Chaos
An Interdisciplinary Journal of Nonlinear Science

RESEARCH ARTICLE | MARCH 19 2024

Information transfers and flows in Markov chains as dynamical causal effects

Dmitry A. Smirnov

Chaos 34, 033130 (2024)
https://doi.org/10.1063/5.0189544

Chaos ARTICLE pubs.aip.org/aip/cha

Information transfers and flows in Markov chains as dynamical causal effects

Cite as: Chaos 34, 033130 (2024); doi: 10.1063/5.0189544
Submitted: 29 November 2023 · Accepted: 22 February 2024 ·
Published Online: 19 March 2024

View Online Export Citation CrossMark

Dmitry A. Smirnov[1,2,a)]

AFFILIATIONS

[1] Institute of Physics, Saratov State University, 155 Moskovskaya Street, Saratov 410012, Russia
[2] Saratov Branch, Kotelnikov Institute of RadioEngineering and Electronics of the Russian Academy of Sciences, 38 Zelyonaya Street, Saratov 410019, Russia

[a)] **Author to whom correspondence should be addressed:** smirnovda@yandex.ru

ABSTRACT

A logical sequence of information-theoretic quantifiers of directional (causal) couplings in Markov chains is generated within the framework of dynamical causal effects (DCEs), starting from the simplest DCEs (in terms of localization of their functional elements) and proceeding step-by-step to more complex ones. Thereby, a system of 11 quantifiers is readily obtained, some of them coinciding with previously known causality measures widely used in time series analysis and often called "information transfers" or "flows" (transfer entropy, Ay–Polani information flow, Liang–Kleeman information flow, information response, etc.,) By construction, this step-by-step generation reveals logical relationships between all these quantifiers as specific DCEs. As a further concretization, diverse quantitative relationships between the transfer entropy and the Liang–Kleeman information flow are found both rigorously and numerically for coupled two-state Markov chains.

There are dozens of approaches for quantification of causal couplings between observed processes since it is an important problem in many scientific disciplines and applications, from nuclear reactors[1] and galactic cosmic rays[2] to climate[3] and environment[4] and so forth. In particular, researchers widely use information-theoretic measures,[5–22] spectral causalities,[23–27] phase-dynamic modeling,[28–33] nonlinear Granger causality,[34–38] cross-recurrent diagrams,[39,40] convergent cross-mapping,[41,42] etc. The diverse causality quantifiers have been (and still are) developed from seemingly independent ideas. Numerical values of different quantifiers of the same coupling may differ drastically, while an appropriate substantial interpretation of any quantifier is often not obvious or even problematic. Therefore, before deciding "how to estimate a causal coupling," one must often realize "which quantifier should be estimated and what it means." The recently proposed framework of dynamical causal effects[43,44] (DCEs) allows one to deduce various quantifiers from a single theoretical principle and formulate them in the same language. So, it provides a regular approach to their interpretation which is useful when a researcher selects between many equally reasonable quantifiers as it may be the case with information-theoretic causality measures. This work demonstrates how the DCE framework brings an order to such multitude of measures: 11 information-theoretic causality quantifiers are deduced step-by-step, and at least five known "information transfers and flows" are encountered among them. Logical relationships between all those quantifiers follow directly from this way of their generation as DCEs. Analytic and numerical relationships between the two most famous of them (transfer entropy and Liang–Kleeman information flow) in a simple stochastic system are obtained as an example. As a perspective, such results should be a contribution to a unifying theory of causality quantifiers for processes.

I. INTRODUCTION

Within the broad field of the detection of causal couplings between processes and estimation of their strengths (e.g., reviews[3,4,8,14,18,27]), some recent works have been devoted to comparisons of various causality quantifiers in different respects (e.g., Refs. 8 and 45–48). This is because many causality quantifiers have already been suggested and widely used, some of them being more easily estimable from time series while others are less accessible. Moreover, difficulties in their interpretation often arise and become a subject of debates (e.g., Refs. 15, 19, and 49–53). Thus, a meaningful organization of the multitude of causality quantifiers for processes seems to have become a topical problem.

The viewpoint of dynamical causal effects (DCEs) has recently been introduced[40,41] and illustrated[53–58] as a useful tool to generate various causality quantifiers from a single first principle (rooted in the concepts of stochastic dynamical system[59,60] and interventional causality)[61] and so to establish logical links and numerical relationships between them. Such systematic links should allow one to "navigate" within the multitude of causality quantifiers and hopefully (after future research) become a tool to select the most appropriate quantifier for a practical problem at hand. However, the previously presented relationships between several popular causality measures within the DCE framework[43,44,53,54] do not reveal how a whole line of meaningful quantifiers can be obtained and so the potential of this framework is not yet fully confirmed. One can expect to encounter such a logical line when dealing with information-theoretic coupling measures, since more than a dozen of such measures (quite different from each other)[5–24] are used in time series analysis, several of them being often called "information flow" (IF) or "information transfer." In particular, the two very popular IFs are the transfer entropy (TE)[5,14] and the Liang–Kleeman IF (LKIF).[7,20] Such a rich and apparently disordered set of different causality quantifiers with often the same name and similar formulas is a challenge for the DCE viewpoint and formalism: Can the latter recognize various information-theoretic measures as specific DCEs and organize them into a logical sequence? This is the question addressed in this work. If the answer is "yes," it would thereby contribute to the development of an intrinsically interrelated system of causality quantifiers and simplify the above mentioned navigation in a large set of existing measures, apparently independent of each other.

Section II describes Markov chains which serve as a "material" to study various IFs since they seem to be conceptually the simplest stochastic dynamical systems. Section III introduces the formalism of DCEs in a concrete form and describes the logics of moving from simpler to more complex DCEs which is the method to study the IFs here. Section IV presents the main part of this work which is a step-by-step development of a sequence of causality quantifiers which includes various above mentioned IFs as specific DCEs. Section V presents both analytic and numerical relationships between the TE and the LKIF. Relations between some information-theoretic DCEs were also given in Ref. 44 for continuous-time and continuous-state systems as an illustration to the DCE formalism. Here, a much fuller picture is obtained via developing a whole logical line of information-theoretic DCEs and finding the place of several known quantifiers in a wider context. Moreover, the discrete-time and discrete-state systems used in this work have their own specific features simplifying the formulation and numerical study of diverse DCEs, while the LKIF has not yet been studied for such systems. Section VI provides a brief discussion and conclusions.

II. COUPLED MARKOV CHAINS

Let two random variables X_n and Y_n characterize the states of two systems **X** and **Y** at a discrete time n. A realization x_n of X_n is some value from a discrete set $A = \{a_1, \ldots, a_M\}$ where a_i are any symbols, and similarly y_n belongs to a set $B = \{b_1, b_2, \ldots, b_N\}$. The random variable $Z_n = (X_n, Y_n)$ with the values $z_n = (x_n, y_n)$ from the set $A \times B$ specifies the state of the combined system **Z**. Denote $p_{XY}^{(n)}(x, y)$ the probability mass function (pmf) of Z_n. The probabilities $p_{XY}^{(n)}\left(a_i, b_j\right) = \mathrm{P}\{X_n = a_i, Y_n = b_j\}$ form the MN-dimensional vector,

$$\mathbf{p}_{XY}^{(n)} = \Big(p_{XY}^{(n)}(a_1, b_1), p_{XY}^{(n)}(a_1, b_2), \ldots, p_{XY}^{(n)}(a_1, b_N), \\ p_{XY}^{(n)}(a_2, b_1), \ldots, p_{XY}^{(n)}(a_M, b_N)\Big)^{\mathrm{T}}, \tag{1}$$

whose elements are non-negative with unit sum and T means transposition. Let the random process Z_n be a first-order Markov chain,[59,62] i.e., the value $z_n = \left(a_i, b_j\right)$ at time n is produced with probability $p_{XY}^{(n)}\left(a_i, b_j\right)$ depending only on the immediate past pmf $p_{XY}^{(n-1)}(x, y)$ of Z_{n-1} as

$$\mathbf{p}_{XY}^{(n)} = \mathbf{Q}\mathbf{p}_{XY}^{(n-1)}, \tag{2}$$

where $\mathbf{Q}$ is the $MN \times MN$ transition probability matrix with constant elements which are the conditional probabilities $q_{i',j'}^{i,j} = P\left\{X_n = a_i, Y_n = b_j | X_{n-1} = a_{i'}, Y_{n-1} = b_{j'}\right\}$, i.e., the probabilities of transitions from a state $(a_{i'}, b_{j'})$ to a state (a_i, b_j) in one time step, located at (i, j)-th row and (i', j')-th column where the values of the double indices (i, j) and (i', j') are ordered as shown in Eq. (1).

The one-step future pmf $p_{XY}^{(1)}(x, y)$ under the initial condition $p_{XY}^{(0)}$ is called the "functionally conditional" pmf in Ref. 44 and denoted $p_{XY}^{(1)}\left[x, y|p_{XY}^{(0)}\right]$ using the square brackets with this special meaning. Denote an initial pmf $p_{XY}^{(0)}(x, y)$ localized at $z_0^* = \left(a_{i*}, b_{j*}\right)$ as $p_{XY}^{(0)}\left(a_i, b_j\right) = \delta_{i,i*}\delta_{j,j*}$ where δ is the Kronecker delta which equals 1 for equal indices in the subscript and 0 otherwise. The values of the pmf $p_{XY}^{(1)}\left[x, y|\delta_{i,i*}\delta_{j,j*}\right]$ for the localized $p_{XY}^{(0)} = \delta_{i,i*}\delta_{j,j*}$ constitute the (i^*, j^*)-th column of $\mathbf{Q}$, i.e., $\mathbf{p}_{XY}^{(1)} = \left(q_{i*,j*}^{1,1}, q_{i*,j*}^{1,2}, \ldots, q_{i*,j*}^{M,N}\right)^{\mathrm{T}}$. Under some general conditions on the matrix $\mathbf{Q}$, the process Z_n possesses a unique stationary pmf $\rho_{XY}^{st}(x, y)$, i.e., an ensemble represented by any initial pmf $p_{XY}^{(0)}(x, y)$ evolves to ρ_{XY}^{st}. A unique ρ_{XY}^{st} is implied below to exist.

The marginal pmf of X_n is obtained from the pmf of Z_n via summation over Y_n values as $p_X^{(n)}(a_i) = \sum_{j=1}^{N} p_{XY}^{(n)}\left(a_i, b_j\right)$. So, the pmf of X_n for a given past pmf of Z_{n-1} reads $\mathbf{p}_X^{(n)} = \mathbf{R}\mathbf{p}_{XY}^{(n-1)}$, where the $M \times MN$ matrix $\mathbf{R}$ has the elements $r_{i',j'}^{i} = P\left\{X_n = a_i | X_{n-1} = a_{i'}, Y_{n-1} = b_{j'}\right\} = \sum_{j=1}^{N} q_{i',j'}^{i,j}$ which are the probabilities of the transitions $\left(a_{i'}, b_{j'}\right) \to a_i$. One naturally says that the influence (or directional coupling, or causality) $Y \to X$ exists if any of these probabilities depends on j', i.e., if $r_{i',j*}^{i} \neq r_{i',j**}^{i}$ for some set $\left(i, i', j^*, j^{**}\right)$. There is no coupling $Y \to X$ if $r_{i',j*}^{i} = r_{i',j**}^{i}$ for any $\left(i, i', j^*, j^{**}\right)$. Everything is similar for the opposite direction $X \to Y$.

For a system **X** isolated from **Y**, denote the individual probabilities of the transitions $a_{i'} \to a_i$ via $r_{i'}^{i}$. Define that introducing the coupling $Y \to X$ makes the above transition probabilities $r_{i',j'}^{i}$ dependent on j' as $r_{i',j'}^{i} = r_{i'}^{i} + \Delta r_{i',j'}^{i}$ where the profile of perturbations $\left\{\Delta r_{i',j'}^{i}\right\}_{j'=1}^{N}$ includes at least one nonzero element and has a zero sum $\sum_{j'=1}^{N} \Delta r_{i',j'}^{i} = 0$ for definiteness. In order to parameterize the directional coupling $Y \to X$ with a single parameter α_{XY},

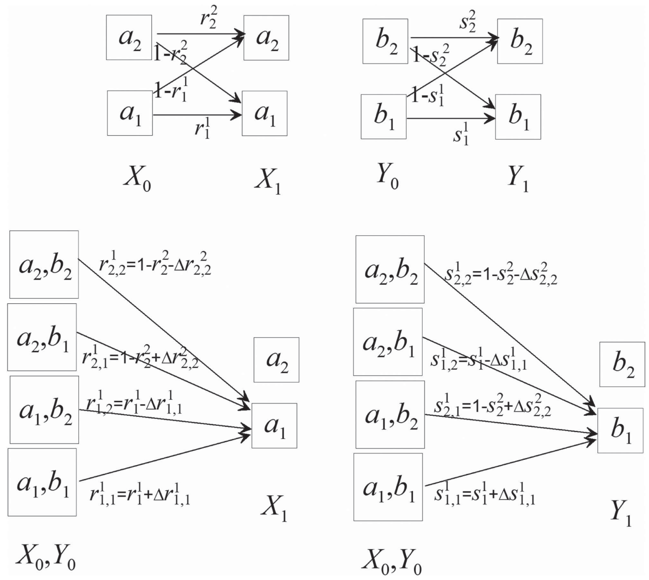

FIG. 1. Illustrations of the transition probabilities for two-state Markov chains **X** and **Y**: individual dynamics (top figures) and coupled dynamics (bottom figures) for mutually independent noises in **X** and **Y** (i.e., $q_{i',j'}^{i,j} = r_{i',j'}^{i} s_{j',i'}^{j}$, see Sec. V B) and $\alpha_{XY} = \alpha_{YX} = 1$.

one can define $r_{i',j'}^{i} = r_{i'}^{i} + \alpha_{XY}\Delta r_{i',j'}^{i}$, so a nonzero α_{XY} means that the coupling $Y \to X$ exists and a greater α_{XY} indicates a stronger dependence of the future pmf of X_n on the previous y_{n-1}. The opposite coupling $X \to Y$ can be parameterized in a similar way: $\mathbf{p}_Y^{(n)} = \mathbf{Sp}_{YX}^{(n-1)}$ with probabilities $s_{j',i'}^{j} = s_{j'}^{j} + \alpha_{YX}\Delta s_{j',i'}^{j}$ for the transitions $(a_{i'}, b_{j'}) \to b_j$. Such a system $\mathbf{Z}$ is called here a system of coupled Markov chains $\mathbf{X}$ and $\mathbf{Y}$ implying that each of the processesX_n and Y_n alone would be a Markov chain if the coupling parameters α_{XY} and α_{YX} were set equal to zero. The full probabilities $q_{i',j'}^{i,j}$ can be defined as the products $q_{i',j'}^{i,j} = r_{i',j'}^{i} s_{j',i'}^{j}$ corresponding to "independent internal noises" in $\mathbf{X}$ and $\mathbf{Y}$ (see Fig. 1 and Sec. V B), or in another way meeting the normalization constraints.

III. FRAMEWORK OF DYNAMICAL CAUSAL EFFECTS AND LOGICS OF THEIR GENERATION

In order to characterize the strength of the influence $Y \to X$ within the DCE framework,[44] one specifies two initial pmfs $p_{XY}^{(0)}$ as $\rho_{XY}^{*} = \rho_X^{*}(x)\rho_{Y|X}^{*}(y|x)$ (the reference initial condition) and $\rho_{XY}^{**} = \rho_X^{*}(x)\rho_{Y|X}^{**}(y|x)$ (the alternative initial condition) with the same marginal pmf ρ_X^{*} of X_0 and quantifies the difference between the respective future pmfs $p_X^{(1)}[x|\rho_{XY}^{*}]$ and $p_X^{(1)}[x|\rho_{XY}^{**}]$. Such an ordered pair of initial pmfs is called the *initial Y-variation* and the respective pair of the future pmfs is the *X-response*. The difference between the X-response components is given by a (continuous) *distinction functional D* which is zero for equal components and may be non-zero otherwise. Its values are finally assembled (e.g., averaged) over diverse initial Y-variations, i.e., over a set Λ of possible values of a certain parameter vector λ entering the initial conditions. Such an *assemblage functional* denoted A_Λ gives the value of a concrete short-term (one-step) DCE,

$$C_{Y\to X} = A_\Lambda\left(D\left(p\left[x|\rho_{XY,\lambda}^{*}\right], p\left[x|\rho_{XY,\lambda}^{**}\right]\right)\right), \tag{3}$$

where explicit dependencies of both initial conditions on λ are indicated in their subscripts. Everything is analogous for the coupling $X \to Y$.

Below, diverse DCEs of the form (3) are developed via considering different Y-variations, different information-theoretic functionals D, and different functionals A_Λ. To be precise, denote the X-response components $p_X^{*}(x) = p_X^{(1)}\left[x|\rho_{XY,\lambda}^{*}\right]$ and $p_X^{**}(x) = p_X^{(1)}\left[x|\rho_{XY,\lambda}^{**}\right]$. A pointwise comparison of the pmfs p_X^{*} and p_X^{**} would give M numbers. The local information $h_X(x) = -\log p_X(x)$ (also called "local entropy") is a basic quantity characterizing p_X in the information theory, so the difference $h_X^{**}(x) - h_X^{*}(x) = \log\frac{p_X^{*}(x)}{p_X^{**}(x)}$ is used to compare p_X^{*} and p_X^{**}. Let us define the distinction functional as its average value over x with some weight function $w(x)$,

$$D\left(p_X^{*}, p_X^{**}\right) = \sum_{i=1}^{M} w(a_i)\log\frac{p_X^{*}(a_i)}{p_X^{**}(a_i)}. \tag{4}$$

Taking $w(x)$ to equal one of the two pmfs as $w(x) = p_X^{*}(x)$ (it is often done below), one obtains a non-negative distinction functional called the Kullback–Leibler divergence (KLD),

$$D\left(p_X^{*}, p_X^{**}\right) = D_{KL}\left(p_X^{*}||p_X^{**}\right) = \sum_{i=1}^{M} p_X^{*}(a_i)\log\frac{p_X^{*}(a_i)}{p_X^{**}(a_i)}. \tag{5}$$

Note that $D_{KL}\left(p_X^{*}||p_X^{**}\right) = -\sum_{i=1}^{M} p_X^{*}(a_i)\log p_X^{**}(a_i) - H\left(p_X^{*}\right)$, where $H\left(p_X^{*}\right)$ is the Shannon entropy of p_X^{*}, i.e., the expectation of the local entropy $H\left(p_X^{*}\right) = -\sum_{i=1}^{M} p_X^{*}(a_i)\log p_X^{*}(a_i)$. Binary logarithms are used in this work, so the units of all such informations or entropies are bits. To define the assemblage functional, let us average the elementary DCE (4) over Y-variations $\left(\rho_{XY,\lambda}^{*}, \rho_{XY,\lambda}^{**}\right)$, i.e., over different values of the parameter λ with some weight function $u(\lambda)$,

$$C_{Y\to X} = \sum_{\lambda\in\Lambda} u(\lambda) D\left(p_X^{*}, p_X^{**}\right). \tag{6}$$

where the response pmfs (p_X^{*}, p_X^{**}) depend on λ since their initial conditions depend on λ.

All the DCE elements (the two initial pmfs in the Y-variation and the weight functions in the two functionals) are introduced below step-by-step from simple (localized) ones to complex (more and more delocalized) ones on the basis of an arbitrary function called "the basic pmf" and denoted $\rho_{XY}(x,y) = \rho_X(x)\rho_{Y|X}(y|x)$. Let us say that "the localized" function (e.g., of y) is the Kronecker delta $(\delta_{j,j*})$, "a less localized" one is the conditional pmf $(\rho_{Y|X}(b_j|a_i))$, and "even less localized" one is the marginal pmf $(\rho_Y(b_j))$. This order is meaningful since the Shannon entropy of the Kronecker delta is zero, i.e., the least possible, and the Shannon entropy of $\rho_{Y|X}$ (averaged over x and called conditional Shannon entropy) is known to be less than or equal to that of ρ_Y. The basic pmf ρ_{XY} needs to be specified to define all DCEs below. It is often taken to equal the stationary pmf $\rho_{XY}^{st}(x,y) = \rho_X^{st}(x)\rho_{Y|X}^{st}(y|x)$ which is a natural choice corresponding to many coupling estimates usually obtained from stationary time series.

IV. FORMAL RESULTS: SEQUENCE OF INFORMATION-THEORETIC DCES

In the sequence of DCEs generated below, several known information transfers and flows are recognized. The proofs that they coincide with some DCEs are given in Ref. 44 for continuous-time and continuous-state systems and would be straightforward here. The novelty here is not in such proofs, but in constructing a united wider picture of causality quantifiers as mentioned above.

A. X- and Y-localized initial conditions

To generate the first DCE, take the simplest initial conditions localized both with respect to x and y, i.e., the reference $\rho^*_{XY}\left(a_i, b_j\right) = \delta_{i,i*}\delta_{j,j*}$ and the alternative $\rho^{**}_{XY}\left(a_i, b_j\right) = \delta_{i,i*}\delta_{j,j**}$, so the parameter vector of this Y-variation is $\lambda = (i^*, j^*, j^{**})$. The one-step X-response is given by the two transition probability vectors $\mathbf{p}^*_X = \left(r^1_{i*,j*}, \ldots, r^M_{i*,j*}\right)^{\mathrm{T}}$ and $\mathbf{p}^{**}_X = \left(r^1_{i*,j**}, \ldots, r^M_{i*,j**}\right)^{\mathrm{T}}$. Below, the distinction functional (4) is the KLD (5) if it is not stated otherwise, i.e., its weight function is the future pmf for the reference initial condition $w(x) = p^{(1)}_X\left[x|\rho^*_{XY}\right]$. To perform the assemblage (6), let us average over i^* with the marginal basic pmf $\rho_X(a_{i*})$ and over (j^*, j^{**}) with the conditional basic pmf $\rho_{Y|X}(b_{j*}|a_{i*})\rho_{Y|X}(b_{j**}|a_{i*})$. In the full explicit form, the obtained DCE (Table I) reads

$$C^{(1)}_{Y\to X} = \sum_{i*=1}^{M}\sum_{j*=1}^{N}\sum_{j**=1}^{N} \rho_X(a_{i*})\rho_{Y|X}(b_{j*}|a_{i*})\rho_{Y|X}(b_{j**}|a_{i*}) \times D_{KL}\left(p^{(1)}_X\left[x|\delta_{i,i*}\delta_{j,j*}\right] \,||p^{(1)}_X\left[x|\delta_{i,i*}\delta_{j,j**}\right]\right), \quad (7)$$

where the superscript indicates the number of a DCE in the generated sequence of DCEs.

The DCE $C^{(1)}_{Y\to X}$ for $\rho_{XY} = \rho^{st}_{XY}$ coincides with the simple short-term DCE $\left(F^{KL}_{Y\to X}\right)^2$ of Ref. 43 (see Sec. III A there). Furthermore, the "local information response" suggested in Ref. 22 is very similar to the elementary DCE $D_{KL}(p^*_X||p^{**}_X)$ in Eq. (7) with the differences that the former implies continuous variables x and y, uses an infinitesimally small distance between the y-locations of the conditional pmfs $\rho^*_{Y|X}$ and $\rho^{**}_{Y|X}$, and is divided by that distance. The authors of Ref. 22 average that local response over the locations of the Dirac delta functions ρ^*_X and $\rho^*_{Y|X}$ using the stationary pmf and obtain "information response" which is similar to the DCE $C^{(1)}_{Y\to X}$ with $\rho_{XY} = \rho^{st}_{XY}$.

To produce the second DCE, change only the factor $\rho_{Y|X}$ of the assemblage weight function u to ρ_Y. It means that the locations of the Kronecker deltas for Y are averaged with weights independent of the location of X, i.e., they are varied with more freedom on average. If X and Y are highly correlated according to the basic pmf ρ_{XY}, e.g., if a stationary pmf for an almost "synchronized" regime is used as the basic pmf (see examples in Sec. V B), then the assemblage with the factor ρ_Y means that Y-locations are varied with much greater freedom and so the resulting DCE $C^{(2)}_{Y\to X}$ may be much greater than $C^{(1)}_{Y\to X}$. Indeed, the latter is then arbitrarily close to zero because an almost deterministic relationship between x and y in ρ_{XY} means small Y-variations (almost equal reference and alternative initial conditions) in $C^{(1)}_{Y\to X}$.

B. X-localized and Y-delocalized initial conditions

To generate the next DCE, take a Y-delocalized alternative conditional pmf $\rho^{**}_{Y|X} = \rho_{Y|X}$ and so the full alternative initial condition $\rho^{**}_{XY}(b_j|a_i) = \delta_{i,i*}\rho_{Y|X}(b_j|a_i) = \delta_{i,i*}\rho_{Y|X}(b_j|a_{i*})$. The reference $\rho^*_{Y|X}(b_j|a_i) = \delta_{j,j*}$ is always same. Take the assemblage weight function coinciding with the alternative pmf $u\left(a_{i*}, b_{j*}\right) = \rho_{XY}(a_{i*}, b_{j*})$.

TABLE I. Information-theoretic DCEs ("transfers and flows") with their elements: the marginal initial pmf of X (the second column), the alternative conditional pmf of Y (the third column), the distinction weight function (the fourth column) and the assemblage weight function (the fifth column). The reference conditional pmf of Y is the same for all these DCEs: $\rho^*_{Y|X}(b_j|a_i) = \delta_{j,j*}$.

No.	$\rho^*_X(a_i)$	$\rho^{**}_{Y\|X}(b_j\|a_i)$	$w(x)$	$u(\lambda) = u\left(i^*, j^*, j^{**}\right)$
1	$\delta_{i,i*}$	$\delta_{j,j**}$	$p^{(1)}_X[x\|\rho^*_{XY}]$	$\rho_X(a_{i*})\rho_{Y\|X}(b_{j*}\|a_{i*})\rho_{Y\|X}(b_{j**}\|a_{i*})$
2	$\delta_{i,i*}$	$\delta_{j,j**}$	$p^{(1)}_X[x\|\rho^*_{XY}]$	$\rho_X(a_{i*})\rho_Y(b_{j*})\rho_Y(b_{j**})$
3	$\delta_{i,i*}$	$\rho_{Y\|X}(b_j\|a_i)$	$p^{(1)}_X[x\|\rho^*_{XY}]$	$\rho_X(a_{i*})\rho_{Y\|X}(b_{j*}\|a_{i*})$
4	$\delta_{i,i*}$	$\rho_Y(b_j)$	$p^{(1)}_X[x\|\rho^*_{XY}]$	$\rho_X(a_{i*})\rho_Y(b_{j*})$
5	$\delta_{i,i*}$	$\rho_Y(b_j)$	$p^{(1)}_X[x\|\rho^*_{XY}]$	$\rho_X(a_{i*})\rho_{Y\|X}(b_{j*}\|a_{i*})$
6	$\delta_{i,i*}$	$\rho_{Y\|X}(b_j\|a_i)$	$p^{(1)}_X[x\|\rho^*_{XY}]$	$\rho_X(a_{i*})\rho_Y(b_{j*})$
7	$\rho_X(a_i)$	$\delta_{j,j**}$	$p^{(1)}_X[x\|\rho^*_{XY}]$	$\rho_Y(b_{j*})\rho_Y(b_{j**})$
8	$\rho_X(a_i)$	$\rho_Y(b_j)$	$p^{(1)}_X[x\|\rho^*_{XY}]$	$\rho_Y(b_{j*})$
9	$\rho_X(a_i)$	$\rho_{Y\|X}(b_j\|a_i)$	$p^{(1)}_X[x\|\rho^*_{XY}]$	$\rho_Y(b_{j*})$
10	$\rho_X(a_i)$	$\rho_{Y\|X}(b_j\|a_i)$	$p^{(1)}_X\left[x\|\rho_{X\|Y}\left(a_i\|b_j\right)\delta_{j,j*}\right]$	$\rho_Y(b_{j*})$
11	$\rho_X(a_i)$	$\rho_Y(b_j)$	$p^{(1)}_X\left[x\|\rho_{X\|Y}\left(a_i\|b_j\right)\delta_{j,j*}\right]$	$\rho_Y(b_{j*})$

Thereby, one gets the DCE $C^{(3)}_{Y\to X}$ (Table I),

$$C^{(3)}_{Y\to X} = \sum_{i*=1}^{M}\sum_{j*=1}^{N} \rho_X(a_{i*})\rho_{Y|X}(b_{j*}|a_{i*}) \times D_{KL}\left(p_X^{(1)}\left[x|\delta_{i,i*}\delta_{j,j*}\right]||p_X^{(1)}\left[x|\delta_{i,i*}\rho_{Y|X}\left(b_j|a_{i*}\right)\right]\right). \quad (8)$$

It is straightforward to show that, since the pmf $\rho^{**}_{Y|X}(b_j|a_i)$ coincides with the respective factor of the assemblage weight function $u_{b|a}(b_{j*}|a_{i*})$, the DCE $C^{(3)}_{Y\to X}$ equals the difference of the Shannon entropies $H\left(p_X^{(1)}\left[x|\rho^{**}_{XY}\right]\right) - H\left(p_X^{(1)}\left[x|\rho^{*}_{XY}\right]\right)$ weighted with $u\left(a_{i*}, b_{j*}\right)$:

$$C^{(3)}_{Y\to X} = \sum_{i*=1}^{M}\sum_{j*=1}^{N} \rho_X(a_{i*})\rho_{Y|X}(b_{j*}|a_{i*}) \left(H\left(p_X^{(1)}\left[x|\delta_{i,i*}\rho_{Y|X}\left(b_j|a_{i*}\right)\right]\right) - H\left(p_X^{(1)}\left[x|\delta_{i,i*}\delta_{j,j*}\right]\right)\right). \quad (9)$$

So, the difference of Shannon entropies can be equivalently taken as the distinction functional for this DCE. Rewriting Eq. (8) in a detailed form useful for comparison with the LKIF, one gets

$$C^{(3)}_{Y\to X} = \sum_{i*=1}^{M}\sum_{j*=1}^{N}\sum_{k=1}^{M} \rho_{XY}(a_{i*}, b_{j*}) p_X^{(1)}\left[a_k|\delta_{i,i*}\delta_{j,j*}\right] \times \log\frac{p_X^{(1)}\left[a_k|\delta_{i,i*}\delta_{j,j*}\right]}{p_X^{(1)}\left[a_k|\delta_{i,i*}\rho_{Y|X}\left(b_j|a_{i*}\right)\right]}. \quad (10)$$

If $\rho_{XY} = \rho^{st}_{XY}$, the DCE $C^{(3)}_{Y\to X}$ just coincides with the famous transfer entropy (TE),[5] so the latter naturally arises in the sequence of quantifiers developed within the DCE framework following the logics of increasing complexity (delocalization) of the DCE elements. The original TE[5] as a DCE was discussed in detail in Refs. 44 and 56. If $\rho_{XY} \neq \rho^{st}_{XY}$, one can use the DCE $C^{(3)}_{Y\to X}$ as an extension of the TE definition and call it a generalized or extended[44] TE for a given basic pmf ρ_{XY}.

To produce the next DCE $C^{(4)}_{Y\to X}$, take a more delocalized alternative conditional pmf for y as the marginal basic pmf $\rho^{**}_{Y|X} = \rho_Y$ and get $\rho^{**}_{XY}\left(a_i, b_j\right) = \delta_{i,i*}\rho_Y(b_j)$. Retain the same w and take the assemblage weight function $u\left(a_{i*}, b_{j*}\right) = \rho_X(a_{i*})\rho_Y(b_{j*})$ with the same factor ρ_Y as that in ρ^{**}_{XY}. Being considered as a joint pmf, u corresponds to statistically independent generation of a_{i*} and b_{j*} which is often called "randomization" in social science applications.[57] Therefore, such pmf or weight function is called "randomized" in Ref. 44. It was shown (Sec. III A of Ref. 44) that for $\rho_{XY} = \rho^{st}_X\rho^{st}_Y$ the DCE $C^{(4)}_{Y\to X}$ coincides with Ay–Polani IF.[9] Such randomization of the stationary basic pmf makes $C^{(4)}_{Y\to X}$ greater than $C^{(3)}_{Y\to X}$ if the subsystems exhibit an "almost synchronized" regime where $\rho^{st}_{Y|X}(y|x)$ is well localized for any given x. The DCE $C^{(4)}_{Y\to X}$ equals the weighted difference of Shannon entropies like the DCE $C^{(3)}_{Y\to X}$ for the same reason.

The factors $\rho^{**}_{Y|X}(b_j|a_i)$ and $u_{b|a}(b_{j*}|a_{i*})$ may also be taken different from each other. Setting $u\left(a_{i*}, b_{j*}\right) = \rho_{XY}\left(a_{i*}, b_{j*}\right)$ and retaining $\rho^{**}_{Y|X}(y|x) = \rho_Y(y)$ produces the DCE $C^{(5)}_{Y\to X}$ (Table I). If $\rho_{XY} = \rho^{st}_{XY}$, this DCE coincides with the "causal strength" of Janzing *et al.*[16] (Ref. 44, Sec. S3 in the Supplementary material, item 13) suggested from a completely different viewpoint considering effects of "removing arrows" in a causal graph. Moreover, Bollt's IF introduced for continuous variables and called forecastability quality metrics[21] is similar to the DCE $C^{(5)}_{Y\to X}$ with a couple of differences[63] (Ref. 44, Sec. S3 in the Supplementary material, item 14).

On the contrary to $C^{(5)}_{Y\to X}$, take the alternative conditional pmf $\rho^{**}_{Y|X} = \rho_{Y|X}$ and the randomized assemblage weight function $u\left(a_{i*}, b_{j*}\right) = \rho_X(a_{i*})\rho_Y(b_{j*})$ and thereby get the next DCE $C^{(6)}_{Y\to X}$ (Table I). For a stationary basic pmf with highly correlated X and Y, the value of y in the alternative initial condition weakly varies for a given $x = a_{i*}$ (like in the TE) but the Y-location b_{j*} of the reference initial condition may strongly vary for a given $x = a_{i*}$ (contrary to the TE) "exploring" a wider region of the state space and making the value of the DCE $C^{(6)}_{Y\to X}$ greater. So, the novel DCE $C^{(6)}_{Y\to X}$ has also a clear statistical sense like the above IFs. Both $C^{(5)}_{Y\to X}$ and $C^{(6)}_{Y\to X}$ are not equal to any weighted difference of two Shannon entropies.

C. X-delocalized and Y-localized initial conditions

To produce the next DCE, take a delocalized marginal pmf of X as $\rho^*_X = \rho_X$ and retain it for all DCEs below. Take the localized conditional pmfs $\rho^*_{XY}\left(a_i, b_j\right) = \rho_X(a_i)\delta_{j,j*}$ and $\rho^{**}_{XY}\left(a_i, b_j\right) = \rho_X(a_i)\delta_{j,j**}$. Then, the assemblage is done over (j^*, j^{**}), so the natural weight function is here $u(b_{j*}, b_{j**}) = \rho_Y(b_{j*})\rho_Y(b_{j**})$. Thereby, one gets the DCE $C^{(7)}_{Y\to X}$ (Table I) which is an analog to $C^{(2)}_{Y\to X}$, but the former DCE compares the evolutions "aggregated over initial x" and should often be less than $C^{(2)}_{Y\to X}$.

D. X- and Y-delocalized initial conditions

To produce the next DCE, take a delocalized alternative conditional pmf of Y equal to the assemblage weight function $u(b_{j*}) = \rho_Y(b_{j*})$, i.e., the initial condition $\rho^{**}_{XY}(x, y) = \rho_X(x)\rho_Y(y)$. Then, one gets the DCE $C^{(8)}_{Y\to X}$ (Table I) which is an analog to the Ay–Polani IF $C^{(4)}_{Y\to X}$ for the stationary basic pmf $\rho_{XY} = \rho^{st}_{XY}$, but the former DCE compares evolutions "aggregated over initial x." It is easy to show that the DCE $C^{(8)}_{Y\to X}$ coincides with the LKIF[7,20] defined for the randomized initial pmf $\rho_{XY} = \rho_X\rho_Y$. In that case, the LKIF appears to be a specific DCE and indeed equals the difference of two Shannon entropies as the authors initially wanted to interpret it.[7] However, this interpretation is not valid for the general LKIF which is met below.

To generate the DCE $C^{(9)}_{Y\to X}$, take another delocalized alternative conditional pmf of Y, i.e., $\rho^{**}_{XY}(x, y) = \rho_X(x)\rho_{Y|X}(y|x)$ with the same other DCE elements as in $C^{(8)}_{Y\to X}$ (Table I). The DCE $C^{(9)}_{Y\to X}$ is not a difference of two Shannon entropies and has the causal meaning similar to $C^{(6)}_{Y\to X}$ with the difference that the former DCE compares evolutions "aggregated over initial x."

E. Another weight function in distinction functional

Developing the line of DCEs, we have reached the Y-variation $\rho^*_{XY}\left(a_i, b_j\right) = \rho_X(a_i)\delta_{j,j*}$ and $\rho^{**}_{XY} = \rho_{XY}$. As a further sophistication of the DCEs, let us replace the distinction weight function $w(x)$

$= p_X^{(1)}\left[x|\rho_{XY}^*\right]$ with a function different from any of the two compared pmfs similarly to the idea of Liang.[20] Namely, take it equal to the one-step future pmf of X for a given initial $y = b_{j*}$ and an initial x distributed with $\rho_{X|Y}$, i.e., $w(x) = p_X^{(1)}\left[x|\rho_{X|Y}\left(x|b_{j*}\right)\delta_{j,j*}\right]$.[64] The distinction functional is then no longer the KLD and depends on the parameter b_{j*}. Taking the assemblage weight function $u(x) = \rho_Y(b_{j*})$ (the product of the two weight functions agree then with the joint distribution ρ_{XY} of the initial x and y), one gets the DCE $C_{Y\to X}^{(10)}$ (Table I),

$$C_{Y\to X}^{(10)} = \sum_{j*=1}^{N}\sum_{k=1}^{M}\rho_Y(b_{j*})p_X^{(1)}\left[a_k|\rho_{X|Y}(x|b_{j*})\delta_{j,j*}\right]\log\frac{p_X^{(1)}\left[a_k|\rho_X\delta_{j,j*}\right]}{p_X^{(1)}\left[a_k|\rho_{XY}\right]}. \quad (11)$$

Using $p_X^{(1)}\left[a_k|\rho_{X|Y}(x|b_{j*})\delta_{j,j*}\right] = \sum_{i*=1}^{M}\rho_{X|Y}\left(a_{i*}|b_{j*}\right)p_X^{(1)}\left[a_k|\delta_{i,i*}\delta_{j,j*}\right]$, it can also be written as

$$C_{Y\to X}^{(10)} = \sum_{i*=1}^{M}\sum_{j*=1}^{N}\sum_{k=1}^{M}\rho_{XY}(a_{i*}, b_{j*})p_X^{(1)}\left[a_k|\delta_{i,i*}\delta_{j,j*}\right]\log\frac{p_X^{(1)}\left[a_k|\rho_X\delta_{j,j*}\right]}{p_X^{(1)}\left[a_k|\rho_{XY}\right]}. \quad (12)$$

It is straightforward to show that it coincides with the general LKIF[20] defined for the initial pmf ρ_{XY}.[64] It is not equal to a weighted difference of two Shannon entropies, but equal to the difference between the specifically averaged local entropy of $p_X^{(1)}\left[x|\rho_X\delta_{j,j*}\right]$ and the Shannon entropy of $p_X^{(1)}\left[x|\rho_{XY}\right]$.[65,66] If the basic pmf equals the stationary pmf, the latter entropy is just the entropy of the stationary marginal pmf. Thus, the LKIF is indeed a causality quantifier exactly expressed as a DCE, similarly to the analysis of Ref. 44 which was performed for continuous-time and continuous-state systems. However, the LKIF is a "more specific" (which also means here "more complex") DCE than the above DCEs since it uses an additional ("non-standard" in a sense) idea for the distinction weight function. The respective modification can be applied to all the above DCEs producing many further DCEs. Such possibilities are very diverse and not considered here.

Finally, change the alternative initial condition to $\rho_{XY}^{**}\left(a_i, b_j\right) = \rho_X(a_i)\rho_Y(b_j)$ to produce the DCE $C_{Y\to X}^{(11)}$ (Table I). It allows more freedom for the variations of y in the alternative conditional pmf in comparison with $C_{Y\to X}^{(10)}$, similarly to the DCEs $C_{Y\to X}^{(4)}$, $C_{Y\to X}^{(5)}$, and $C_{Y\to X}^{(8)}$.

V. QUANTITATIVE RESULTS: TRANSFER ENTROPY VS LIANG–KLEEMAN INFORMATION FLOW

Relationships between different above DCEs are of interest since each of them is meaningful as information-theoretic causality quantifier. Some of them are better known and wider used, in particular, the TE seems to be the most famous one (see, e.g., "its own" monograph[14]) and the LKIF is also quite respectable (see, e.g., climate and other applications[67,68]). Their relationship was briefly mentioned as unknown in the original work on the LKIF.[7] In Ref. 44, it has got a theoretical consideration within the DCE framework for continuous-time stochastic systems. In Ref. 48, some formal relationships between them were established from another viewpoint for a special class of continuous-time systems. As for the numerical values of the TE and the LKIF for the same system in the same direction, they can be either strongly different or close to each other as demonstrated in Ref. 44. However, a fuller quantitative analysis is of interest in respect of the following questions. What are rigorous conditions for deterministic dependencies or inequalities relating these two DCEs (if any)? What are "most probable" numerical ratios of these two DCEs for some classes of systems? Such questions are studied below for a class of simple stochastic systems-coupled two-state Markov chains determined by Eq. (1) with $M = N = 2$.

A. Rigorously proven inequality for randomized basic pmf

Consider the randomized basic pmf $\rho_{XY}(a_i, b_j) = \rho_X(a_i)\rho_Y(b_j)$. Denote $\rho_X(a_1) = \alpha$ and $\rho_Y(b_1) = \beta$. For convenience, denote $f(p)$ a function of the real number $0 \le p \le 1$ whose value is the Shannon entropy of a binary variable with the probability of one of its two states equal to p: $H = f(p) = -p\log p - (1-p)\log(1-p)$. Then, the Shannon entropy of the random variable X_1 conditioned by the initial condition localized at (a_{i*}, b_{j*}) reads $H\left(p_X^{(1)}\left[x|\delta_{i,i*}\delta_{j,j*}\right]\right) = f\left(r_{i*,j*}^1\right)$, see Fig. 2. Averaging it over j^* for a given $X_0 = a_{i*}$ yields $\overline{H}_{i*}^X = \beta f\left(r_{i*,1}^1\right) + (1-\beta)f\left(r_{i*,2}^1\right)$. Next, the Shannon entropy of X_1 conditioned by the initial pmf localized only with respect to X_0 at a_{i*} reads $H_{i*}^X = H\left(p_X^{(1)}\left[x|\delta_{i,i*}\rho_Y\left(b_j\right)\right]\right) = f\left(\beta r_{i*,1}^1 + (1-\beta)r_{i*,2}^1\right)$. As it follows from Eq. (9), the difference $T_{Y\to X,i*} = H_{i*}^X - \overline{H}_{i*}^X$ is the extended TE for the basic pmf $\delta_{i,i*}\rho_Y$. For convenience, denote $g(p_1, p_2) = f(\beta p_1 + (1-\beta)p_2) - \beta f(p_1) - (1-\beta)f(p_2)$ and get $T_{Y\to X,i*} = g\left(r_{i*,1}^1, r_{i*,2}^1\right)$. From (9), the extended TE for the basic pmf $\rho_X\rho_Y$ is written as $T_{Y\to X} = \alpha T_{Y\to X,1} + (1-\alpha)T_{Y\to X,2}$, i.e.

$$C_{Y\to X}^{(3)} = C_{Y\to X}^{(4)} = T_{Y\to X} = \alpha g\left(r_{1,1}^1, r_{1,2}^1\right) + (1-\alpha)g\left(r_{2,1}^1, r_{2,2}^1\right). \quad (13)$$

Similar considerations apply to the LKIF. The Shannon entropy of X_1 conditioned by the initial pmf localized with respect to

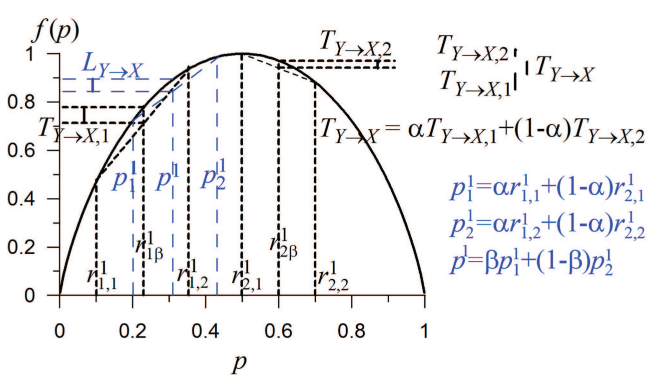

FIG. 2. Illustration to the inequality for the extended TE and the LKIF in the direction $Y \to X$ for the randomized basic pmf with $p_X(a_1) = \alpha = 1/4$ and $p_Y(b_1) = \beta = 1/2$. The following notations are used: $r_{1\beta}^1 = \beta r_{1,1}^1 + (1-\beta)r_{1,2}^1$, $r_{2\beta}^1 = \beta r_{2,1}^1 + (1-\beta)r_{2,2}^1$.

Y_0 at b_{j*} reads $H\left(p_X^{(1)}\left[x|\rho_X(a_i)\delta_{j,j*}\right]\right) = f\left(p_{j*}^1\right)$ where $p_{j*}^1 = \alpha r_{1,j*}^1 + (1-\alpha) r_{2,j*}^1$. The average of $H\left(p_X^{(1)}\left[x|\rho_X(a_i)\delta_{j,j*}\right]\right)$ over j^* is $\beta f\left(p_1^1\right) + (1-\beta) f\left(p_2^1\right)$. The Shannon entropy of X_1 for the initial pmf $\rho_X\rho_Y$ is $H(p_X^{(1)}[x|\rho_X\rho_Y]) = f(\beta p_1^1 + (1-\beta)p_2^1)$. As follows from Eq. (11), the difference $L_{Y\to X} = f(\beta p_1^1 + (1-\beta)p_2^1) - \beta f\left(p_1^1\right) - (1-\beta) f\left(p_2^1\right)$ is the LKIF for the initial pmf $\rho_X\rho_Y$ which then equals

$$C_{Y\to X}^{(8)} = C_{Y\to X}^{(10)} = L_{Y\to X} = g(p_1^1, p_2^1) = g\left(\alpha r_{1,1}^1 + (1-\alpha) r_{2,1}^1, \alpha r_{1,2}^1 + (1-\alpha) r_{2,2}^1\right). \tag{14}$$

Due to convexity of the function $f(p)$ on the interval (0,1), the function $g\left(p_1^1, p_2^1\right)$ which is the difference of the "intermediate" and the "average" values of f is non-negative. It proves that the TE (13) and the LKIF (14) are non-negative. As for the difference $T_{Y\to X} - L_{Y\to X}$, it appears to equal the difference of the "average" and the "intermediate" values of $g(p_1, p_2)$: $T_{Y\to X} - L_{Y\to X} = \alpha g(\mathbf{p}_1) + (1-\alpha) g(\mathbf{p}_2) - g(\alpha \mathbf{p}_1 + (1-\alpha)\mathbf{p}_2)$, where $\mathbf{p}_1 = (r_{1,1}^1, r_{1,2}^1), \mathbf{p}_2 = (r_{2,1}^1, r_{2,2}^1)$. The sign of $T_{Y\to X} - L_{Y\to X}$ depends on whether g is convex as the function of two variables, concave, or arbitrary in the square $(0,1)\times(0,1)$. To find it out, let us find the Hessian (determinant of the partial second derivatives matrix) of g. Via some algebra, it is straightforward to obtain

$$\left|\frac{\partial^2 g}{\partial p_i \partial p_j}\right|_{(p_1^1, p_2^1)} = \frac{\beta^2(1-\beta)^2}{p_1^1(1-p_1^1)p_2^1(1-p_2^1)(\beta p_1^1 + (1-\beta)p_2^1)(\beta(1-p_1^1) + (1-\beta)p_2^1)}\left(p_1^1 - p_2^1\right)^2. \tag{15}$$

This Hessian is nonnegative, so g is concave, rather than convex as one might suppose in analogy with f. Hence, the difference $T_{Y\to X} - L_{Y\to X}$ is nonnegative. Thereby, we have proven the following theorem:

In a system of coupled two-state Markov chains (1), the extended TE for a randomized basic pmf is greater than or equal to the respective LKIF:

$$C_{Y\to X}^{(3)} = C_{Y\to X}^{(4)} = T_{Y\to X} \geq L_{Y\to X} = C_{Y\to X}^{(8)} = C_{Y\to X}^{(10)}. \tag{16}$$

A plausible conjecture is that the inequality (16) is valid for arbitrary M and N, but its proof would be more time-consuming and it is not done here. The equality in (16) holds true if $r_{1,1}^1 = r_{2,1}^1$ and $r_{1,2}^1 = r_{2,2}^1$ (i.e., if the transition probability to any future x does not depend on the initial x), and if $\beta = 0$ or $\beta = 1$ or $\alpha = 0$ or $\alpha = 1$ (i.e., for a localized pmf of either Y_0 or X_0). The inequality (16) may well be violated for a non-randomized basic pmf as will be seen below.

B. Numerical values and their ratios for symmetric Markov chains

A system of coupled two-state Markov chains has 16 parameters (transition probabilities). Only 12 of them can be specified independently (free parameters) due to 4 normalization constrains. Consider here only "independent internal noises" in the subsystems $\mathbf{X}$ and $\mathbf{Y}$, i.e., each full transition probability $q_{i',j'}^{i,j}$ is the product of the "marginal" ones $q_{i',j'}^{i,j} = r_{i',j'}^{i} s_{j',i'}^{j}$, so X_n and Y_n are mutually independent given (x_{n-1}, y_{n-1}). Then, the system $\mathbf{Z}$ has 8 free parameters. As defined in Sec. II (see Fig. 1), $r_1^1 = (r_{1,1}^1 + r_{1,2}^1)/2$ and $r_2^1 = \left(r_{2,1}^1 + r_{2,2}^1\right)/2$ characterize the individual dynamics of $\mathbf{X}$, i.e., zero coupling $Y \to X$ corresponds to the isolated subsystem $\mathbf{X}$ with the transition probabilities r_1^1 and r_2^1. Similarly, $s_1^1 = \left(s_{1,1}^1 + s_{1,2}^1\right)/2$ and $s_2^1 = \left(s_{2,1}^1 + s_{2,2}^1\right)/2$ describe the individual dynamics of $\mathbf{Y}$, $\Delta r_{1,1}^1 = r_{1,1}^1 - r_1^1 = -\Delta r_{1,2}^1$ and $\Delta r_{2,1}^1 = r_{2,1}^1 - r_2^1 = -\Delta r_{2,2}^1$ describe the coupling $Y \to X$, and $\Delta s_{1,1}^1 = s_{1,1}^1 - s_1^1 = -\Delta s_{1,2}^1$ and $\Delta s_{2,1}^1 = s_{2,1}^1 - s_2^1 = -\Delta s_{2,2}^1$ describe the coupling $X \to Y$. To reveal the TE and the LKIF values and relationships, these parameters are varied for different basic pmfs.

Consider first the stationary basic pmf as an important case where $C_{Y\to X}^{(3)}$ coincides with the original[5,56] TE (denote it $T_{Y\to X}^{st}$ to indicate explicitly that $\rho_{XY} = \rho_{XY}^{st}$) and $C_{Y\to X}^{(10)}$ with the most often used version of the LKIF[20] (analogously, denote it $L_{Y\to X}^{st}$). For the most vivid analysis, consider only symmetric subsystems where the properties of the two states of each subsystem are the same: $r_{1,1}^1 = r_{2,2}^2$, $r_{1,2}^1 = r_{2,1}^2$, $s_{1,1}^1 = s_{2,2}^2$, $s_{1,2}^1 = s_{2,1}^2$. So, only 4 free parameters remain. Due to symmetry $\rho_X^{st}(a_i) = \rho_Y^{st}(b_j) = 1/2$. Denote $\rho_{XY}^{st}\left(a_i, b_j\right) = 1/4 + \Delta\rho_{i,j}$. From $\boldsymbol{\rho}_{XY}^{st} = \mathbf{Q}\boldsymbol{\rho}_{XY}^{st}$, it is straightforward to derive $\Delta\rho_{1,1} = \Delta\rho_{2,2} = \Delta\rho_{st}$ and $\Delta\rho_{1,2} = \Delta\rho_{2,1} = -\Delta\rho_{st}$ where

$$\Delta\rho_{st} = \frac{1}{2}\frac{\left(s_1^1 - 1/2\right)\Delta r_{1,1}^1 + \left(r_1^1 - 1/2\right)\Delta s_{1,1}^1}{r_1^1\left(1 - s_1^1\right) + s_1^1\left(1 - r_1^1\right) - 2\Delta r_{1,1}^1 \Delta s_{1,1}^1}. \tag{17}$$

Using the definitions of the TE (10) and the LKIF (12) and some algebra, one derives

$$T_{Y\to X}^{st} = f\left(r_1^1 + 4\Delta\rho_{st}\Delta r_{1,1}^1\right) - \left(\frac{1}{2} + 2\Delta\rho_{st}\right) f\left(r_1^1 + \Delta r_{1,1}^1\right) - \left(\frac{1}{2} - 2\Delta\rho_{st}\right) f\left(r_1^1 - \Delta r_{1,1}^1\right), \tag{18}$$

whose truncated Taylor expansion for small $\Delta r_{1,1}^1$ gives $T_{Y\to X}^{st} \approx \frac{\left(\Delta r_{1,1}^1\right)^2\left(1 - 16(\Delta\rho_{st})^2\right)}{2 r_1^1\left(1 - r_1^1\right)\ln 2}$ bits, and

$$L_{Y\to X}^{st} = \frac{1}{2}\log\left(\left(1 + 2\Delta r_{1,1}^1\right)\left(1 - 2\Delta r_{1,1}^1\right)\right) - \left(4\Delta\rho_{st}\left(r_1^1 - 1/2\right) + \Delta r_{1,1}^1\right)\log\frac{1 - 2\Delta r_{1,1}^1}{1 + 2\Delta r_{1,1}^1} \tag{19}$$

TABLE II. TE and LKIF for different parameters of the coupled two-state Markov chains which are symmetric ($r^1_{1,2} = r^2_{2,1}$, etc.,) and individually identical ($1/2 \le r^1_1 = s^1_1 < 1$).

Coupling	ind. par.	$T^{st}_{Y\to X}$	$L^{st}_{Y\to X}$	$\frac{T^{st}_{Y\to X}}{L^{st}_{Y\to X}}$	Relation
$\Delta s^1_{1,1} = -\Delta r^1_{1,1}$small $\Delta r^1_{1,1}$	$1 - r^1_1 << 1$	$\frac{(\Delta r^1_{1,1})^2}{2(1-r^1_1)\ln 2}$	$\frac{2(\Delta r^1_{1,1})^2}{\ln 2}$	$\frac{1}{4(1-r^1_1)}$	$T^{st}_{Y\to X} >> L^{st}_{Y\to X}$
$\Delta s^1_{1,1} = -\Delta r^1_{1,1}$small $\Delta r^1_{1,1}$	$r^1_1 = 3/4$	$\frac{8(\Delta r^1_{1,1})^2}{3\ln 2}$	$\frac{2(\Delta r^1_{1,1})^2}{\ln 2}$	4/3	$T^{st}_{Y\to X} > L^{st}_{Y\to X}$ moderately
$\Delta s^1_{1,1} = -\Delta r^1_{1,1}$small $\Delta r^1_{1,1}$	$r^1_1 - 1/2 << 1$	$\frac{2(\Delta r^1_{1,1})^2}{\ln 2}$	$\frac{2(\Delta r^1_{1,1})^2}{\ln 2}$	1	$T^{st}_{Y\to X} \approx L^{st}_{Y\to X}$
$\Delta s^1_{1,1} = 0$ small $\Delta r^1_{1,1}$	$r^1_1 - 1/2 << 1$	$\frac{2(\Delta r^1_{1,1})^2}{\ln 2}$	$\frac{2(\Delta r^1_{1,1})^2}{\ln 2}$	1	$T^{st}_{Y\to X} \approx L^{st}_{Y\to X}$
$\Delta s^1_{1,1} = 0$ small $\Delta r^1_{1,1}$	$r^1_1 = 3/4$	$\frac{8(\Delta r^1_{1,1})^2}{3\ln 2}$	$\frac{10(\Delta r^1_{1,1})^2}{3\ln 2}$	4/5	$L^{st}_{Y\to X} > T^{st}_{Y\to X}$ moderately
$\Delta s^1_{1,1} = 0$ small $\Delta r^1_{1,1}$	$1 - r^1_1 << 1$	$\frac{(\Delta r^1_{1,1})^2}{2(1-r^1_1)\ln 2}$	$\frac{(\Delta r^1_{1,1})^2}{(1-r^1_1)\ln 2}$	1/2	$L^{st}_{Y\to X} \approx 2T^{st}_{Y\to X}$
$\Delta s^1_{1,1} > 0$ small $\Delta r^1_{1,1}$	$1/2 < r^1_1 < 1$	$\propto(\Delta r^1_{1,1})^2$	$\propto\Delta r^1_{1,1}$	$\propto\Delta r^1_{1,1}$	$T^{st}_{Y\to X} << L^{st}_{Y\to X}$
$\Delta r^1_{1,1} = \Delta s^1_{1,1}$ $\approx 1 - r^1_1 > 0$	$1/2 < r^1_1 < 1$	close to 0	$(r^1_1 - 1/2)(1 - r^1_1)$	close to 0	$T^{st}_{Y\to X} << L^{st}_{Y\to X}$
$\Delta s^1_{1,1} < 0$small $\Delta r^1_{1,1}$	$1/2 < r^1_1 < 1$	$\propto(\Delta r^1_{1,1})^2$	$\propto-\Delta r^1_{1,1}$	$\propto-\Delta r^1_{1,1}$	$T^{st}_{Y\to X} << \lvert L^{st}_{Y\to X}\rvert$
$\Delta s^1_{1,1} < 0 \Delta r^1_{1,1} < \lvert\Delta s^1_{1,1}\rvert$	$1/2 < r^1_1 < 1$	Eq. (18)	Eq. (19)	about 1	$T^{st}_{Y\to X} \approx L^{st}_{Y\to X}$
$\Delta r^1_{1,1} > \lvert\Delta s^1_{1,1}\rvert$	$1/2 < r^1_1 < 1$	Eq. (18)	Eq. (19)	> 1	$T^{st}_{Y\to X} > L^{st}_{Y\to X}$

with $L^{st}_{Y\to X} \approx (16\Delta\rho_{st}\Delta r^1_{1,1}(r^1_1 - 1/2) + 2(\Delta r^1_{1,1})^2)/\ln 2$ bits for small $\Delta r^1_{1,1}$. From these expressions, one can see characteristic relationships between the TE and the LKIF. For simplicity, consider only identical individual parameters $r^1_1 = s^1_1$ which appear to suffice for illustrations of all diverse relationships between the two IFs. They are summarized in Table II and briefly described below.

(i) For the antisymmetric coupling $\Delta r^1_{1,1} = -\Delta s^1_{1,1}$ [rows 2, 3, 4 in Table II, Fig. 3(a)], one has $\Delta\rho_{st} = 0$ and $T^{st}_{Y\to X}/L^{st}_{Y\to X} \approx 1/\left(4r^1_1\left(1 - r^1_1\right)\right)$ for small enough (i.e., infinitesimally small) $\Delta r^1_{1,1}$, i.e., the TE is much greater than the LKIF for $1 - r^1_1 << 1/4$ and $T^{st}_{Y\to X} = L^{st}_{Y\to X}$ for $r^1_1 = 0.5$.

(ii) For the unidirectional coupling $Y \to X$ [Table II, Fig. 3(b)], one has $\Delta\rho_{st} = \frac{(r^1_1 - 1/2)\Delta r^1_{1,1}}{4r^1_1(1-r^1_1)}$. Then, $16(\Delta\rho_{st})^2 << 1$ if r^1_1 is not close to 1. So, $\frac{L^{st}_{Y\to X}}{T^{st}_{Y\to X}} \approx 2 - 4r^1_1\left(1 - r^1_1\right)$ and the LKIF exceeds the TE by a moderate factor of 1 to 2.

(iii) For a finite (i.e., not infinitesimally small) opposite coupling $\Delta s^1_{1,1} > 0$, a finite $1 - r^1_1 > 0$, and a small $|\Delta r^1_{1,1}| << 1 - r^1_1$ [Table II, Figs. 3(c) and 3(d)], the LKIF is much greater than the TE because $\Delta\rho_{st} \approx \frac{1}{4}\frac{(r^1_1 - 1/2)\Delta s^1_{1,1}}{r^1_1(1-r^1_1)}$ and so $T^{st}_{Y\to X}/L^{st}_{Y\to X} \approx \frac{1-16(\Delta\rho_{st})^2}{32r^1_1(1-r^1_1)(r^1_1-1/2)\Delta\rho_{st}}\Delta r^1_{1,1}$ is infinitesimally small. The LKIF is much greater than the TE also for the almost maximally possible finite couplings $\Delta r^1_{1,1} = \Delta s^1_{1,1} \approx 1 - r^1_1 > 0$, since then $\Delta\rho_{st} \approx 1/4$ (almost deterministic relationship between simultaneous x and y) and so $T^{st}_{Y\to X} \propto 1 - 16(\Delta\rho_{st})^2 \approx 0$ while $L^{st}_{Y\to X} \approx (r^1_1 - 1/2)(1 - r^1_1) \neq 0$.

(iv) The negative LKIF is encountered for a finite opposite coupling $\Delta s^1_{1,1} < 0$, a finite $1 - r^1_1 > 0$, and a small $0 < \Delta r^1_{1,1} << 1 - r^1_1$ [Table II, Figs. 3(e) and 3(f)]. Then, $T^{st}_{Y\to X}/L^{st}_{Y\to X} \approx -c\cdot\Delta r^1_{1,1}$ with a finite constant $c > 0$, i.e., the TE is much less than $|L^{st}_{Y\to X}|$. Under an increase of $\Delta r^1_{1,1}$, the TE gets first equal to the LKIF absolute value, then the LKIF decreases in absolute value and gets equal to zero, and finally both DCEs gets positive with the TE exceeding the LKIF by a moderate factor.

These four characteristic situations represent a full set of diverse ratios between the two IFs.

C. Statistical numerical results for arbitrary Markov chains

To reveal the "most typical" values of the TE and LKIF, let us generate an ensemble of pairs of coupled Markov chains with randomly chosen parameters. Each of the 8 transition probabilities ($r^1_{i,j}$ and $s^1_{i,j}$) is independently generated as a random number uniformly distributed in the interval [0,1]. Let us first consider the case of the stationary basic pmf which is uniquely determined by the 8-dimensional parameter vector from $\boldsymbol{\rho}^{st}_{XY} = \mathbf{Q}\boldsymbol{\rho}^{st}_{XY}$. Given all parameters and this basic pmf, the TE and the LKIF in a chosen direction are

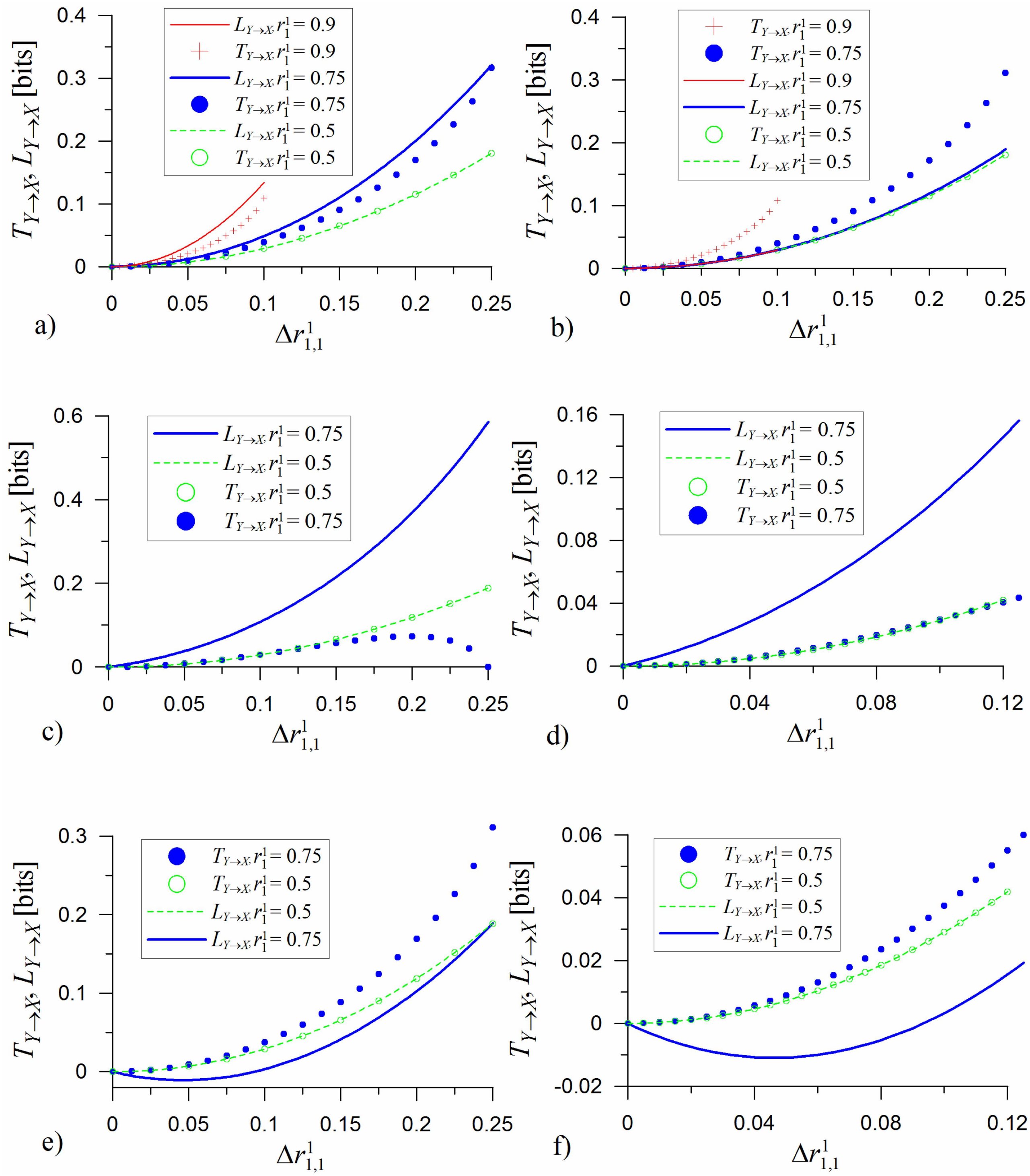

FIG. 3. The TE $T^{st}_{Y\to X}$ and the LKIF $L^{st}_{Y\to X}$ for the coupled symmetric identical two-state Markov chains, different lines (LKIF) and symbols (TE) correspond to different r_1^1: (a) antisymmetric coupling $\Delta s_{1,1}^1 = -\Delta r_{1,1}^1$, (b) unidirectional coupling $Y \to X$, i.e., $\Delta s_{1,1}^1 = 0$, (c) and (d) finite $\Delta s_{1,1}^1 = 0.25$, (e) and (f) finite $\Delta s_{1,1}^1 = -0.25$. The panels (c) and (e) are magnified segments of the panels (d) and (f), respectively.

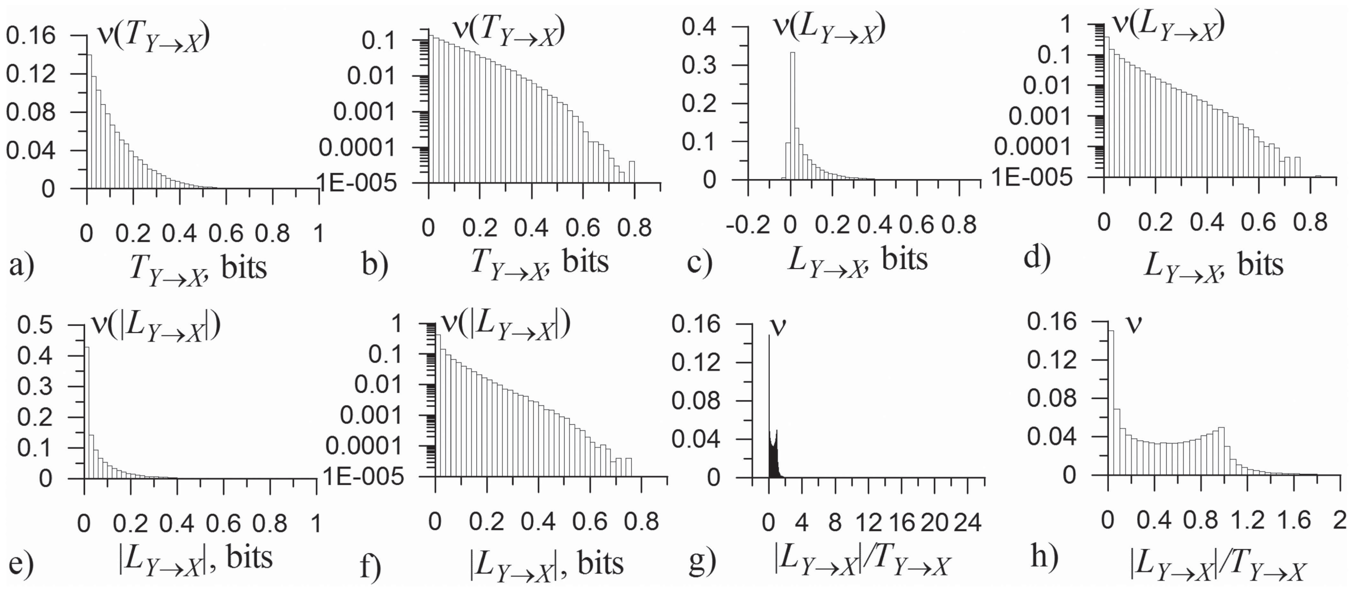

FIG. 4. Histograms of the DCE values for an ensemble of transition probabilities uniformly distributed in the interval [0,1] and the stationary basic pmf: (a) and (b) for the TE $T^{st}_{Y\to X}$ in the linear and semi-logarithmic scales, (c) and (d) for the LKIF $L^{st}_{Y\to X}$, (d) and (e) for the LKIF absolute value $|L^{st}_{Y\to X}|$, (g) and (h) for the ratio of the LKIF absolute value $|L^{st}_{Y\to X}|$ to the TE $T^{st}_{Y\to X}$ and its magnified segment. ν is probability of a bin, bin sizes are 0.02 (a)–(f) and 0.05 (g) and (h).

uniquely computed via Eqs. (10) and (12). Thereby, one obtains an arbitrarily large sample of the TE and the LKIF values and, hence, their empirical pmfs and statistical momenta. For a sample of the size of 10^5, the histograms are shown in Figs. 4(a)–4(f). The shape of the histogram $\nu(T^{st}_{Y\to X})$ is exponential up to about $T^{st}_{Y\to X} = 0.4$bits [Figs. 4(a) and 4(b)], more precisely $\nu(T^{st}_{Y\to X}) \propto exp(-8T^{st}_{Y\to X})$. This is similar for the LKIF along the positive semi-axis except for the vicinity of zero [Figs. 4(c) and 4(d)]: $\nu(L^{st}_{Y\to X}) \propto exp(-11L^{st}_{Y\to X})$, and almost the same for the LKIF absolute value [Figs. 4(e) and 4(f)]. The ratio $\frac{|L^{st}_{Y\to X}|}{T^{st}_{Y\to X}}$ is distributed with two local maxima (0 and close to 1) and rare large values (up to the values in the interval from 20 to 30 in the sample at hand) as shown in Figs. 4(g) and 4(h). The expectation of the TE is estimated as $E[T^{st}_{Y\to X}] = 0.1244 \pm 0.0007$ bits, while the expectation of the LKIF is $E[L^{st}_{Y\to X}] = 0.0626 \pm 0.0006$ bits and of its absolute value is $E|L^{st}_{Y\to X}| = 0.0638 \pm 0.0006$ bits. The ratio of the expectations is 0.51 and the expectation of the ratio $\frac{|L^{st}_{Y\to X}|}{T^{st}_{Y\to X}}$ is almost the same: 0.53. The expectation of the inverse ratio $\frac{T^{st}_{Y\to X}}{|L^{st}_{Y\to X}|}$ is much greater than unity (since the LKIF is often much less than the TE, this is also the case in all examples below), while its median is 2.1, i.e., close to the ratio 2:1. Table III summarizes these values and other cases indicated below.

The values $r^1_1 \geq 0.5$ and $s^1_1 \geq 0.5$ correspond to persistence of both subsystems. This is typical for continuous-time systems sampled with sufficiently small time steps and so is of special interest. Consider the above example with transition probabilities uniformly distributed in the interval [0.5,1] under other equal conditions. Then, the median of $\frac{|L^{st}_{Y\to X}|}{T^{st}_{Y\to X}}$ is 0.6 which is close to the ratio of expectations. The median of the inverse ratio $\frac{T^{st}_{Y\to X}}{|L^{st}_{Y\to X}|}$ is 1.67, i.e., not

TABLE III. The estimates of the expectations of the TE and the LKIF for ensembles of pairs of coupled two-state Markov chains with random parameters uniformly distributed in the intervals [0,1] or [0.5,1] (the second column) and different specifications of the basic pmf (the first column).

Basic pmf	Param. pdf	$E[T_{Y\to X}]$ bits	$E\|L_{Y\to X}\|$ bits	$\frac{E\|L_{Y\to X}\|}{E[T_{Y\to X}]}$	$E\frac{\|L_{Y\to X}\|}{T_{Y\to X}}$	Median of $\frac{\|L_{Y\to X}\|}{T_{Y\to X}}$	Median of $\frac{T_{Y\to X}}{\|L_{Y\to X}\|}$
Stationary	[0,1]	0.12	0.064	0.52	0.53	0.48	2.1
Stationary	[0.5,1]	0.04	0.028	0.65	0.77	0.60	1.67
Arbitrary	[0,1]	0.12	0.071	0.60	0.94	0.56	1.8
Arbitrary	[0.5,1]	0.038	0.027	0.72	1.9	0.67	1.5
Randomized	[0,1]	0.10	0.06	0.62	0.56	0.63	1.6
Randomized	[0.5,1]	0.032	0.016	0.50	0.47	0.49	2.0

very different from the previous case, though the mean values of both DCEs are about two or three times as small.

To compare the DCEs for an arbitrary basic pmf, the transition probabilities are generated as in the first example, while $\rho_{XY}(a_1, b_1)$, $\rho_{XY}(a_1, b_2)$, $\rho_{XY}(a_2, b_1)$, and $\rho_{XY}(a_2, b_2)$ are taken to be mutually independent uniformly distributed in [0,1] with subsequent division of each of the four values by their sum to provide normalization of the pmf. The median of $\frac{T_{Y\to X}}{|L_{Y\to X}|}$ is then about 1.8. For $r_1^1 \geq 0.5$, $s_1^1 \geq 0.5$ and other equal conditions, the median of $\frac{T_{Y\to X}}{|L_{Y\to X}|}$ is about 1.5.

Finally, consider the transition probabilities as in the first example and a randomized basic pmf $\rho_X \rho_Y$, where $\rho_X(a_1)$ and $\rho_Y(b_1)$ are drawn mutually independently from [0,1]. Then, the median of $\frac{T_{Y\to X}}{|L_{Y\to X}|}$ is about 1.6. For $r_1^1 \geq 0.5$ and $s_1^1 \geq 0.5$, the median of $\frac{T_{Y\to X}}{|L_{Y\to X}|}$ is about 2.05.

To summarize, all three cases (differing by the basic pmfs) give overall similar results. The typical (in terms of medians) ratio of the extended TE to the LKIF absolute value and the ratio of their expectations range from 1.5 to 2 depending on the ensemble constraints. The situations where the TE is much greater than the LKIF and vice versa are quite possible but somewhat less probable to be encountered in the parameter space.

Relationships between the other DCEs deduced in Sec. IV can be readily investigated in the same way. In particular, such computations for the coupled two-state Markov chains show that $C^{(3)}_{Y\to X} \leq C^{(4)}_{Y\to X} \leq C^{(5)}_{Y\to X} \leq C^{(6)}_{Y\to X}$ in a wide range of parameters including all cases of Table II. The corresponding results will be published elsewhere.

As a further extension of this research, one can consider Markov chains of higher orders when the next values of X and Y are generated based on their several previous values. Then, everything will be formally the same in Sec. IV, but the number of states M and N of the systems X and Y will be greater than two even for binary variables, e.g., M and N equal 2^k for k-th order Markov chains [see also the discussion of the original TE and the (k,l)-history TE in Ref. 56]. So, the coupling profiles $\Delta r^i_{i',j'}$ and $\Delta s^i_{i',j'}$ will be more diverse and numerical conditions for different relationships between the TE and the LKIF will take more complex forms than those in Secs. V B and V C. However, principal diversity of the relationships is shown already for the simplest Markov chains here.

Another continuation of the present research could systematically compare performance of the TE and the LKIF (or any other pair or set of quantifiers) in solving different applied problems related to quantification of directional couplings. For example, if one needs a quantifier which most reliably detects a non-zero coupling, i.e., roughly speaking, which takes greater values, then one can conclude from Fig. 3 that sometimes the TE is advantageous in this sense [Fig. 3(b), large couplings in Figs. 3(e) and 3(f)] and sometimes the LKIF works better [Figs. 3(a), 3(c), and 3(d), small couplings in Figs. 3(e) and 3(f)]. However, for other purposes, one could require a quantifier which should be small in almost synchronized regimes (as the original TE) or which can take both positive and negative values depending on the relationship between the basic pmf and the coupling profiles in two directions (as the LKIF), etc., The DCE viewpoint explicates an important circumstance that different quantifiers can be appropriate and advantageous to answer different questions about the coupling role in dynamics and there is no universally best quantifier for all cases. Special studies are necessary to say where each concrete quantifier is better than some others or which quantifier is better for a given problem setting.

VI. DISCUSSION AND CONCLUSIONS

As shown above, quite a parsimonious set of possible initial variations (produced from a single basic pmf which is most often the stationary pmf), distinction functionals (average difference of local entropies with a relevant weight function), and assemblage functionals (weighted sum using the basic pmf to specify the weight function) allows one to formulate many DCEs, each of which meaningfully quantifies the corresponding directional coupling (causality) in a Markov chain. Among the 11 derived DCEs, there are 5 well-known information-theoretic quantifiers ("information transfers and flows") previously suggested from quite different viewpoints (Sec. IV). Here, all of them are produced from the same first principle without much effort spent to formulate such quantifiers in the original works. The proofs that each of those 5 quantifiers is a DCE are basically given in Ref. 44 and briefly mentioned here, while the purpose of this work is to show how those "transfers and flows" arise in the simple logical line within the DCE framework and, thereby, how they are logically related to each other and to many further possible DCEs.

It is easy to extend the set of DCEs by considering, e.g., other distinction functionals. In particular, the DCEs $C^{(3)}_{Y\to X}$, $C^{(4)}_{Y\to X}$, and $C^{(8)}_{Y\to X}$ are equivalently formulated for the distinction functional in the form of the Shannon entropies difference $D(p^*_X, p^{**}_X) = H^{**}_X - H^*_X$. For their initial Y-variations, this difference of Shannon entropies is non-negative. One can take the difference of Shannon entropies to be the distinction functional for any initial Y-variation and obtain another version of each functional in Table I. Then, the non-negativity of a DCE may be violated. Still, such modified DCEs make a clear sense quantifying the change in the "aggregated" uncertainty of the future X rather than a change in the pmf of X. Diverse other versions of initial variations, distinctions and assemblages are obviously possible, e.g., Jensen–Shannon distance instead of the KLD turns the DCE $C^{(5)}_{Y\to X}$ into Bollt's IF.[63] However, the 11 DCEs presented here suffice to demonstrate how the DCE viewpoint allows one to generate a large set of causality quantifiers and enrich interpretations of existing "information transfers and flows" showing their place in such a set.

Apart from the formal derivations, it is studied here how strongly some apparently similar DCEs may differ from each other with respect to their numerical values. In particular, the debated question about the quantitative relationships between the TE and the LKIF is considered for coupled Markov chains which represent possibly the simplest model of coupled stochastic dynamical systems. It is shown that such relationships are diverse depending on the parameters of the subsystems and couplings and on the basic pmf used. For the study of such dependence, the extended (or generalized) TE for an arbitrary basic pmf is taken as it naturally arises within the DCE framework. It is rigorously shown that the extended TE is always greater than or equal to the LKIF for a randomized basic pmf (Sec. V A). But in general the relationships are diverse, including a much greater TE, a much greater LKIF, and both values of the same order, while the LKIF may be either negative or positive; conditions for all such situations are formulated for the coupled two-state Markov chains and a stationary basic pmf (Sec. V B). The distributions of the generalized TE and the LKIF values (in bits) are presented for ensembles of pairs of coupled Markov chains (Sec. V C); they also show that the generalized TE values are "typically" (e.g., in terms of medians) 1.5-2 times as great as the LKIF absolute values. All those results together seem to clarify the question "What is the quantitative relationship between the TE and the LKIF?" to a significant extent: indeed, even the simple system under study shows that the relationships are diverse and the conditions for different relationships can be meaningfully formulated, i.e., it is not reasonable to expect a single simple relationship or a couple of such relationships. Some nearest steps extending the studies in the field of interrelating and comparing various "information transfers and flows" are indicated in the end of Sec. V C.

As a perspective, developing a unified quantitative theory of diverse causality measures for processes is a goal which seems to be attractive even for practitioners, since such a theory should be useful for reliable interpretations of the numerical values of coupling characteristics obtained from data. A detailed study of possible and "typical" values of DCEs (e.g., the TE and the LKIF) and their ratios for the simplest (paradigmatic) stochastic dynamical systems as done in this work is a necessary step to achieve that goal, since such a study can serve as a reference point for a further research in that direction.

ACKNOWLEDGMENTS

This work was carried out under the financial support of the Russian Science Foundation (Grant No. 23-12-00241).

AUTHOR DECLARATIONS

Conflict of Interest

The author has no conflicts to disclose.

Author Contributions

Dmitry A. Smirnov: Conceptualization (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Writing – original draft (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] D. Chionis, A. Dokhane, H. Ferroukhi, and A. Pautz, Chaos **29**, 043126 (2019).
[2] A. Tsonis, E. R. Deyle, R. M. May, G. Sugihara, K. Swanson, J. D. Verbeten, and G. Wang, Proc. Natl. Acad. Sci. U.S.A. **112**, 3253 (2015).
[3] J. Runge, A. Gerhardus, G. Varando, V. Eyring, and G. Camps-Valls, Nat. Rev. Earth Environ. **4**, 487–505 (2023).
[4] J. Runge, Nat. Commun. **14**, 1917 (2023).
[5] T. Schreiber, Phys. Rev. Lett. **85**, 461 (2000).
[6] M. Palus, V. Komarek, Z. Hrncir, and K. Sterbova, Phys. Rev. E **63**, 046211 (2001).
[7] X. S. Liang and R. Kleeman, Phys. Rev. Lett. **95**, 244101 (2005).
[8] K. Hlavackova-Schindler, M. Palus, M. Vejmelka, and J. Bhattacharya, Phys. Rep. **441**, 1 (2007).
[9] N. Ay and D. Polani, Adv. Complex Syst. **11**, 17 (2008).
[10] M. Vejmelka and M. Palus, Phys. Rev. E **77**, 026214 (2008).
[11] M. Staniek and K. Lehnertz, Phys. Rev. Lett. **100**, 158101 (2008).
[12] L. Barnett, A. B. Barrett, and A. K. Seth, Phys. Rev. Lett. **103**, 238701 (2009).
[13] J. T. Lizier and M. Prokopenko, Eur. Phys. J. B **73**, 605 (2010).
[14] T. Bossomaier, L. Barnett, M. Harre, and J. T. Lizier, *An Introduction to Transfer Entropy. Information Flow in Complex Systems* (Springer, Switzerland, 2016).
[15] J. Runge, J. Heitzig, V. Petoukhov, and J. Kurths, Phys. Rev. Lett. **108**, 258701 (2012).
[16] D. Janzing, D. Balduzzi, M. Grosse-Wentrup, and B. Schoelkopf, Ann. Stat. **41**, 2324 (2013).
[17] J. Sun and E. M. Bollt, Phys. D **267**, 49 (2014).
[18] M. Wibral, R. Vicente, and J. T. Lizier, in *Directed Information Measures in Neuroscience* (Springer-Verlag, Berlin, 2014).
[19] J. Runge, Phys. Rev. E **92**, 062829 (2015).
[20] X. San Liang, Phys. Rev. E **94**, 052201 (2016).
[21] E. M. Bollt, Chaos **28**, 075309 (2018).
[22] A. Auconi, B. M. Friedrich, and A. Giansanti, Europhys. Lett. **135**, 28002 (2021).
[23] J. Geweke, J. Am. Stat. Assoc. **77**, 304 (1982).
[24] M. Kaminski, M. Ding, W. A. Truccolo, and S. L. Bressler, Biol. Cybern. **85**, 145 (2001).
[25] L. Baccala and K. Sameshima, Biol. Cybern. **84**, 463 (2001).
[26] M. Dhamala, G. Rangarajan, and M. Ding, Phys. Rev. Lett. **100**, 018701 (2008).
[27] S. L. Bressler and A. K. Seth, NeuroImage **58**, 323 (2011).
[28] M. G. Rosenblum and A. S. Pikovsky, Phys. Rev. E **64**, 045202R (2001).
[29] M. Paluš and A. Stefanovska, Phys. Rev. E **67**, 055201(R) (2003).
[30] D. A. Smirnov and B. P. Bezruchko, Phys. Rev. E **68**, 046209 (2003).
[31] J. Brea, D. F. Russell, and A. B. Neiman, Chaos **16**, 026111 (2006).
[32] B. Kralemann, A. Pikovsky, and M. Rosenblum, New J. Phys. **16**, 085013 (2014).
[33] I. Vlachos, D. Kugiumtzis, and M. Paluš, Chaos **32**, 053111 (2022).
[34] D. Marinazzo, M. Pellicoro, and S. Stramaglia, Phys. Rev. Lett. **100**, 144103 (2008).
[35] I. Vlachos and D. Kugiumtzis, Phys. Rev. E **82**, 016207 (2010).
[36] L. Faes, G. Nollo, and A. Porta, Phys. Rev. E **83**, 051112 (2011).
[37] A. Montalto, S. Stramaglia, L. Faes, G. Tessitore, R. Prevete, and D. Marinazzo, Neural Net. **71**, 159 (2015).
[38] B. Wahl, U. Feudel, J. Hlinka, M. Wächter, J. Peinke, and J. A. Freund, Phys. Rev. E **93**, 022213 (2016).
[39] M. C. Romano, M. Thiel, J. Kurths, and C. Grebogi, Phys. Rev. E **76**, 036211 (2007).
[40] J. F. Donges, Y. Zou, N. Marwan, and J. Kurths, Europhys. Lett. **87**, 48007 (2009).
[41] G. Sugihara, R. May, H. Ye, C. Hsieh, E. Deyle, M. Fogarty, and S. Munch, Science **338**, 496 (2012).
[42] A. Ghouse, L. Faes, and G. Valenza, Phys. Rev. E **104**, 064208 (2021).
[43] D. A. Smirnov, Phys. Rev. E **90**, 062921 (2014).
[44] D. A. Smirnov, Phys. Rev. E **105**, 034209 (2022).
[45] M. Palus, A. Krakovska, J. Jakubik, and M. Chvostekova, Chaos **28**, 075307 (2018).
[46] A. Krakovská, J. Jakubík, M. Chvosteková, D. Coufal, N. Jajcay, and M. Paluš, Phys. Rev. E **97**, 042207 (2018).
[47] 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, and J. Zscheischler, Nat. Commun. **10**, 2553 (2019).
[48] H. Kiwata, Phys. Rev. E **105**, 044130 (2022).
[49] R. G. James, N. Barnett, and J. P. Crutchfield, Phys. Rev. Lett. **116**, 238701 (2016).
[50] P. A. Stokes and P. L. Purdon, Proc. Natl. Acad. Sci. U.S.A. **114**, 7063 (2017).
[51] M. Dhamala, H. Liang, S. L. Bressler, and M. Ding, NeuroImage **175**, 460 (2018).
[52] L. Barnett, A. B. Barrett, and A. K. Seth, NeuroImage **178**, 744 (2018).

[53] D. A. Smirnov, Europhys. Lett. **128**, 20006 (2019).
[54] D. A. Smirnov and I. I. Mokhov, Phys. Rev. E **92**, 042138 (2015).
[55] D. A. Smirnov, Chaos **28**, 075303 (2018).
[56] D. A. Smirnov, Phys. Rev. E **102**, 062139 (2020).
[57] D. A. Smirnov, Chaos **31**, 073127 (2021).
[58] I. I. Mokhov and D. A. Smirnov, Chaos **32**, 063128 (2022).
[59] R. L. Stratonovich, *Topics in Theory of Random Noise* (Gordon and Breach, New York, 1963).
[60] L. Arnold, *Random Dynamical Systems* (Springer-Verlag, Berlin, 1998).
[61] J. Pearl, *Causality: Models, Reasoning, and Inference* (Cambridge University Press, Cambridge, 2000).
[62] V. I. Tikhonov and M. A. Mironov, *Markov Processes* (Sov.: Radio, Moscow, 1977) (in Russian).
[63] Bollt's IF[21] uses the Jensen–Shannon distance instead of the KLD as the distinction functional and does not use any averaging over b_{j*}, but implies either a fixed location of the alternative conditional pmf or zeroing of the coupling parameter (the two versions are equivalent for the discrete maps considered in Ref. 21 as discussed in Supplementary material to Ref. 44). That IF was developed to overcome difficulties which arise when one tries to use the KLD for Dirac delta distributions to quantify couplings between deterministically chaotic dynamical systems.
[64] Such a choice corresponds to the LKIF definition[20] since the joint distribution of $(x_2(\tau), x_1(\tau+1))$ equivalent to our notation (y_0, x_1) is implied in Sec. II of Ref. 20 in the sense $p_{YX}^{(0,1)}(b_{j*}, x_1) = \sum_{i*=1}^{M} \rho_{XY}\left(a_{i*}, b_{j*}\right) p_X^{(1)}[x_1|\delta_{i,i*}\delta_{j,j*}] \equiv \rho_Y(b_{j*}) p_X^{(1)}[x_1|\rho_{X|Y}(x|b_{j*})\delta_{j,j*}]$. There, it is then approximated to make computations easier for continuous-state systems. For discrete systems here, no such approximations are needed and so one can implement the idea of Liang exactly.
[65] For continuous-time systems and infinitesimally small time step (sampling time), this DCE (i.e., the LKIF) reduces to the local entropy difference averaged with even simpler distinction weight function $\rho_{X|Y}(x|b_{j*})$ (as obtained in Appendix D2 in Ref. 44) instead of $p_X^{(1)}[x|\rho_{X|Y}(x|b_{j*})\delta_{j,j*}]$.
[66] A. Stips, D. Macias, C. Coughlan, E. Garcia-Gorriz, and X. S. Liang, Sci. Rep. **6**, 21691 (2016).
[67] H. Xiao, F. Zhang, L. Miao, X. S. Liang, K. Wu, and R. Liu, Clim. Dyn. **55**, 1443–1456 (2020).
[68] G. Wang, C. Zhao, M. Zhang, Y. Zhang, M. Lin, and F. Qiao, Sci. Rep. **10**, 17141 (2020).