{2 \rho_X(x_0^*)} \frac{\partial^2[g_{xx}(x_0^*, y_0^*) \rho_X(x_0^*)]}{\partial x_0^{*2}}. \end{aligned} \quad (\text{D15})$$

Let us define the distinction functional depending on the parameter x_0^* as

$$\{[X_t | \rho_{XY, y_0^*}^*] || [X_t | \rho_{XY}^*]\}_{x_0^*} = h_{x_0^*}[X_t | \rho_{XY}^*] - h_{x_0^*}[X_t | \rho_{XY, y_0^*}^*]. \quad (\text{D16})$$

So $\{[X_t | \rho_{XY, y_0^*}^*] || [X_t | \rho_{XY}^*]\}_{x_0^*}$ compares the local entropies $h_{x_0^*}(X_t)$ under the conditions of y_0 distributed with $\delta(y_0 - y_0^*)$ and with $\rho_{Y|X}(y_0 | x_0)$. The value of this functional depends also on y_0^* as a parameter of the reference initial condition $\rho_{XY, y_0^*}^*$. Let us assemble this distinction over y_0^* via averaging with $\rho_{Y|X}(y_0^* | x_0^*)$ and get the DCE depending on the parameter x_0^* as

$$\begin{aligned} & \{ \{ [X_t | \rho_{XY, y_0^*}^*] || [X_t | \rho_{XY}^*] \}_{x_0^*} \}_{\rho_{Y|X}(y_0^* | x_0^*)} \\ &= \int (h_{x_0^*}[X_t | \rho_{XY}^*] - h_{x_0^*}[X_t | \rho_{XY, y_0^*}^*]) \rho_{Y|X}(y_0^* | x_0^*) dy_0^*. \end{aligned} \quad (\text{D17})$$

Now, let us assemble it as the average over x_0^* with $\rho_X(x_0^*)$ and get

$$\begin{aligned} & \{ \{ [X_t | \rho_{XY, y_0^*}^*] || [X_t | \rho_{XY}^*] \}_{x_0^*} \}_{\rho_{XY}(x_0^*, y_0^*)} \\ &= \iint (h_{x_0^*}[X_t | \rho_{XY}^*] - h_{x_0^*}[X_t | \rho_{XY, y_0^*}^*]) \rho_{XY} dx_0^* dy_0^*. \end{aligned} \quad (\text{D18})$$

Dividing Eq. (E20) by t and taking the limit $t \rightarrow 0$, one gets finally

$$\begin{aligned} & \lim_{t \rightarrow 0} \frac{\langle \{ [X_t | \rho_{XY, y_0^*}^*] || [X_t | \rho_{XY}^*] \}_{x_0^*} \}_{\rho_{XY}(x_0^*, y_0^*)} \rangle}{t} \\ &= \iint (\dot{h}_{x_0^*}[X_t | \rho_{XY}^*] - \dot{h}_{x_0^*}[X_t | \rho_{XY, y_0^*}^*]) \rho_{XY} dx_0^* dy_0^*. \end{aligned} \quad (\text{D19})$$

The difference of Eq. (D19) from $l_{Y \rightarrow X}^{\text{dir}}$ is that $l_{Y \rightarrow X}^{\text{dir}}$ involves the term (D11) where the integration is done with the randomized $\rho_X \rho_Y$ because the full Shannon entropy (D8) is first found as the quantity functionally conditional on $\rho_X(x_0) \delta(y_0 - y_0^*)$ and independent of any x_0^* , and then the average over y_0^* is computed without an opportunity to include any dependence on x_0^* .

To find explicitly the right-hand side of Eq. (D19), note that it consists of the two terms (D14) and (D15) multiplied by $\rho_X(x_0^*) \rho_{Y|X}(y_0^* | x_0^*)$ and integrated over x_0^* and y_0^* . The term (D14) does not depend on y_0^* , so only multiplication by $\rho_X(x_0^*)$ and the integration over x_0^* remain:

$$\begin{aligned} & \int \dot{h}_{x_0^*}[X_t | \rho_{XY}^*] \rho_X(x_0^*) dx_0^* \\ &= \iint \left(\frac{\partial(\rho_{XY} f_x)}{\partial x_0^*} - \frac{1}{2} \frac{\partial^2(\rho_{XY} g_{xx}^2)}{\partial x_0^{*2}} \right) dy_0 dx_0^*, \end{aligned} \quad (\text{D20})$$

where all functions under the integral sign are functions of (x_0^*, y_0) . This quantity is equal to zero since the derivatives with respect to x_0^* are integrated over x_0^* . The term (D15) after

multiplication and integration becomes

$$\begin{aligned} & \iint \dot{h}_{x_0^*}[X_t|\rho_{XY}^*(x_0^*, y_0^*)]\rho_{XY}(x_0^*, y_0^*) dx_0^* dy_0^* \\ &= \iint \left(\frac{\partial(\rho_X f_x)}{\partial x_0^*} - \frac{1}{2} \frac{\partial^2(\rho_X g_{xx}^2)}{\partial x_0^{*2}} \right) \rho_{Y|X} dx_0^* dy_0^*, \end{aligned} \quad (\text{D21})$$

where all functions under the integral sign are functions of (x_0^*, y_0^*) . In such a form, it is clear that via subtracting (D21) from zero (D20) to get the DCE (D19), one gets the right-hand side of Eq. (14) for the information flow $l_{Y \rightarrow X}$. So the latter (LKIF) reads

$$l_{Y \rightarrow X} = \lim_{t \rightarrow 0} \frac{\langle \{ [X_t|\rho_{XY}^*(x_0^*, y_0^*)] | [X_t|\rho_{XY}^{**}(x_0^*, y_0^*)] \}_{x_0^*} \rangle_{\rho_{XY}(x_0^*, y_0^*)}}{t}. \quad (\text{D22})$$

This proves Theorem 2 showing that $l_{Y \rightarrow X}$ is a DCE equal to the difference of the local entropy rates at x_0^* (which is the parameter of the distinction) averaged over y_0^* with $\rho_{Y|X}(y_0^*|x_0^*)$ and over x_0^* with $\rho_X(x_0^*)$.

3. Matching with original derivation of LKIF

To obtain Eq. (14), the author of [75] finds $\dot{H}(X_t)$ from the equation for the marginal PDF of X_t (D2) and gets it both in the forms (D6) and (D7). So it is the rate of the functionally conditional Shannon entropy $H(X_t) = H[X_t|\rho_{XY}^{**}]$ under the condition $\rho_{XY}^{**} = \rho_{XY}$. However, to define the quantity \dot{H}_X^* , the author does not consider the Fokker-Planck equation with any initial condition as an initial value problem for an evolving ensemble, but uses another formal consideration.

Still, the Fokker-Planck equation for $p_X^{(t)}$ is used in Ref. [75] with $y = y^*$ considered as a constant parameter, i.e., Eq. (D2) with $p_{XY}^{(t)}(x, y) = p_X^{(t)}(x)\delta(y - y^*)$:

$$\frac{\partial p_X^{(t)}(x)}{\partial t} = - \frac{\partial [f_x(x, y^*) p_X^{(t)}(x)]}{\partial x} + \frac{1}{2} \frac{\partial^2 [g_{xx}^2(x, y^*) p_X^{(t)}(x)]}{\partial x^2}. \quad (\text{D23})$$

This is given as Eq. (11) in Ref. [75], only the notations here somewhat differ. Further, the author claims that the entropy rate of interest \dot{H}_X^* “cannot be obtained from the Fokker-Planck equation (D23), where the dynamics is consistent through time” (the words before Eq. (9) in Ref. [75]) and suggests to return to the definition of the derivative of a stochastic process. It looks like a heuristic reasoning, and the motivation is not explained in more detail. So the author transforms Eq. (D23) into the equation for $-\ln p_X^{(t)}(x)$ and gets the above Eq. (D15) [the formula after Eq. (12) in Ref. [75]]. Let us rewrite it here for $t = 0$ as

$$\begin{aligned} - \frac{\partial \ln p_X^{(t)}(x)}{\partial t} \Big|_{t=0} &= \frac{1}{\rho_X(x)} \frac{\partial [f_x(x, y^*) \rho_X(x)]}{\partial x} \\ &\quad - \frac{1}{2\rho_X(x)} \frac{\partial^2 [g_{xx}^2(x, y^*) \rho_X(x)]}{\partial x^2}. \end{aligned} \quad (\text{D24})$$

Then the author goes to a finite difference on the left-hand side of (D24) via the approximation $\partial \ln p_X^{(t)}(x)/\partial t|_{t=0} \approx$

$[\ln p_X^{(\Delta t)}(x) - \ln \rho_X(x)]/\Delta t$ getting

$$\begin{aligned} - \ln p_X^{(\Delta t)}(x) + \ln \rho_X(x) &= \frac{\Delta t}{\rho_X(x)} \frac{\partial [f_x(x, y^*) \rho_X(x)]}{\partial x} \\ &\quad - \frac{\Delta t}{2\rho_X(x)} \frac{\partial^2 [g_{xx}^2(x, y^*) \rho_X(x)]}{\partial x^2}. \end{aligned} \quad (\text{D25})$$

After that, a stochastic realization $x_{\Delta t}$ occurred after the present value x_0 is substituted into Eq. (D25) instead of x . In the term $\ln \rho(x_{\Delta t})$ on the left-hand side, the argument $x_{\Delta t}$ is replaced with $x_{\Delta t} = x_0 + f_x(x_0, y^*)\Delta t + g_{xx}(x_0, y^*)dw_x$ and the resulting term $\ln \rho_X[x_0 + f_x(x_0, y^*)\Delta t + g_{xx}(x_0, y^*)dw_x]$ is expanded into the Taylor series in Δt and dw_x at x_0 . The argument $x_{\Delta t}$ on the right-hand side is replaced in the same way, but finally $x_{\Delta t}$ appears to be just replaced with x_0 after retaining only the lowest-order terms. It gives

$$\begin{aligned} & - \ln p_X^{(\Delta t)}(x_{\Delta t}) + \ln \rho_X(x_0) - A \\ &= \frac{\Delta t}{\rho_X(x_0)} \frac{\partial [f_x(x_0, y^*) \rho_X(x_0)]}{\partial x_0} \\ &\quad - \frac{\Delta t}{2\rho_X(x_0)} \frac{\partial^2 [g_{xx}(x_0, y^*) \rho_X(x_0)]}{\partial x_0^2}, \end{aligned} \quad (\text{D26})$$

where

$A = - \frac{\partial \ln \rho_X(x_0)}{\partial x_0} [f_x \Delta t + g_{xx}(x_0, y^*) dw_x] - \frac{1}{2} \frac{\partial^2 \ln \rho_X(x_0)}{\partial x_0^2} [f_x \Delta t + g_{xx}(x_0, y^*) dw_x]^2$. Then the author takes expectations of both sides of (D26), i.e., averages over an ensemble of realizations (x_0, y^*, dw_x) [whether y^* is fixed as in Eq. (D23) or not is considered below]. The average $-\langle \ln p_X^{(\Delta t)}(x_{\Delta t}) \rangle$ is further *assumed* to give the entropy $H(X_{\Delta t})$ “as if Y is frozen,” while $-\langle \ln \rho(x_0) \rangle$ to give the entropy $H(X_0)$. If so, one writes $\dot{H}_X^* = (-\langle \ln p_X^{(\Delta t)}(x_{\Delta t}) \rangle + \langle \ln \rho(x_0) \rangle)/\Delta t$. It is further derived [75] that $\langle A \rangle/\Delta t = \dot{H}[X_t|\rho_{XY}^{**}]$ with $\dot{H}[X_t|\rho_{XY}^{**}]$ in the above form (D6). It results in

$$\begin{aligned} \dot{H}_X^* &= \dot{H}[X_t|\rho_{XY}^{**}] \\ &+ \iint \frac{\rho_{XY}(x_0, y^*)}{\rho_X(x_0)} \frac{\partial [f_x(x_0, y^*) \rho_X(x_0)]}{\partial x_0} dx_0 dy^* \\ &- \iint \frac{\rho_{XY}(x_0, y^*)}{2\rho_X(x_0)} \frac{\partial^2 [g_{xx}^2(x_0, y^*) \rho_X(x_0)]}{\partial x_0^2} dx_0 dy^*. \end{aligned} \quad (\text{D27})$$

Subtracting the right-hand side of Eq. (D27) from $\dot{H}[X_t|\rho_{XY}^{**}]$ (D6), the author gets $l_{Y \rightarrow X}$ in the form (14).

The obstacle is that the average $-\langle \ln p_X^{(\Delta t)}(x_{\Delta t}) \rangle$ would be the Shannon entropy only if one used the relevant ensemble [i.e., the PDF of stochastic realizations (x_0, y^*)] over which this average is done. If this ensemble is described with $\rho_{XY, y^*}^* = \rho_X(x_0)\delta(y - y^*)$, then $-\langle \ln p_X^{(\Delta t)}(x_{\Delta t}) \rangle$ is indeed the Shannon entropy at a given y^* as in $l_{Y \rightarrow X}^{\text{dir}}$. The subsequent average with $\rho_Y(y^*)$ would give the Shannon entropy, and the weighting function $\rho_X(x_0)\rho_Y(y^*)$ would enter the right-hand side of Eq. (D27) instead of $\rho_{XY}(x_0, y^*)$ resulting in the above $l_{Y \rightarrow X}^{\text{dir}}$. However, the author [75] averages with the PDF $\rho_{XY}(x_0, y^*)$ in Eq. (D27) violating the condition of fixed y^* (“ Y frozen”), so the general Eq. (D2) should be used

instead of Eq. (D23) to obtain the Shannon entropy $H(X_t)$ of the ensemble starting from $\rho_{XY}^{**} = \rho_{XY}(x_0, y_0)$, since the violated condition was the condition for the applicability of Eq. (D23). Then $-\langle \ln p_X^{(\Delta t)}(x_{\Delta t}) \rangle$ would equal $H[X_t | \rho_{XY}^{**}]$ giving $\dot{H}_X^* = \dot{H}[X_t | \rho_{XY}^{**}]$. So the only possible understanding is that the author averages over realizations (x_0, y^*) distributed with $\rho_{XY}(x_0, y^*)$ but using the simpler FPE corresponding to the frozen $y = y^*$. Hence, $-\langle \ln p_X^{(\Delta t)}(x_{\Delta t}) \rangle$ is *not the Shannon entropy of any ensemble* of time realizations of the system (6) starting from some initial condition for the relevant FPE and taken at Δt , but a mix (over various x_0^*) of the local entropy rates at x_0^* for the ensembles starting from $\rho_{XY, y^*}^{**} = \rho_X(x_0)\delta(y_0 - y^*)$ with different y^* . This is not even the average of any Shannon entropies due to such ‘‘mixing’’: at each x_0^* the local entropy rates for the ensembles with different y^* enter the aggregate quantity with different weights distributed as $\rho_{Y|X}(y^* | x_0^*)$. So it follows from Eqs. (D27) and (D20) that

$$\dot{H}_X^* = \langle \dot{h}_{x_0^*}[X_t | \rho_X(x_0)\delta(y_0 - y^*)] \rangle_{\rho_{XY}(x_0^*, y^*)} + \dot{H}[X_t | \rho_{XY}^{**}]. \quad (\text{D28})$$

Hence, the author first averages the local entropy $h_{x_0^*}(X_t)$ at each x_0^* over a set of y^* distributed as $\rho_{Y|X}(y^* | x_0^*)$ and then averages the result over x_0^* with $\rho_X(x_0^*)$, plus the addendum $\langle A \rangle$ summing the small terms in the expansion of $\ln p_X^{(\Delta t)}(x_{\Delta t})$ in Taylor series at x_0 and averaged over x_0, y^*, dw_x . This is exactly the derivation performed within the DCE framework in Appendix D 2 above.

So \dot{H}_X^* is not the Shannon entropy rate of any ensemble for the SDS (6). From Eq. (D28) one immediately sees $I_{Y \rightarrow X} = \dot{H}[X_t | \rho_{XY}^{**}] - \dot{H}_X^* = 0 - \langle \dot{h}_{x_0^*}[X_t | \rho_X(x_0)\delta(y_0 - y^*)] \rangle_{\rho_{XY}(x_0^*, y^*)}$ as in Appendix D 2. Therefore, the efforts with stochastic realizations in Ref. [75] are unnecessary, the FPE with an appropriate initial condition applies consistently, and the DCE viewpoint reveals the meaning of \dot{H}_X^* . Interestingly, the author of Ref. [75] has found the nice and useful formula (14) even semi-intuitively: the intuitive basis was noted by the authors in their previous work [5]. The explicating DCE viewpoint ‘‘deciphers’’ the meaning of this result as a concrete DCE and, thereby, provides its further theoretical underpinning.

APPENDIX E: DETAILS OF THE NUMERICAL EXAMPLE

This Appendix provides analytic expressions for the DCEs in the example of overdamped oscillators (1).

To find the stationary second-order moments σ_X^2 , σ_Y^2 , and σ_{XY} (zero-lag covariance) for a linear SDS $\dot{\mathbf{z}} = \mathbf{A}\mathbf{z} + \xi(t)$, one solves a linear matrix equation (e.g., [29])

$$\mathbf{A}\mathbf{C}_{zz} + \mathbf{C}_{zz}\mathbf{A}^T + \mathbf{\Gamma} = 0, \quad (\text{E1})$$

where \mathbf{C}_{zz} is the stationary cross-covariance matrix of the state vector \mathbf{z} , \mathbf{T} denotes transposition, and $\mathbf{\Gamma}$ is the noise intensity matrix. In our case, the state vector is two-dimensional $\mathbf{z} = (x, y)$ and Eq. (E1) becomes

$$\begin{aligned} -2a_x\sigma_X^2 + 2a_{xy}\sigma_{XY} + \Gamma_{xx} &= 0, \\ -2a_y\sigma_Y^2 + 2a_{yx}\sigma_{XY} + \Gamma_{yy} &= 0, \\ a_{yx}\sigma_X^2 + a_{xy}\sigma_Y^2 - (a_x + a_y)\sigma_{XY} &= 0, \end{aligned} \quad (\text{E2})$$

By solving it explicitly, one finds the stationary moments as [79]

$$\begin{aligned} \sigma_X^2 &= \frac{\Gamma_{xx}}{2a_x} + \frac{a_x a_{xy}^2 \Gamma_{yy} + a_y a_{xy} a_{yx} \Gamma_{xx}}{2a_x(a_x + a_y)(a_x a_y - a_{xy} a_{yx})}, \\ \sigma_Y^2 &= \frac{\Gamma_{yy}}{2a_y} + \frac{a_y a_{yx}^2 \Gamma_{xx} + a_x a_{xy} a_{yx} \Gamma_{yy}}{2a_y(a_x + a_y)(a_x a_y - a_{xy} a_{yx})}, \\ \sigma_{XY} &= \frac{a_y a_{yx} \Gamma_{xx} + a_x a_{xy} \Gamma_{yy}}{2(a_x + a_y)(a_x a_y - a_{xy} a_{yx})}. \end{aligned} \quad (\text{E3})$$

Recalling the notations $\beta_{xy} = a_{xy}\sigma_{Y,0}/(a_x\sigma_{X,0})$, $\beta_{yx} = a_{yx}\sigma_{X,0}/(a_y\sigma_{Y,0})$, and $r_{st} = \sigma_{XY}/(\sigma_X\sigma_Y)$, one finds $S_{Y \rightarrow X}$ from the first equation of (E3) as

$$S_{Y \rightarrow X} = \frac{\beta_{xy}^2 + \beta_{xy}\beta_{yx}}{(1 - \beta_{xy}\beta_{yx})(1 + m_{xy})}. \quad (\text{E4})$$

Let us consider several simpler quantifiers (see also Sec. I of Ref. [90]) as steps to finding the TE. Note that a future state of the SDS (1) relates to an initial state as

$$\begin{aligned} X_t &= \alpha_X^{(t)} X_0 + \alpha_Y^{(t)} Y_0 + \varepsilon_X^{(t)}, \\ Y_t &= \alpha_Y^{(t)} Y_0 + \alpha_X^{(t)} X_0 + \varepsilon_Y^{(t)}, \end{aligned} \quad (\text{E5})$$

where all α 's are found via solving ordinary differential equations for the conditional expectations

$$\begin{aligned} \frac{dE(X_t | x_0, y_0)}{dt} &= -a_x E(X_t | x_0, y_0) + a_{xy} E(Y_t | x_0, y_0), \\ \frac{dE(Y_t | x_0, y_0)}{dt} &= -a_y E(Y_t | x_0, y_0) + a_{yx} E(X_t | x_0, y_0), \end{aligned} \quad (\text{E6})$$

starting from the initial state (x_0, y_0) at $t = 0$. The solution to these ODEs reads

$$\begin{aligned} E(X_t | x_0, y_0) &= \alpha_X^{(t)} x_0 + \alpha_Y^{(t)} y_0, \\ E(Y_t | x_0, y_0) &= \alpha_Y^{(t)} y_0 + \alpha_X^{(t)} x_0, \end{aligned} \quad (\text{E7})$$

where all α 's do not depend on x_0 and y_0 , being analytically expressed via $(a_x, a_{xy}, a_y, a_{yx}, t)$. Their explicit formulas are needed further only in a simple version for the infinitesimal t . Then the path coefficient [90] $\alpha_X^{(t)}$ describing the influence $Y \rightarrow X$ reads

$$\alpha_X^{(t)} = a_{xy}t, \quad (\text{E8})$$

up to an error of the order $O(t^2)$. The ACE of the Dirac δ initial condition variation $[\delta(x_0 - x_0^*)\delta(y_0 - y_0^*), \delta(x_0 - x_0^*)\delta(y_0 - y_0^{**})]$ is

$$E_{Y \rightarrow X}^{(t)}(x_0^*, y_0^*, y_0^{**}) = a_{xy}(y_0^* - y_0^{**})t. \quad (\text{E9})$$

The contribution of Y_0 to the variance of X_t , given $X_0 = x_0^*$, is obtained using linearity of the system and the conditional distribution $p_{Y|X}(y_0 | x_0^*)$ as in Eq. (E7) and reads

$$\gamma_{Y \rightarrow X}^{(t)} = \alpha_X^{(t)2} \sigma_{Y|X}^2 = a_{xy}^2 \sigma_{Y|X}^2 t^2. \quad (\text{E10})$$

Let us denote the variances and covariance of the noise $(\varepsilon_X^{(t)}, \varepsilon_Y^{(t)})$ in Eq. (E5) as $\sigma_X^{(t)2}$, $\sigma_Y^{(t)2}$, and $\sigma_{XY}^{(t)}$, since they are just conditional variances and covariance of the vector (X_t, Y_t) given any initial state (x_0, y_0) . They are found from the linear ordinary differential equations (e.g., [29])

$$\dot{\mathbf{C}}_{zz} = \mathbf{A}\mathbf{C}_{zz} + \mathbf{C}_{zz}\mathbf{A}^T + \mathbf{\Gamma}, \quad (\text{E11})$$

which read for the SDS (1)

$$\begin{aligned}\frac{d\sigma_X^{(t)2}}{dt} &= -2a_x\sigma_X^{(t)2} + 2a_{xy}\sigma_{XY}^{(t)} + \Gamma_{xx}, \\ \frac{d\sigma_Y^{(t)2}}{dt} &= -2a_y\sigma_Y^{(t)2} + 2a_{yx}\sigma_{XY}^{(t)} + \Gamma_{yy}, \\ \frac{d\sigma_{XY}^{(t)}}{dt} &= a_{yx}\sigma_X^{(t)2} + a_{xy}\sigma_Y^{(t)2} - (a_x + a_y)\sigma_{XY}^{(t)},\end{aligned}\quad (\text{E12})$$

starting from the initial state (0,0,0) at $t = 0$. They are solved exactly, and one gets for the infinitesimally small t and a nonzero Γ_{xx}

$$\sigma_X^{(t)2} = \text{var}(\varepsilon_X^{(t)}) = \Gamma_{xx}t, \quad (\text{E13})$$

up to an error of the order $O(t^2)$. Then the relative contribution of the initial state Y_0 to the variance of X_t , given $X_0 = x_0^*$, reads

$$\kappa_{Y \rightarrow X}^{(t)} = \frac{a_{xy}^2\sigma_{Y|X}^2 t}{\Gamma_{xx}}, \quad (\text{E14})$$

up to an error of the order $O(t^2)$. Its rate at $t = 0$ is

$$\kappa_{Y \rightarrow X} = \frac{a_{xy}^2\sigma_{Y|X}^2}{\Gamma_{xx}}. \quad (\text{E15})$$

It equals the double TE rate $\kappa_{Y \rightarrow X} = 2\tau_{Y \rightarrow X}$; see, e.g., [15,81] or Sec. I in Ref. [90]. Note that $\sigma_{Y|X}^2 = \sigma_Y^2(1 - r_{st}^2)$ and use $\sigma_Y^2 = \sigma_{Y,0}^2(1 + S_{X \rightarrow Y})$ to get

$$\tau_{Y \rightarrow X} = \frac{a_x\beta_{xy}^2(1 - r_{st}^2)(1 + S_{X \rightarrow Y})}{4}. \quad (\text{E16})$$

The LKIF can be found directly from Eq. (14). It has been done in Ref. [76], which gives

$$l_{Y \rightarrow X} = \frac{a_{xy}\sigma_{XY}}{\sigma_X^2}. \quad (\text{E17})$$

To relate this expression to the asymptotic DCE $S_{Y \rightarrow X}$, note that the left-hand side of Eq. (D6) $H[X_t|\rho_{XY}^{**}]$ is zero for $\rho_{XY}^{**} = p_{XY}^{st}$ and recall that p_{XY}^{st} is a two-dimensional Gaussian PDF with zero expectations and second-order moments σ_X^2 , σ_Y^2 , and $\sigma_{XY} = r_{st}\sigma_X\sigma_Y$. Substituting $f_x = -a_x x$, $g_{xx}^2 = \Gamma_{xx}$, $d \ln \rho_X/dx = -x/\sigma_X^2$, and $d^2 \ln \rho_X/dx^2 = -1/\sigma_X^2$ into the right-hand side of Eq. (D6), one gets

$$-a_x + \frac{a_{xy}\sigma_{XY}}{\sigma_X^2} + \frac{\Gamma_{xx}}{2\sigma_X^2} = 0. \quad (\text{E18})$$

Using $\Gamma_{xx} = 2a_x\sigma_{X,0}^2$, one further gets

$$\begin{aligned}l_{Y \rightarrow X} &= \frac{a_{xy}\sigma_{XY}}{\sigma_X^2} = a_x - \frac{\Gamma_{xx}}{2\sigma_X^2} \\ &= a_x \frac{\sigma_X^2 - \sigma_{X,0}^2}{\sigma_X^2} = \frac{a_x S_{Y \rightarrow X}}{1 + S_{Y \rightarrow X}}.\end{aligned}\quad (\text{E19})$$

Thereby, Theorem 4 is proven. Alternatively, this relationship can be checked by directly expressing the covariance from Eq. (E2) as

$$\sigma_{XY} = \frac{\sigma_{X,0}\sigma_{Y,0}(\beta_{xy} + m_{xy}\beta_{yx})}{(1 + m_{xy})(1 - \beta_{xy}\beta_{yx})}, \quad (\text{E20})$$

substituting it into Eq. (E17) and using Eq. (E4) and the definition of β_{xy} .

APPENDIX F: EXTENSIONS

This Appendix discusses extensions of the DCE formalism to the case of more than two subsystems and the inverse problems of DCE estimation and causal discovery.

1. More than two subsystems

Consider the case when three subsystems constitute the entire SDS. Ref. [30] has generalized the TE $T_{Y \rightarrow X}$ to *causation entropy* taking into account the state of Z as $T_{Y \rightarrow X|Z}$. A similar approach is used in [95] where the initial states are specified by intervention on the basis of the randomized stationary PDF. A similar concept of *complete* TE [13,17] takes into account the state of a third subsystem, while the bivariate TE is called then an *apparent* TE [13,17]. There are many studies with partial or conditional characteristics, e.g., the conditional Granger causality [146] and the conditional Granger-Geweke spectra [100,147].

The general rule of how to define the three-subsystems DCE $C_{Y \rightarrow X|Z}^{(t)}$ based on the two-subsystems DCE $C_{Y \rightarrow X}^{(t)}$ (4) may also be readily formulated as follows: (1) in the initial conditions, replace x_0 with the vector (x_0, z_0) to get $\rho^*(x_0, y_0, z_0) = \rho_{XZ}(x_0, z_0)\rho_{Y|XZ}^*(y_0|x_0, z_0)$ and $\rho^{**}(x_0, y_0, z_0) = \rho_{XZ}(x_0, z_0)\rho_{Y|XZ}^{**}(y_0|x_0, z_0)$; (2) a possible parameter variation is again that of a_y and/or a_{xy} ; (3) define the distinction only in respect of X_t as $\{[X_t|\rho^*]||[X_t|\rho^{**}]\}$, rather than in respect of (X_t, Z_t) , i.e., z_0 is only a conditioning variable required to control confounders; and (iv) the assemblage parameter λ may involve parameters determining the initial PDF of Z_0 , e.g., the location of the Dirac δ , $\delta(z_0 - z_0^*)$.

As an example, consider how the causation entropy [30] is naturally produced as a generalization of the TE within the DCE formalism. Consider an arbitrary PDF $\rho(x_0, y_0, z_0) = \rho_{XZ}(x_0, z_0)\rho_{Y|XZ}(y_0|x_0, z_0)$ (in particular, it can be a stationary PDF $p_{XYZ}^{st}(x, y, z)$ of an SDS under study) to define both the initial conditions and the assemblage. Take the reference initial condition $\rho_{x_0^*, y_0^*, z_0^*}^* = \delta(x_0 - x_0^*)\delta(z_0 - z_0^*)\delta(y_0 - y_0^*)$, where y_0 is fixed to y_0^* as in the extended TE $I_{Y \rightarrow X}^{(t)}$. Take the alternative initial condition $\rho_{x_0^*, y_0^*, z_0^*}^{**} = \delta(x_0 - x_0^*)\delta(z_0 - z_0^*)\rho_{Y|XZ}(y_0|x_0^*, z_0^*)$, i.e., y_0 is distributed according to the conditional PDF as in $I_{Y \rightarrow X}^{(t)}$, but with an additional condition $z_0 = z_0^*$. The response of X on a finite horizon t is $([X_t|\rho_{x_0^*, y_0^*, z_0^*}^*], [X_t|\rho_{x_0^*, y_0^*, z_0^*}^{**}])$ exactly as in $I_{Y \rightarrow X}^{(t)}$, only the initial conditions here depend also on the third vector z_0 , i.e., on all confounders. The distinction functional is again the difference of the Shannon entropies $\{[X_t|\rho_{x_0^*, y_0^*, z_0^*}^*]||[X_t|\rho_{x_0^*, y_0^*, z_0^*}^{**}]\} = H[X_t|\rho_{x_0^*, y_0^*, z_0^*}^{**}] - H[X_t|\rho_{x_0^*, y_0^*, z_0^*}^*]$. The assemblage is the average over (x_0^*, y_0^*, z_0^*) with $\rho(x_0^*, y_0^*, z_0^*)$ similarly to $I_{Y \rightarrow X}^{(t)}$, but with an additional variable z_0^* . Thereby, one gets the three-subsystem extended TE $I_{Y \rightarrow X|Z}^{(t)}$ as

$$I_{Y \rightarrow X|Z}^{(t)} = \langle H[X_t|\rho_{x_0^*, y_0^*, z_0^*}^{**}] - H[X_t|\rho_{x_0^*, y_0^*, z_0^*}^*] \rangle_{\rho(x_0^*, y_0^*, z_0^*)}. \quad (\text{F1})$$

If one sets $\rho = p_{XYZ}^{st}$, Eq. (F1) gives the three-subsystems TE $T_{Y \rightarrow X|Z}^{(t)}$. Selecting $t = 1$, one gets the quantity $T_{Y \rightarrow X|Z}^{(1)}$ which exactly coincides with the causation entropy suggested by Sun and Bollt and given by Eq. (32) in Ref. [30] where

Z and Y are interchanged as compared to Eq. (F1) here. To see that coincidence, just notice that Eq. (32) uses Eq. (10) in Ref. [30].

If the knowledge about an SDS involves other factors beyond initial states of X and Y , one can include it into the generalized initial conditions of the DCE in a formal way similar to how it is done for the third subsystem. Such additional factors can be (1) an influence of a temporal segment $Y_{(0,t)}$ over an interval $(0, t)$ on X_t relevant for a unidirectional coupling $Y \rightarrow X$; (2) an influence of a temporal segment $(X_{(-\Delta,0)}, Y_{(-\Delta,0)})$ (i.e., a longer past) on X_t relevant for time-delayed systems and couplings; and (3) an influence of a noise realization segment $\xi_{y,(-\Delta,0)}$ on X_t relevant to describe separate influences of “unique events” in the subsystem Y at different past instants as formalized with momentary information transfer [22] which is a fruitful and actively used causality quantifier [66]. These considerations will readily apply to studying couplings in larger ensembles of dynamical systems which problem is the subject of a huge number of works, e.g., [6,40,66,148,149].

2. Inverse problems

a. Estimation

Estimation of DCEs from time series is a practically important issue when one must answer the question “How to compute?” It should be studied as an *inverse problem* separately from the *direct problem* of defining a DCE for a given SDS which addresses the question “What to compute?” The estimation essentially uses the solution to the direct problem, since one must clearly understand “what to compute” before “how to compute.” This work is devoted to the former question, while the latter one is briefly commented on below.

If an SDS under study is indeed fully available, one can perform sufficiently many experiments starting from different initial states and parameters, observing future responses, computing distinctions, and assembling elementary DCEs. The difference from the theoretical computation of DCE (4) is only in a finite number of experiments contrary to infinite number of experiments implied by the expectations typically used in the definition (4).

However, one often has only a single time series for a given parameter value a without possibilities to perform desired experiments (interventions). Then quite strict conditions must be satisfied to apply the definition of a DCE directly: (i) an SDS under study is ergodic, (ii) a DCE of interest involves only initial condition variations, (iii) full state vectors (x_t, y_t) of the SDS are observed, and (iv) the observed time realization is long enough, i.e., it returns (within a small enough distance) to any state (x_0^*, y_0^*) relevant to determine the DCE of interest sufficiently many times separated by significant time intervals to assure that the evolutions after each return are mutually independent. Then one creates the ensembles ρ_{XY}^* and ρ_{XY}^{**} from the observed states and compares the respective futures.

Any of the four conditions may be violated. As a violation of (iv), not all relevant states may be visited. Then one should have an opportunity to perform interventions and impose initial states absent from (or not well represented in) a time series at hand. Then the observed responses to such interventions

together with the original time series can be used to estimate a DCE according to its general definition. Such interventions are used, e.g., in Ref. [19] to estimate phase dynamics-based quantifiers of causal couplings and in dynamical causal modeling [57] to estimate causal coupling coefficients.

As a violation of (iii), a parameter variation may be involved in a DCE while a time series is recorded at a single parameter value. Then one should have an opportunity to perform a parameter intervention imposing an alternative parameter value, observe an SDS response, and use again the DCE definition (4). Such time series recorded at different parameter values were used in Ref. [6] to reconstruct causal structure of an ensemble of phase oscillators from stationary phase shifts between oscillators observed at different frequency mismatches.

As a violation of (i), an SDS may not be ergodic. Then a single time series is generally insufficient to estimate a DCE. A direct solution of this problem requires an opportunity to perform interventions and impose necessary initial states to create ensembles ρ_{XY}^* and ρ_{XY}^{**} (approximately) and observe responses of the SDS under study.

As a violation of (ii) and (iii), only some components of a state vector may be observed (so an observed process is non-Markovian) or one may not possess a time series for the alternative parameter value. Then the DCE is not directly estimable, but two indirect ways are possible. First, for a DCE of a parameter variation (e.g., $S_{Y \rightarrow X}$), its relations to more available quantifiers (e.g., $\tau_{Y \rightarrow X}$ and $l_{Y \rightarrow X}$) for a class of SDS (as in Sec. III D) can be used. Second, a parametrized model can be identified (e.g., [101,102]) with the subsequent use of the definition (4). In more detail, one assumes a state space model (a model SDS) which includes full states and parameters and constructs such a model from a time series, e.g., [101,102]. Then the DCE is defined for the obtained model via the definition (4). The model-based approach is universal and can be applied in the case of any of the above violations. However, the causal information is encoded in the model structure (a parametrized set of models [101]) selected for the identification. If the model structure is not adequate, the causality quantifier may not be a meaningful characteristic, i.e., any nonzero causality quantifier (4) is not *per se* a reliable basis for *causal discovery*.

In agreement with this perspective, Stokes and Purdon [45] suggest to focus on model identification and criticize the causality spectra for their disagreement with “intuitive notion of causality,” which is further discussed in [49]. Identification of coupled systems is used in the dynamical causal modeling [57] and the coupling function analysis [44] for concrete research purposes.

If a system under study is not a Markovian RDS, but a general RDS, then the interventional and passively observed PDFs $p(X_t | \text{do}(x_0, y_0))$ and $p(X_t | x_0, y_0)$ generally differ due to different PDFs of hidden factors (confounders). To define a DCE, one should either perform real interventions or to specify a parametrized model for a full system under study either in the form of an SDS with a higher-dimensional full state vector (reducing this case to the previous ones) or in the form of an SDS coupled to some non-Markovian process (see also Sec. III in Ref. [90]). In practice, the latter implies further assumptions about a system under study.

Diverse DCEs differ from each other in their availability. Easier estimable DCEs have practical advantages. Knowing relations between DCEs, one can also compute a desired DCE (e.g., an asymptotic DCE of a parameter variation $S_{Y \rightarrow X}$) from an estimate of a more available DCE (e.g., a short-term DCE of initial condition variations $\tau_{Y \rightarrow X}$) for a certain class of SDS, e.g., linear overdamped stochastic oscillators (1). Concrete ways to compute DCE estimates may strongly differ depending on the class of SDS under study, a DCE of interest, and data at hand. Many techniques have already been suggested for estimation of the TE and other quantifiers; see, e.g., Refs. [1–41]. Others may well represent an avenue for a further research within the DCE framework.

b. Causal discovery

Since a full SDS is known *a priori* in all above considerations, causal discovery has not been needed. In a more general setting where the causal structure must be *discovered* from data, one needs the full machinery of the SCM techniques, which is a focus of many works, e.g., Refs. [27,40,119–122]. This issue belongs to the field of inverse problems and identifiability.

The result of causal discovery relevant for the problems of causality quantification considered here is a plausible SDS

describing processes under study. Such an SDS may also be formulated from some conceptual considerations. Anyway, such a model (inferred) SDS can then be used to compute a desired DCE directly from the general definition (4). Since such an SDS is itself only a model (approximation) for a system under study, an obtained DCE value is an estimate of a desired DCE.

If one does not desire to formulate explicitly any SDS as a basic model, then possibilities to correctly and precisely define DCEs are strongly limited and anyway depend on the assumptions about the system under study. If such assumptions remain implicit, then no causal statements can be justified: no quantifier can itself be a sufficient basis for causal statements, it must anyway rely on an *explicitly* stated SDS, exactly as such causal statements are required to rely explicitly on a structural causal model in Ref. [42]. Indeed, Pearl [42] stresses that it is the very syntax of a structural causal model (i.e., the fact that an author declares that model as “structural”) that encodes its causal content, but not the data analysis which maintains model assumptions implicit. In the field of processes, an SDS *is* a general structural causal model encoding the causal content due to the arrow of time, i.e., due to the “flow of causation from past into future” [103].

-
- [1] T. Schreiber, *Phys. Rev. Lett.* **85**, 461 (2000).
 - [2] M. Palus, V. Komarek, Z. Hrnčir, and K. Sterbova, *Phys. Rev. E* **63**, 046211 (2001).
 - [3] M. G. Rosenblum and A. S. Pikovsky, *Phys. Rev. E* **64**, 045202(R) (2001).
 - [4] D. A. Smirnov and B. P. Bezruchko, *Phys. Rev. E* **68**, 046209 (2003).
 - [5] X. S. Liang and R. Kleeman, *Phys. Rev. Lett.* **95**, 244101 (2005).
 - [6] M. Timme, *Phys. Rev. Lett.* **98**, 224101 (2007)
 - [7] I. T. Tokuda, S. Jain, I. Z. Kiss, and J. L. Hudson, *Phys. Rev. Lett.* **99**, 064101 (2007).
 - [8] K. Hlavackova-Schindler, M. Palus, M. Vejmelka, and J. Bhattacharya, *Phys. Rep.* **441**, 1 (2007).
 - [9] A. Bahraminasab, F. Ghasemi, A. Stefanovska, P. V. E. McClintock, and H. Kantz, *Phys. Rev. Lett.* **100**, 084101 (2008).
 - [10] M. Dhamala, G. Rangarajan, and M. Ding, *Phys. Rev. Lett.* **100**, 018701 (2008).
 - [11] D. Marinazzo, M. Pellicoro, and S. Stramaglia, *Phys. Rev. Lett.* **100**, 144103 (2008).
 - [12] M. Staniek and K. Lehnertz, *Phys. Rev. Lett.* **100**, 158101 (2008).
 - [13] J. T. Lizier, M. Prokopenko, and A. Y. Zomaya, *Phys. Rev. E* **77**, 026110 (2008).
 - [14] M. Vejmelka and M. Palus, *Phys. Rev. E* **77**, 026214 (2008).
 - [15] L. Barnett, A. B. Barrett, and A. K. Seth, *Phys. Rev. Lett.* **103**, 238701 (2009).
 - [16] D. Chicharro and R. G. Andrzejak, *Phys. Rev. E* **80**, 026217 (2009).
 - [17] J. T. Lizier and M. Prokopenko, *Eur. Phys. J. B* **73**, 605 (2010).
 - [18] D. W. Hahs and S. D. Pethel, *Phys. Rev. Lett.* **107**, 128701 (2011).
 - [19] Z. Levnajic and A. Pikovsky, *Phys. Rev. Lett.* **107**, 034101 (2011).
 - [20] G. Sugihara, R. May, H. Ye, C. Hsieh, E. Deyle, M. Fogarty, and S. Munch, *Science* **338**, 496 (2012).
 - [21] L. Barnett and T. Bossomaier, *Phys. Rev. Lett.* **109**, 138105 (2012).
 - [22] J. Runge, J. Heitzig, V. Petoukhov, and J. Kurths, *Phys. Rev. Lett.* **108**, 258701 (2012).
 - [23] J. Runge, J. Heitzig, N. Marwan, and J. Kurths, *Phys. Rev. E* **86**, 061121 (2012).
 - [24] T. Stankovski, A. Duggento, P. V. E. McClintock, and A. Stefanovska, *Phys. Rev. Lett.* **109**, 024101 (2012).
 - [25] L. Barnett, J. T. Lizier, M. Harre, A. K. Seth, and T. Bossomaier, *Phys. Rev. Lett.* **111**, 177203 (2013).
 - [26] D. A. Smirnov, *Phys. Rev. E* **87**, 042917 (2013).
 - [27] D. Janzing, D. Balduzzi, M. Grosse-Wentrup, and B. Schoelkopf, *Ann. Stat.* **41**, 2324 (2013).
 - [28] M. Palus, *Phys. Rev. Lett.* **112**, 078702 (2014).
 - [29] D. A. Smirnov, *Phys. Rev. E* **90**, 062921 (2014).
 - [30] J. Sun and E. M. Bollt, *Physica D* **267**, 49 (2014).
 - [31] J. Runge, *Phys. Rev. E* **92**, 062829 (2015).
 - [32] R. G. James, N. Barnett, and J. P. Crutchfield, *Phys. Rev. Lett.* **116**, 238701 (2016).
 - [33] D. Harnack, E. Laminski, M. Schunemann, and K. R. Pawelzik, *Phys. Rev. Lett.* **119**, 098301 (2017).
 - [34] M. Nitzan, J. Casadiego, and M. Timme, *Sci. Adv.* **3**, e1600396 (2017).
 - [35] P. Laiou and R. G. Andrzejak, *Phys. Rev. E* **95**, 012210 (2017).
 - [36] L. Faes, G. Nollo, S. Stramaglia, and D. Marinazzo, *Phys. Rev. E* **96**, 042150 (2017).

- [37] H. Ashikaga and R. G. James, *Chaos* **28**, 075306 (2018).
- [38] E. M. Bollt, *Chaos* **28**, 075309 (2018).
- [39] X. S. Liang, *Chaos* **28**, 075311 (2018).
- [40] J. Runge, S. Bathiany, E. Bollt, G. Camps-Valls, D. Coumou, E. Deyle, C. Glymour, M. Kretschmer, M. D. Mahecha, J. Munoz-Mari, E. H. van Nes, J. Peters, R. Quax, M. Reichstein, M. Scheffer, B. Scholkopf, P. Spirtes, G. Sugihara, J. Sun, K. Zhang *et al.*, *Nat. Commun.* **10**, 2553 (2019).
- [41] J. Runge, P. Nowack, M. Kretschmer, S. Flaxman, and D. Sejdinovic, *Sci. Adv.* **5**, eaau4996 (2019).
- [42] J. Pearl, *Causality: Models, Reasoning, and Inference* (Cambridge University Press, Cambridge, 2000).
- [43] P. Spirtes, C. Glymour, and R. Scheines, *Causation, Prediction, and Search* (MIT Press, Cambridge, MA, 2000).
- [44] T. Stankovski, T. Pereira, P. V. E. McClintock, and A. Stefanovska, *Rev. Mod. Phys.* **89**, 045001 (2017).
- [45] P. A. Stokes and P. L. Purdon, *Proc. Natl. Acad. Sci. USA* **114**, 7063 (2017).
- [46] M. Dhamala, H. Liang, S. L. Bressler, and M. Ding, *NeuroImage* **175**, 460 (2018).
- [47] L. Barnett, A. B. Barrett, and A. K. Seth, *NeuroImage* **178**, 744 (2018).
- [48] P. A. Stokes and P. L. Purdon, *Proc. Natl. Acad. Sci. USA* **115**, E6678 (2018).
- [49] D. A. Smirnov, *Europhys. Lett.* **128**, 20006 (2019).
- [50] T. Bossomaier, L. Barnett, M. Harre, and J. T. Lizier, *An Introduction to Transfer Entropy. Information Flow in Complex Systems* (Springer, Cham, 2016).
- [51] N. Wiener, in *Modern Mathematics for Engineers*, edited by E. F. Beckenbach (McGraw-Hill, New York, 1956), pp. 125–139.
- [52] C. W. J. Granger, *Inf. Control* **6**, 28 (1963).
- [53] M. Prokopenko and J. T. Lizier, *Sci. Rep.* **4**, 5394 (2014).
- [54] D. Chionis, A. Dokhane, H. Ferroukhi, and A. Pautz, *Chaos* **29**, 043126 (2019).
- [55] Y.-C. Hung and C.-K. Hu, *Phys. Rev. Lett.* **101**, 244102 (2008).
- [56] A. Tsonis, E. R. Deyle, R. M. May, G. Sugihara, K. Swanson, J. D. Verbeten, and G. Wang, *Proc. Natl. Acad. Sci. USA* **112**, 3253 (2015).
- [57] K. J. Friston, L. Harrison, and W. Penny, *NeuroImage* **19**, 1273 (2003).
- [58] A. Brovelli, M. Ding, A. Ledberg, Y. Chen, R. Nakamura, and S. L. Bressler, *Proc. Natl. Acad. Sci. USA* **101**, 9849 (2004).
- [59] B. P. Bezruchko, V. I. Ponomarenko, M. D. Prokhorov, D. A. Smirnov, and P. A. Tass, *Phys. Usp.* **51**, 304 (2008).
- [60] B. Schelter, J. Timmer, and M. Eichler, *J. Neurosci. Methods* **179**, 121 (2009).
- [61] S. L. Bressler and A. K. Seth, *NeuroImage* **58**, 323 (2011).
- [62] M. Wibral, R. Vicente, and J. T. Lizier (Eds.), *Directed Information Measures in Neuroscience* (Springer-Verlag, Berlin, 2014).
- [63] I. I. Mikhov and D. A. Smirnov, *Geophys. Res. Lett.* **33**, L03708 (2006).
- [64] I. I. Mikhov, D. A. Smirnov, P. I. Nakonechny, S. S. Kozlenko, Ye. P. Seleznev, and J. Kurths, *Geophys. Res. Lett.* **38**, L00F04 (2011).
- [65] A. Attanasio, A. Pasini, and U. Triacca, *Atmos. Climate Sci.* **3**, 515 (2013).
- [66] J. Runge, V. Petoukhov, J. F. Donges, J. Hlinka, N. Jajcay, M. Vejmelka, D. Hartman, N. Marwan, M. Palus, and J. Kurths, *Nat. Commun.* **6**, 8502 (2015).
- [67] A. Stips, D. Macias, C. Coughlan, E. Garcia-Gorriz, and X. S. Liang, *Sci. Rep.* **6**, 21691 (2016).
- [68] D. A. Smirnov, S. F. M. Breitenbach, G. Feulner, F. A. Lechleitner, K. M. Prufer, J. U. L. Baldini, N. Marwan, and J. Kurths, *Sci. Rep.* **7**, 11131 (2017).
- [69] M. Palus, A. Krakovska, J. Jakubik, and M. Chvostekova, *Chaos* **28**, 075307 (2018).
- [70] P. Nowack, J. Runge, V. Eyring, and J. D. Haigh, *Nat. Commun.* **11**, 1415 (2020).
- [71] H. Xiao, F. Zhang, L. Miao, X. S. Liang, K. Wu, and R. Liu, *Clim. Dyn.* **55**, 1443 (2020).
- [72] G. Wang, C. Zhao, M. Zhang, Y. Zhang, M. Lin, and F. Qiao, *Sci. Rep.* **10**, 17141 (2020).
- [73] X. S. Liang and R. Kleeman, *Physica D* **231**, 1 (2007).
- [74] X. S. Liang and R. Kleeman, *Physica D* **227**, 173 (2007).
- [75] X. S. Liang, *Phys. Rev. E* **78**, 031113 (2008).
- [76] X. S. Liang, *Phys. Rev. E* **90**, 052150 (2014).
- [77] X. S. Liang, *Phys. Rev. E* **92**, 022126 (2015).
- [78] X. S. Liang, *Phys. Rev. E* **94**, 052201 (2016).
- [79] D. A. Smirnov and I. I. Mikhov, *Phys. Rev. E* **92**, 042138 (2015).
- [80] D. A. Smirnov, *Chaos* **28**, 075303 (2018).
- [81] D. A. Smirnov, *Phys. Rev. E* **102**, 062139 (2020).
- [82] D. A. Smirnov, *Chaos* **31**, 073127 (2021).
- [83] L. Arnold, *Random Dynamical Systems* (Springer-Verlag, Berlin, 1998).
- [84] A. A. Andornov, A. A. Vitt, and S. E. Khaikin, *Teoriya kolebaniy*, 2nd ed. with N. A. Zheleztsov's supplements (Fizmatlit, Moscow, 1959). *Theory of Oscillators*, translated by F. Immirzi (Pergamon, Oxford, 1966).
- [85] J. Guckenheimer and P. Holmes, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields* (Springer-Verlag, Berlin, 1983).
- [86] A. Katok and B. Hasselblatt, *Introduction to the Modern Theory of Dynamical Systems* (Cambridge University Press, Cambridge, 1995).
- [87] S. M. Rytov, *Introduction to Statistical Radiophysics. Vol. 1. Stochastic Processes* (Fizmatlit, Moscow, 1976).
- [88] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1981).
- [89] J. Honerkamp, *Stochastic Dynamical Systems: Concepts, Numerical Methods, Data Analysis* (John Wiley and Sons, New York, 1994).
- [90] Supplemental material at <http://link.aps.org/supplemental/10.1103/PhysRevE.105.034209> provides technical details on causal graphs and SCM characteristics related to the DCE formalism, a table containing 23 causality quantifiers with all DCE elements indicated, and possible mathematical extensions of the suggested formalism.
- [91] K. Hasselmann, *Tellus* **28**, 473 (1976).
- [92] K. Sekimoto, *Stochastic Energetics*, Lectures Notes in Physics Vol. 799 (Springer, Berlin, 2010).
- [93] P. W. Holland, in *Sociological Methodology*, edited by C. Clogg (American Sociological Association, Washington, DC, 1988), p. 449.
- [94] M. F. Jansen, D. Dommengot, and N. Keenlyside, *J. Climate* **22**, 550 (2009).

- [95] N. Ay and D. Polani, *Adv. Complex Syst.* **11**, 17 (2008).
- [96] A. S. Pikovsky, M. G. Rosenblum, and J. Kurths, *Synchronization: A Universal Concept in Nonlinear Sciences* (Cambridge University Press, Cambridge, 2001).
- [97] A. Kolmogoroff, *Math. Ann.* **104**, 415 (1931).
- [98] H. Risken, *The Fokker-Planck Equation: Methods of Solution, Applications* (Springer-Verlag, Berlin, 1984).
- [99] J. Geweke, *J. Am. Stat. Assoc.* **77**, 304 (1982).
- [100] J. Geweke, *J. Am. Stat. Assoc.* **79**, 907 (1984).
- [101] L. Ljung, *System Identification: Theory for the User* (Prentice-Hall, Englewood Cliffs, NJ, 1987).
- [102] B. P. Bezruchko and D. A. Smirnov, *Extracting Knowledge from Time Series. An Introduction to Nonlinear Empirical Modeling* (Springer-Verlag, Berlin, 2010).
- [103] R. E. Kalman, P. L. Falb, and M. A. Arbib, *Topics in Mathematical System Theory* (McGraw-Hill, New York, 1969).
- [104] J. Brea, D. F. Russell, and A. B. Neiman, *Chaos* **16**, 026111 (2006).
- [105] D. Smirnov, U. B. Barnikol, T. T. Barnikol, B. P. Bezruchko, C. Hauptmann, C. Buehrle, M. Maarouf, V. Sturm, H.-J. Freund, and P. A. Tass, *Europhys. Lett.* **83**, 20003 (2008).
- [106] P. Tass, D. Smirnov, A. Karavaev, U. Barnikol, T. Barnikol, I. Adamchic, C. Hauptmann, N. Pawelczyk, M. Maarouf, V. Sturm, H.-J. Freund, and B. Bezruchko, *J. Neural Eng.* **7**, 016009 (2010).
- [107] B. Kralemann, M. Rosenblum, and A. Pikovsky, *Chaos* **21**, 025104 (2011).
- [108] D. A. Smirnov and B. P. Bezruchko, *Phys. Rev. E* **79**, 046204 (2009).
- [109] L. I. Mandelshtam, *Complete Works*, Vol. 4, *Lektsii po kolebaniyam (Lectures on Oscillations)* (Izd. Acad. Nauk USSR, Moscow, 1955).
- [110] G. S. Gorelik, *Kolebaniya i volny (Oscillations and Waves)*, 2nd ed. (Fizmatlit, Moscow, 1959).
- [111] A. Pechenkin, *Leonid Isaakovich Mandelstam* (Springer, Cham, 2014), pp. 135–149.
- [112] M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics*, Vol. 2, *Inference and Relationship*, 2nd ed. (Charles Griffin and Co., London, 1967).
- [113] G. A. F. Seber, *Linear Regression Analysis* (John Wiley and Sons, New York, 1977).
- [114] S. Wright, *J. Agric. Res.* **20**, 557 (1921)
- [115] T. Haavelmo, *Econometrica* **11**, 1 (1943).
- [116] J. Neyman, “Sur les applications de la tour des probabilités aux expériences agraires: Essays des principe” (1923). Excerpts reprinted in English (D. Dabrowska and T. Speed, translators), *Stat. Sci.* **5**, 463 (1990).
- [117] D. B. Rubin, *J. Educ. Psychol.* **66**, 688 (1974).
- [118] T. J. Vanderweele, *Explanation in Causal Inference* (Oxford University Press, Oxford, 2015).
- [119] D. Dash, in *Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics—AISTATS 2005*, edited by R. G. Cowell and Z. Ghahramani (Society for Artificial Intelligence and Statistics, London, 2005), p. 81.
- [120] M. Voortman, D. Dash, and M. J. Druzdzel, in *JMLR Workshop and Conference Proceedings (NIPS 2008 Workshop on Causality)*, edited by I. Guyon, D. Janzing, and B. Schoelkopf (PMLR, 2010), Vol. 6, pp. 257–266.
- [121] M. Eichler, in *Causality: Statistical Perspectives and Applications*, edited by C. Berzuini, A. P. Dawid, and L. Bernardinelli (Wiley, Chichester, 2012), p. 327.
- [122] M. Eichler, R. Dahlhaus, and J. Dueck, *J. Time Ser. Anal.* **38**, 225 (2017).
- [123] M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics*, Vol. 3, *Deqian and Analysis, and Time Series*, 2nd ed. (Charles Griffin and Co., London, 1968).
- [124] Yu. I. Rozanov, *Stationary Random Processes* (Holden-Day, San Francisco, 1967).
- [125] I. I. Gihman and A. V. Skorohod, *The Theory of Stochastic Processes* (Springer, Berlin, 1975).
- [126] A. M. Yaglom, Introduction to the theory of stationary random functions, *Usp. Mat. Nauk* **7**, 3 (1952) English edition (R. A. Silverman, translator) (Prentice-Hall, Englewood Cliffs, NJ, 1962).
- [127] G. M. Jenkins and D. G. Watts, *Spectral Analysis and Its Applications* (Holden-Day, San Francisco, 1968).
- [128] M. J. Kaminski and K. J. Blinowska, *Biol. Cybern.* **65**, 203 (1991).
- [129] L. A. Baccala and K. Sameshima, *Biol. Cybern.* **84**, 463 (2001).
- [130] C. W. J. Granger, *J. Econ. Dyn. Control* **2**, 329 (1980).
- [131] P. Biller, J. Honerkamp, and F. Petruccione, *Stochastic Dynamical Systems* (Istituto Italiano per gli Studi Filosofici, Naples, 1991).
- [132] A. Barrett and L. Barnett, *Front. Neuroinform.* **7**, 6 (2013).
- [133] A. Hannart, J. Pearl, F. Otto, P. Naveau, and M. Ghil, *Bull. Am. Meteorol. Soc.* **97**, 99 (2016).
- [134] M. Baldovin, F. Cecconi, and A. Vulpiani, *Phys. Rev. Research* **2**, 043436 (2020).
- [135] A. Auconi, B. M. Friedrich, and A. Giansanti, *Europhys. Lett.* **135**, 28002 (2021).
- [136] G. E. P. Box and G. M. Jenkins, *Time Series Analysis. Forecasting and Control* (Holden-Day, San Francisco, 1970).
- [137] A. N. Shiryaev, *Probability* (Springer-Verlag, Berlin, 1996).
- [138] A. A. Borovkov, *Probability Theory* (Springer-Verlag, London, 2013).
- [139] M. Loeve, *Probability Theory*, 4th ed. (Springer-Verlag, New York, 1991).
- [140] V. S. Pugachev, *Theory of Probabilities and Mathematical Statistics*, 2nd ed. (Fizmatlit, Moscow, 2002).
- [141] V. S. Pugachev and I. N. Sinityn, *Stochastic Differential Systems. Analysis and Filtering* (Wiley, Chichester, 1987).
- [142] Y. I. Molokov, E. M. Loskutov, D. N. Mukhin, and A. M. Feigin, *Phys. Rev. E* **85**, 036216 (2012).
- [143] D. Mukhin, E. Loskutov, A. Mukhina, A. Feigin, I. Zaliapin, and M. Ghil, *J. Climate* **28**, 1940 (2015).
- [144] A. B. Barrett, L. Barnett, and A. K. Seth, *Phys. Rev. E* **81**, 041907 (2010).
- [145] V. I. Tikhonov and M. A. Mironov, *Markov Processes* (Sovetskoye Radio, Moscow, 1977).
- [146] Y. Chen, G. Rangarajan, J. Feng, and M. Ding, *Phys. Lett. A* **324**, 26 (2004).
- [147] L. Barnett and A. K. Seth, *Phys. Rev. E* **91**, 040101(R) (2015).
- [148] I. V. Sysoev, M. D. Prokhorov, V. I. Ponomarenko, and B. P. Bezruchko, *Phys. Rev. E* **89**, 062911 (2014).
- [149] I. V. Sysoev, V. I. Ponomarenko, and A. Pikovsky, *Commun. Nonlinear Sci. Numer. Simul.* **57**, 342 (2018).