SCIENTIFIC REPORTS

OPEN

Received: 28 June 2016
Accepted: 21 October 2016
Published: 11 November 2016

Backward transfer entropy: Informational measure for detecting hidden Markov models and its interpretations in thermodynamics, gambling and causality

Sosuke Ito[1,2]

The transfer entropy is a well-established measure of information flow, which quantifies directed influence between two stochastic time series and has been shown to be useful in a variety fields of science. Here we introduce the transfer entropy of the backward time series called the backward transfer entropy, and show that the backward transfer entropy quantifies how far it is from dynamics to a hidden Markov model. Furthermore, we discuss physical interpretations of the backward transfer entropy in completely different settings of thermodynamics for information processing and the gambling with side information. In both settings of thermodynamics and the gambling, the backward transfer entropy characterizes a possible loss of some benefit, where the conventional transfer entropy characterizes a possible benefit. Our result implies the deep connection between thermodynamics and the gambling in the presence of information flow, and that the backward transfer entropy would be useful as a novel measure of information flow in nonequilibrium thermodynamics, biochemical sciences, economics and statistics.

In many scientific problems, we consider directed influence between two component parts of complex system. To extract meaningful influence between component parts, the methods of time series analysis have been widely used[1–3]. Especially, time series analysis based on information theory[4] provides useful methods for detecting the directed influence between component parts. For example, the transfer entropy (TE)[5–7] is one of the most influential informational methods to detect directed influence between two stochastic time series. The main idea behind TE is that, by conditioning on the history of one time series, informational measure of correlation between two time series represents the information flow that is actually transferred at the present time. Transfer entropy has been well adopted in a variety of research areas such as economics[8], neural networks[9–11], biochemical physics[12–14] and statistical physics[15–19]. Several efforts to improve the measure of TE have also been done[20–22].

In a variety of fields, a similar concept of TE has been discussed for a long time. In economics, the statistical hypothesis test called as the Granger causality (GC) has been used to detect the causal relationship between two time series[23,24]. Indeed, for Gaussian variables, the statement of GC is equivalent to TE[25]. In information theory, nearly the same informational measure of information flow called the directed information (DI)[26,27] has been discussed as a fundamental bound of the noisy channel coding under causal feedback loop. As in the case of GC, DI can be applied to an economic situation[28,29], that is the gambling with side information[4,30].

In recent studies of a thermodynamic model implementing the Maxwell's demon[31,32], which reduces the entropy change in a small subsystem by using information, TE has attracted much attention[13–15,18,33–38]. In this context, TE from a small subsystem to other systems generally gives a lower bound of the entropy change in a

[1]Department of Physics, Tokyo Institute of Technology, Oh-okayama 2-12-1, Meguro-ku, Tokyo 152-8551, Japan. [2]FOM Institute AMOLF, Science Park 104, 1098 XG Amsterdam, The Netherlands. Correspondence and requests for materials should be addressed to S.I. (email: sosuke@stat.phys.titech.ac.jp)

subsystem[15,18,33]. As a tighter bound of the entropy change for Markov jump process, another directed informational measure called the dynamic information flow (DIF)[34] has also been discussed[33–43].

In this article, we provide the unified perspective on different measures of information flow, i.e., TE, DI, and DIF. To introduce TE for backward time series[13,38], called *backward transfer entropy* (BTE), we clarify the relationship between these informational measures. By considering BTE, we also obtain a tighter bound of the entropy change in a small subsystem even for non Markov process. In the context of time series analysis, this BTE has a proper meaning: an informational measure for detecting a hidden Markov model. From the view point of the statistical hypothesis test, BTE quantifies an anti-causal prediction. These fact implies that BTE would be a useful directed measure of information flow as well as TE.

Furthermore, we also discuss the analogy between thermodynamics for a small system[32,44,45] and the gambling with side information[4,30]. To considering its analogy, we found that TE and BTE play similar roles in both settings of thermodynamics and gambling: BTE quantifies a loss of some benefit while TE quantifies some benefit. Our result reveals the deep connection between two different fields of science, thermodynamics and gambling.

Results

Setting. We consider stochastic dynamics of interacting systems $\mathcal{X}$ and $\mathcal{Y}$, which are not necessarily Markov processes. We consider a discrete time k ($=1, \ldots, N$), and write the state of $\mathcal{X}$ ($\mathcal{Y}$) at time k as x_k (y_k). Let $x_k^{(l)} := \{x_k, \ldots, x_{k-l+1}\}$ ($y_k^{(l)} := \{y_k, \ldots, y_{k-l+1}\}$) be the path of system $\mathcal{X}$ ($\mathcal{Y}$) from time $k-l+1$ to k where $l \geq 1$ is the length of the path. The probability distribution of the composite system at time k is represented by $p(X_k = x_k, Y_k = y_k)$, and that of paths is represented by $p(X_k^{(k)} = x_k^{(k)}, Y_k^{(k)} = y_k^{(k)})$, where capital letters (e.g., X_k) represent random variables of its states (e.g., x_k).

The dynamics of composite system are characterized by the conditional probability $p(X_{k+1} = x_{k+1}, Y_{k+1} = y_{k+1} | X_k^{(k)} = x_k^{(k)}, Y_k^{(k)} = y_k^{(k)})$ such that

$$
\begin{aligned}
& p(X_{k+1}^{(k+1)} = x_{k+1}^{(k+1)}, Y_{k+1}^{(k+1)} = y_{k+1}^{(k+1)}) \\
& = p(X_{k+1} = x_{k+1}, Y_{k+1} = y_{k+1} | X_k^{(k)} = x_k^{(k)}, Y_k^{(k)} = y_k^{(k)}) p(X_k^{(k)} = x_k^{(k)}, Y_k^{(k)} = y_k^{(k)}),
\end{aligned}
\tag{1}
$$

where $p(A = a | B = b) := p(A = a, B = b)/p(B = b)$ is the conditional probability of a under the condition of b.

Transfer entropy. Here, we introduce conventional TE as a measure of directed information flow, which is defined as the conditional mutual information[4] between two time series under the condition of the one's past. The mutual information characterizes the static correlation between two systems. The mutual information between X and Y at time k is defined as

$$
I(X_k; Y_k) := \sum_{x_k, y_k} p(X_k = x_k, Y_k = y_k) \ln \frac{p(X_k = x_k, Y_k = y_k)}{p(X_k = x_k) p(Y_k = y_k)}.
\tag{2}
$$

This mutual information is nonnegative quantity, and vanishes if and only if x_k and y_k are statistically independent (i.e., $p(X_k = x_k, Y_k = y_k) = p(X_k = x_k) p(Y_k = y_k)$)[4]. This mutual information quantifies how much the state of y_k includes the information about x_k, or equivalently the state of x_k includes the information about y_k. In a same way, the mutual information between two paths $x_k^{(l)}$ and $y_{k'}^{(l')}$ is also defined as

$$
I(X_k^{(l)}; Y_{k'}^{(l')}) := \sum_{x_k^{(l)}, y_{k'}^{(l')}} p(X_k^{(l)} = x_k^{(l)}, Y_{k'}^{(l')} = y_{k'}^{(l')}) \ln \frac{p(X_k^{(l)} = x_k^{(l)}, Y_{k'}^{(l')} = y_{k'}^{(l')})}{p(X_k^{(l)} = x_k^{(l)}) p(Y_{k'}^{(l')} = y_{k'}^{(l')})}.
\tag{3}
$$

While the mutual information is very useful in a variety fields of science[4], it only represents statistical correlation between two systems in a symmetric way. In order to characterize the directed information flow from X to Y, Schreiber[5] introduced TE defined as

$$
T_{X_k^{(l)} \to Y_{k'+1}^{(l'+1)}} := I(X_k^{(l)}; Y_{k'+1}^{(l'+1)}) - I(X_k^{(l)}; Y_{k'}^{(l')}),
\tag{4}
$$

with $k \leq k'$. Equation (4) implies that TE $T_{X_k^{(l)} \to Y_{k'+1}^{(l'+1)}}$ is an informational difference about the path of the system $\mathcal{X}$ that is newly obtained by the path of the system $\mathcal{Y}$ from time k' to $k'+1$. Thus, TE $T_{X_k^{(l)} \to Y_{k'+1}^{(l'+1)}}$ can be regarded as a directed information flow from $\mathcal{X}$ to $\mathcal{Y}$ at time k'. This TE can be rewritten as the conditional mutual information[4] between the paths of $\mathcal{X}$ and the state of $\mathcal{Y}$ under the condition of the history of $\mathcal{Y}$:

$$
\begin{aligned}
T_{X_k^{(l)} \to Y_{k'+1}^{(l'+1)}} &= I(X_k^{(l)}; Y_{k'+1} | Y_{k'}^{(l')}) \\
&:= \sum_{x_k^{(l)}, y_{k'+1}^{(l'+1)}} p(X_k^{(l)} = x_k^{(l)}, Y_{k'+1}^{(l'+1)} = y_{k'+1}^{(l'+1)}) \\
&\quad \times \ln \frac{p(Y_{k'+1} = y_{k'+1} | X_k^{(l)} = x_k^{(l)}, Y_{k'}^{(l')} = y_{k'}^{(l')})}{p(Y_{k'+1} = y_{k'+1} | Y_{k'}^{(l')} = y_{k'}^{(l')})},
\end{aligned}
\tag{5}
$$

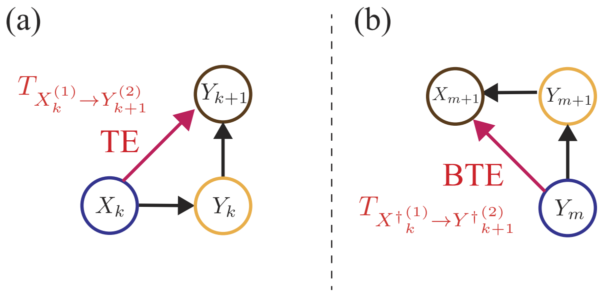

Figure 1. Schematics of TE and BTE. Two graphs (**a**) and (**b**) are the Bayesian networks corresponding to the joint probabilities $p(X_k = x_k, Y_k = y_k, Y_{k+1} = y_{k+1}) = p(X_k = x_k)p(Y_k = y_k|X_k = x_k)p(Y_{k+1} = y_{k+1}|X_k = x_k, Y_k = y_k)$ and $p(X_{m+1} = x_{m+1}, Y_m = y_m, Y_{k+1} = y_{k+1}) = p(Y_m = y_m)p(Y_{m+1} = y_{m+1}|Y_m = y_m)p(X_{m+1} = x_{m+1}|Y_{m+1} = y_{m+1}, Y_m = y_m)$, respectively (see also refs 15, 37 and 60). (**a**) Transfer entropy $T_{X_k^{(1)} \to Y_{k+1}^{(2)}}$ corresponds to the edge from X_k to Y_{k+1} on the Bayesian network. If TE $T_{X_k^{(1)} \to Y_{k+1}^{(2)}}$ is zero, the edge from X_k to Y_{k+1} vanishes, i.e., $p(X_k = x_k, Y_k = y_k, Y_{k+1} = y_{k+1}) = p(X_k = x_k)p(Y_k = y_k|X_k = x_k)p(Y_{k+1} = y_{k+1}|Y_k = y_k)$. (**b**) Backward transfer entropy $T_{X_k^{\dagger(1)} \to Y_{k+1}^{\dagger(2)}}$ corresponds to the edge from Y_m to X_{m+1} on the Bayesian network. If BTE $T_{X_k^{\dagger(1)} \to Y_{k+1}^{\dagger(2)}}$ is zero, the edge from Y_m to X_{m+1} vanishes, i.e., $p(X_{m+1} = x_{m+1}, Y_m = y_m, Y_{m+1} = y_{m+1}) = p(Y_m = y_m)p(Y_{m+1} = y_{m+1}|Y_m = y_m)p(X_{m+1} = x_{m+1}|Y_{m+1} = y_{m+1})$.

which implies that TE is nonnegative quantity, and vanishes if and only if the transition probability in $\mathcal{Y}$ from $y_{k'}^{(l')}$ to $y_{k'+1}$ does not depend on the time series $x_k^{(l)}$, i.e., $p(Y_{k'+1} = y_{k'+1}|X_k^{(l)} = x_k^{(l)}, Y_{k'}^{(l')} = y_{k'}^{(l')}) = p(Y_{k'+1} = y_{k'+1}|X_k^{(l)} = x_k^{(l)}, Y_{k'}^{(l')} = y_{k'}^{(l')})$). [see also Fig. 1(a)].

Backward transfer entropy. Here, we introduce BTE as a novel usage of TE for the backward paths. We first consider the backward path of the system $\mathcal{X}$ ($\mathcal{Y}$); $x_k^{\dagger(l)} := \{x_{N-k+1}, \ldots, x_{N-k+l}\}$ ($y_k^{\dagger(l)} = \{y_{N-k+1}, \ldots, y_{N-k+l}\}$), which is the time-reversed trajectories of the system $\mathcal{X}$ ($\mathcal{Y}$) from time $N-k+l$ to $N-k+1$. We now introduce the concept of BTE defined as TE for the backward paths

$$T_{X_k^{\dagger(l)} \to Y_{k'+1}^{\dagger(l'+1)}} = I(X_k^{\dagger(l)}; Y_{k'+1}^{\dagger(l'+1)}) - I(X_k^{\dagger(l)}; Y_{k'}^{\dagger(l')}) = I(X_{m+l}^{(l)}; Y_{m'}|Y_{m'+l'}^{(l')}), \tag{6}$$

with $m = N-k$, $m' = N-k'$ and $k \leq k'$. In this sense, BTE may represent "the time-reversed directed information flow from the future to the past." However BTE is well defined as the conditional mutual information, it is nontrivial if such a concept makes any sense information-theoretically or physically where stochastic dynamics of composite system itself do not necessarily have the time-reversal symmetry.

To clarify the proper meaning of BTE, we compare BTE $T_{X_k^{\dagger(1)} \to Y_{k+1}^{\dagger(2)}}$ with TE $T_{X_k^{(1)} \to Y_{k+1}^{(2)}}$ [see Fig. 1]. Transfer entropy quantifies the dependence of X_k in the transition from time Y_k to Y_{k+1} [see Fig. 1(a)]. In the same way, BTE quantifies the dependence of Y_m in the correlation between X_{m+1} and Y_{m+1} [see Fig. 1(b)]. Thus, BTE implies how X_{m+1} depends on Y_{m+1} without the dependence of the past state Y_m. In other words, BTE $T_{X_k^{\dagger(1)} \to Y_{k+1}^{\dagger(2)}}$ is nonnegative and vanishes if and only if a Markov chain $Y_m \to Y_{m+1} \to X_{m+1}$ exists, which implies that dynamics of X are given by a hidden Markov model. In general, BTE $T_{X_k^{\dagger(l)} \to Y_{k'}^{\dagger(l')}}$ is nonnegative and vanishes if and only if a Markov chain

$$\begin{aligned} &p(X_{m+l}^{(l)} = x_{m+l}^{(l)}, Y_{m'} = y_{m'}, Y_{m'+l'}^{(l')} = y_{m'+l'}^{(l')}) \\ &\quad = p(Y_{m'} = y_{m'})p(Y_{m'+l'}^{(l')} = y_{m'+l'}^{(l')}|Y_{m'} = y_{m'})p(X_{m+l}^{(l)} = x_{m+l}^{(l)}|Y_{m'+l'}^{(l')} = y_{m'+l'}^{(l')}), \end{aligned} \tag{7}$$

exists. Therefore, BTE from $\mathcal{X}$ to $\mathcal{Y}$ quantifies how far it is from composite dynamics of $\mathcal{X}$ and $\mathcal{Y}$ to a hidden Markov model in $\mathcal{X}$.

Thermodynamics of information. We next discuss a thermodynamic meaning of BTE. To clarify the interpretation of BTE in nonequilibrium stochastic thermodynamics, we consider the following non-Markovian interacting dynamics

$$\begin{aligned} &p(X_{k+1} = x_{k+1}, Y_{k+1} = y_{k+1}|X_k^{(k)} = x_k^{(k)}, Y_k^{(k)} = y_k^{(k)}) = P_{k+1}^{\mathcal{X}} P_{k+1}^{\mathcal{Y}}, \\ &P_{k+1}^{\mathcal{X}} = \begin{cases} p(X_{k+1} = x_{k+1}|X_k = x_k, Y_{k-n} = y_{k-n}) & (k \geq n+1), \\ p(X_{k+1} = x_{k+1}|X_k = x_k, Y_1 = y_1) & (k \leq n), \end{cases} \\ &P_{k+1}^{\mathcal{Y}} = \begin{cases} p(Y_{k+1} = y_{k+1}|Y_k = y_k, X_{k-n} = x_{k-n}) & (k \geq n+1), \\ p(Y_{k+1} = y_{k+1}|Y_k = y_k, X_1 = x_1) & (k \leq n), \end{cases} \end{aligned} \tag{8}$$

where a nonnegative integer n represents the time delay between $\mathcal{X}$ and $\mathcal{Y}$. The stochastic entropy change in heat bath $\mathcal{B}$ attached to the system $\mathcal{X}$ from time 1 to N in the presence of $\mathcal{Y}$[15] is defined as

$$\Delta s_{\mathcal{B}} := \sum_{k=1}^{N-1} \ln \frac{P^{\mathcal{X}}_{k+1}}{Q^{\mathcal{X}}_{k+1}} \tag{9}$$

where

$$Q^{X}_{k+1} = \begin{cases} p_{\mathrm{B}}(X_k = x_k | X_{k+1} = x_{k+1},\ Y_{k-n} = y_{k-n}) & (k \geq n+1), \\ p_{\mathrm{B}}(X_k = x_k | X_{k+1} = x_{k+1},\ Y_1 = y_1) & (k \leq n), \end{cases} \tag{10}$$

is the transition probability of backward dynamics, which satisfies the normalization of the probability $\sum_{x_k} Q^{X}_{k+1} = 1$. For example, if the system $\mathcal{X}$ and $\mathcal{Y}$ does not include any odd variable that changes its sign with the time-reversal transformation, the backward probability is given by $p_{\mathrm{B}}(X_k = x_k | X_{k+1} = x_{k+1}, Y_{k-n} = y_{k-n}) = p(X_{k+1} = x_k | X_k = x_{k+1}, Y_{k-n} = y_{k-n})$ with $k \geq n+1$ ($p_{\mathrm{B}}(X_k = x_k | X_{k+1} = x_{k+1}, Y_1 = y_1) = p(X_{k+1} = x_k | X_k = x_{k+1}, Y_1 = y_1)$ with $k \leq n$). This definition of the entropy change in the heat bath Eq. (9) is well known as the local detailed balance or the detailed fluctuation theorem[45]. We define the entropy change in $\mathcal{X}$ and heat bath as

$$\Delta S_{\mathcal{X}B} := \sum_{x_N^{(N)}, y_N^{(N)}} p(X_N^{(N)} = x_N^{(N)},\ Y_N^{(N)} = y_N^{(N)})[\Delta s_{\mathcal{B}} + \Delta s_{\mathcal{X}}], \tag{11}$$

where $\Delta s_{\mathcal{X}} := -\ln\ p(X_N = x_N) + \ln\ p(X_1 = x_1)$ is the stochastic Shannon entropy change in $\mathcal{X}$.

For the non-Markovian interacting dynamics Eq. (8), we have the following inequality (see Method);

$$\begin{aligned} -\Delta S_{\mathcal{X}B} \leq &- \sum_{k=1}^{n}\left[T_{X_1^{\dagger(1)} \to Y_{k+1}^{\dagger\,(k+1)}} - T_{X_1^{(1)} \to Y_{k+1}^{(k+1)}}\right] - \sum_{k=n+1}^{N-1}\left[T_{X_{k-n}^{\dagger\,(1)} \to Y_{k+1}^{\dagger\,(k+1)}} - T_{X_{k-n}^{(1)} \to Y_{k+1}^{(k+1)}}\right] \\ &- I(X_N;\ Y_N) + I(X_1;\ Y_1) \end{aligned} \tag{12}$$

$$\leq \sum_{k=1}^{n} T_{X_1^{(1)} \to Y_{k+1}^{(k+1)}} + \sum_{k=n+1}^{N-1} T_{X_{k-n}^{(1)} \to Y_{k+1}^{(k+1)}} + I(X_1;\ Y_1). \tag{13}$$

We add that the term $\sum_{k=1}^{n}[T_{X_1^{\dagger\,(1)} \to Y_{k+1}^{\dagger\,(k+1)}} - T_{X_1^{(1)} \to Y_{k+1}^{(k+1)}}]$ vanishes for the Markovian interacting dynamics ($n = 0$).

These results [Eqs (12) and (13)] can be interpreted as a generalized second law of thermodynamics for the subsystem $\mathcal{X}$ in the presence of information flow from $\mathcal{X}$ to $\mathcal{Y}$. If there is no interaction between $\mathcal{X}$ and $\mathcal{Y}$, informational terms vanish, i.e., $T_{X_1^{\dagger(1)} \to Y_{k+1}^{\dagger(k+1)}} = 0$, $T_{X_1^{(1)} \to Y_{k+1}^{(k+1)}} = 0$, $T_{X_{k-n}^{(1)} \to Y_{k+1}^{(k+1)}} = 0$, $T_{X_{k-n}^{(1)} \to Y_{k+1}^{(k+1)}} = 0$, $I(X_N; Y_N) = 0$ and $I(X_1; Y_1) = 0$. Thus these results reproduce the conventional second law of thermodynamics $\Delta S_{\mathcal{X}B} \geq 0$, which indicates the nonnegativity of the entropy change in $\mathcal{X}$ and bath[45]. If there is some interaction between $\mathcal{X}$ and $\mathcal{Y}$, $\Delta S_{\mathcal{X}B}$ can be negative, and its lower bound is given by the sum of TE from X to Y and mutual information between $\mathcal{X}$ and $\mathcal{Y}$ at initial time;

$$I^n(X_N^{(N)} \to Y_N^{(N)}) := \begin{cases} I(X_1;\ Y_1) + \sum_{k=1}^{n} T_{X_1^{(1)} \to Y_{k+1}^{(k+1)}} + \sum_{k=n+1}^{N-1} T_{X_{k-n}^{(1)} \to Y_{k+1}^{(k+1)}} & (n \geq 1) \\ I(X_1;\ Y_1) + \sum_{k=1}^{N-1} T_{X_k^{(1)} \to Y_{k+1}^{(k+1)}} & (n = 0), \end{cases} \tag{14}$$

which is a nonnegative quantity $I^n(X_N^{(N)} \to Y_N^{(N)}) \geq 0$. In information theory, this quantity $I^0(X_N^{(N)} \to Y_N^{(N)})$ is known as DI from $\mathcal{X}$ to $\mathcal{Y}$[27]. Intuitively speaking, $-\Delta S_{\mathcal{X}B}$ quantifies a kind of thermodynamic benefit because its negativity is related to the work extraction in $\mathcal{X}$ in the presence of $\mathcal{Y}$[32]. Thus, a weaker bound (13) implies that the sum of TE quantifies a possible thermodynamic benefit of $\mathcal{X}$ in the presence of $\mathcal{Y}$.

We next consider the sum of TE for the time-reversed trajectories;

$$I^n(X_N^{\dagger(N)} \to Y_N^{\dagger(N)}) = \begin{cases} I(X_N;\ Y_N) + \sum_{k=1}^{n} T_{X_1^{\dagger(1)} \to Y_{k+1}^{\dagger(k+1)}} + \sum_{k=n+1}^{N-1} T_{X_{k-n}^{\dagger(1)} \to Y_{k+1}^{\dagger(k+1)}} & (n \geq 1) \\ I(X_N;\ Y_N) + \sum_{k=1}^{N-1} T_{X_k^{\dagger(1)} \to Y_{k+1}^{\dagger(k+1)}} & (n = 0), \end{cases} \tag{15}$$

which is given by the sum of BTE and the mutual information between $\mathcal{X}$ and $\mathcal{Y}$ at final time. A tighter bound (12) can be rewritten as the difference between the sum of TE and BTE;

$$-\Delta S_{\mathcal{X}B} \leq I^n(X_N^{(N)} \to Y_N^{(N)}) - I^n(X_N^{\dagger\,(N)} \to Y_N^{\dagger\,(N)}) \leq I^n(X_N^{(N)} \to Y_N^{(N)}). \tag{16}$$

This result implies that a possible benefit $I^n(X_N^{(N)} \to Y_N^{(N)})$ should be reduced by up to the sum of BTE $I^n(X_N^{\dagger(N)} \to Y_N^{\dagger(N)})$. Thus, the sum of BTE means a loss of thermodynamic benefit. We add that a tighter bound $I^n(X_N^{(N)} \to Y_N^{(N)}) - I^n(X_N^{\dagger(N)} \to Y_N^{\dagger(N)})$ is not necessarily nonnegative while a weaker bound $I^n(X_N^{(N)} \to Y_N^{(N)})$ is nonnegative.

We here consider the case of Markovian interacting dynamics ($n = 0$). For Markovian interacting dynamics, we have the following additivity for a tighter bound [see Supplementary information (SI)]

$$\begin{aligned} & I^0(X_N^{(N)} \to Y_N^{(N)}) - I^0(X_N^{\dagger(N)} \to Y_N^{\dagger(N)}) \\ & = \sum_{k=1}^{N-1} [I^0(X_{k+1}^{(2)} \to Y_{k+1}^{(2)}) - I^0(X_{N-k+1}^{\dagger(2)} \to Y_{N-k+1}^{\dagger(2)})], \end{aligned} \tag{17}$$

where the sum of TE and BTE for a single time step $I^0(X_{k+1}^{(2)} \to Y_{k+1}^{(2)})$ and $I^0(X_{N-k+1}^{\dagger(2)} \to Y_{N-k+1}^{\dagger(2)})$ are defined as $I^0(X_{k+1}^{(2)} \to Y_{k+1}^{(2)}) := I(X_k; Y_k) + T_{X_k^{(1)} \to Y_{k+1}^{(2)}}$ and $I^0(X_{N-k+1}^{\dagger(2)} \to Y_{N-k+1}^{\dagger(2)}) := I(X_{k+1}; Y_{k+1}) + T_{X_{N-k}^{\dagger(1)} \to Y_{N-k+1}^{\dagger(2)}}$, respectively. This additivity implies that a tighter bound for multi time steps is equivalent to the sum of a tighter bound for a single time step $I^0(X_{k+1}^{(2)} \to Y_{k+1}^{(2)}) - I^0(X_{N-k+1}^{\dagger(2)} \to Y_{N-k+1}^{\dagger(2)}) = I(X_k; \{Y_k, Y_{k+1}\}) - I(X_{k+1}; \{Y_k, Y_{k+1}\})$. We stress that a tighter bound for a single time step has been derived in ref. 13. We next consider the continuous limit $x_k = x(t = k\Delta t)$, $y_k = y(t = k\Delta t)$, and $N = O(\Delta t^{-1})$, where t denotes continuous time, $\Delta t \ll 1$ is an infinitesimal time interval and the symbol O is the Landau notation. Here we clarify the relationship between a tighter bound (16) and DIF[34] (or the learning rate[18]) defined as $I_{\text{flow}}^k := I(X_{k+1}; Y_k) - I(X_k; Y_k)$. For the bipartite Markov jump process[18] or two dimensional Langevin dynamics without any correlation between thermal noises in X and Y[15], we have the following relationship [see also SI]

$$I^0(X_k^{(2)} \to Y_{k+1}^{(2)}) - I^0(X_{N-k+1}^{\dagger\,(2)} \to Y_{N-k+1}^{\dagger\,(2)}) = -I_{\text{flow}}^k + O(\Delta t^2). \tag{18}$$

Thus a bound by TE and BTE is equivalent to a bound by DIF for such systems in the continuous limit, i.e., $-\Delta S_{\mathcal{X}B} \le I^0(X_N^{(N)} \to Y_N^{(N)}) - I^0(X_N^{\dagger(N)} \to Y_N^{\dagger(N)}) = \sum_{k=1}^{N-1} I_{\text{flow}}^{\text{k}} + O(\Delta t)$.

Gambling with side information. In classical information theory, the formalism of the gambling with side information has been well known as another perspective of information theory based on the data compression over a noisy communication channel[4,30]. In the gambling with side information, the mutual information between the result in the gambling and the side information gives a bound of the gambler's benefit.

This formalism of gambling is similar to the above-mentioned result in thermodynamics of information. In thermodynamics, thermodynamic benefit (e.g., the work extraction) can be obtained by using information. On the other hand, the gambler obtain the benefit by using side information. We here clarify the analogy between gambling and thermodynamics in the presence of information flow. To clarify the analogy between thermodynamics and gambling, BTE plays a crucial role as well as TE.

We introduce the basic concept of the gambling with side information given by the horse race[4,30]. Let y_k be the horse that won the k-th horse race. Let $f_k \ge 0$ and $o_k \ge 0$ be the bet fraction and the odds on the k-th race, respectively. Let m_k be the gambler's wealth before the k-th race. Let s_k be the side information at time k. We consider the set of side information $x_{k-1} = \{s_1, \ldots, s_{k-1}\}$, which the gambler can access before the k-th race. The bet fraction f_k is given by the function $f_k(y_k|y_{k-1}^{(k-1)}, x_{k-1})$ with $k \ge 2$, and $f_1(y_1|x_1)$. The conditional dependence $\{y_{k-1}^{(k-1)}, x_{k-1}\}$ ($\{x_1\}$) of $f_k(y_k|y_{k-1}^{(k-1)}, x_{k-1})$ ($f(y_1|x_1)$) implies that the gambler can decide the bet fraction f_k (f_1) by considering the past information $\{y_{k-1}^{(k-1)}, x_{k-1}\}$ ($\{x_1\}$). We assume normalizations of the bet fractions $\sum_{y_k} f_k(y_k|y_{k-1}^{(k-1)}, x_{k-1}) = 1$ and $\sum_{y_1} f_1(y_1|x_1) = 1$, which mean that the gambler bets all one's money in every race. We also assume that $\sum_{y_k} 1/o_k(y_k) = 1$. This condition satisfies if the odds in every race are fair, i.e., $1/o_k(y_k)$ is given by a probability of Y_k.

The stochastic gambler's wealth growth rate at k-th race is given by

$$g_k := \ln \frac{m_{k+1}}{m_k} = \ln[f_k(y_k|y_{k-1}^{(k-1)}, x_{k-1}) o_k(y_k)], \tag{19}$$

with $k \ge 2$ [$g_1 := \ln(m_2/m_1) = f_1(y_1|x_1) o_1(y_1)$], which implies that the gambler's wealth stochastically changes due to the bet fraction and odds. The information theory of the gambling with side information indicates that the ensemble average of total wealth growth $G := \sum_{x_N^{(N)}, y_N^{(N)}} p(X_N^{(N)} = x_N^{(N)}, Y_N^{(N)} = y_N^{(N)}) [\sum_{k=1}^N g_k]$ is bounded by the sum of TE (or DI) from X to Y[28,29] (see Method);

$$G \le \sum_{k=1}^N \langle \ln\ o_k \rangle - S(Y_N^{(N)}) + I^0(X_N^{(N)} \to Y_N^{(N)}) \tag{20}$$

$$\le I^0(X_N^{(N)} \to Y_N^{(N)}), \tag{21}$$

where $\langle \cdots \rangle = \sum_{x_N^{(N)}, y_N^{(N)}} p(X_N^{(N)} = x_N^{(N)}, Y_N^{(N)} = y_N^{(N)}) \cdots$ indicates the ensemble average, and $S(Y_N^{(N)}) := -\langle \ln\ p(Y_N^{(N)} = y_N^{(N)}) \rangle$ is the Shannon entropy of $Y_N^{(N)}$. This result (21) implies that the sum of TE can be interpreted as a possible benefit of the gambler.

We discuss the analogy between thermodynamics of information and the gambling with side information. A weaker bound in the gambling with side information (21) is similar to a weaker bound in thermodynamics of information (16), where the negative entropy change $-\Delta S_{\mathcal{X}B}$ corresponds to the total wealth growth G. On the other hand, a tighter bound in the gambling with side information (20) is rather different from a tighter bound by the sum of BTE in thermodynamics of information (16). We show that a tighter bound in the gambling is also given by the sum of BTE if we consider the special case that the bookmaker who decides the odds o_k cheats in the

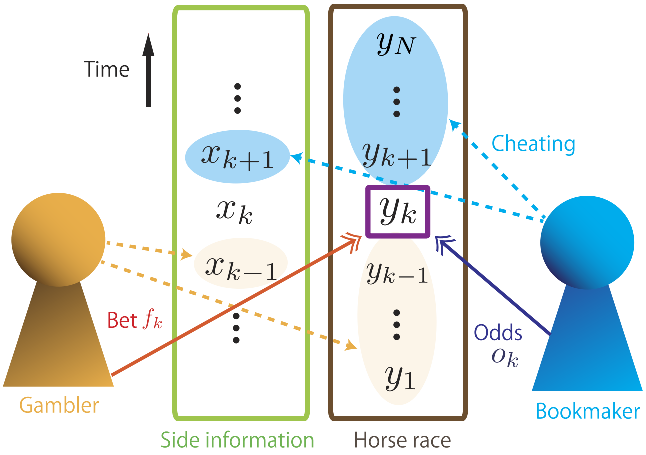

Figure 2. Schematic of the special case of the horse race. The gambler can only access the past side information x_{k-1} and the past races $y_{k-1}^{(k-1)} = \{y_1, \ldots, y_{k-1}\}$, and decides the bet fraction f_k on the k-th race. The bookmaker makes some cheating which can access the future side information x_{k+1} and the future races $y_N^{(N-k)} = \{y_{k+1}, \ldots, y_N\}$, and decides the odds on the k-th race.

horse race; The odds o_k can be decided by the unaccessible side information x_{k+1} and information of the future races $y_N^{(N-k)}$ [see also Fig. 2]. In this special case, the fair odds of the k-th race o_k can be the conditional probability of the future information $1/o_k(y_k) = p(Y_k = y_k | Y_N^{(N-k)} = y_N^{(N-k)}, X_{k+1} = x_{k+1})$ with $k \le N-1$, and $1/o_N(y_N) = p(Y_N = y_N | X_N = x_N)$. The inequality (20) can be rewritten as

$$G \le I^0(X_N^{(N)} \to Y_N^{(N)}) - I^0(X_N^{\dagger\,(N)} \to Y_N^{\dagger\,(N)}) \le I^0(X_N^{(N)} \to Y_N^{(N)}). \tag{22}$$

which implies that the sum of BTE $I^0(X_N^{(N)} \to Y_N^{(N)}) - I^0(X_N^{\dagger(N)} \to Y_N^{\dagger(N)})$ represents a loss of the gambler's benefit because of the cheating by the bookmaker who can access the future information with anti-causality. We stress that Eq. (22) has a same form of the thermodynamic inequality (16) for Markovian interacting dynamics ($n = 0$). This fact implies that thermodynamics of information can be interpreted as the special case of the gambling with side information; The gambler uses the past information and the bookmaker uses the future information. If we regard thermodynamic dynamics as the gambling, anti-causal effect should be considered.

Causality. We here show that BTE itself is related to anti-causality without considering the gambling. From the view point of the statistical hypothesis test, TE is equivalent to GC for Gaussian variables[25]. Therefore, it is naturally expected that BTE can be interpreted as a kind of the causality test.

Suppose that we consider two linear regression models

$$y_{k'+1}^{(1)} = \alpha + (y_{k'}^{(l')} \oplus x_k^{(l)}) \cdot A + \epsilon, \tag{23}$$

$$y_{k'+1}^{(1)} = \alpha' + (y_{k'}^{(l')}) \cdot A' + \epsilon', \tag{24}$$

where α (α') is a constant term, A (A') is the vector of regression coefficients, $\oplus$ denotes concatenation of vectors, and ϵ (ϵ') is an error term. The Granger causality of $\mathcal{X}$ to $\mathcal{Y}$ quantifies how the past time series of $\mathcal{X}$ in the first model reduces the prediction error of $y_{k'+1}^{(1)}$ compared to the error in the second model. Performing ordinary mean squares to find the regression coefficients A (A') and α (α') that minimize the variance of ϵ (ϵ'), the standard measure of GC is given by

$$\mathcal{F}_{X_k^{(l)} \to Y_{k'+1}^{(l'+1)}} := \ln \frac{\mathrm{var}(\epsilon')}{\mathrm{var}(\epsilon)}, \tag{25}$$

where var(ϵ) denotes the variance of ϵ. Here we assume that the joint probability $p(X_k^{(l)} = x_k^{(l)}, Y_{k'+1}^{(l'+1)} = y_{k'+1}^{(l'+1)})$ is Gaussian. Under Gaussian assumption, TE and GC are equivalent up to a factor of 2,

$$2T_{X_k^{(l)} \to Y_{k'+1}^{(l'+1)}} = \mathcal{F}_{X_k^{(l)} \to Y_{k'+1}^{(l'+1)}}. \tag{26}$$

In the same way, we discuss BTE from the view point of GC. Here we assume that the joint probability $p(X_{m+l}^{(l)} = x_{m+l}^{(l)}, Y_{m'+l'}^{(m'+l')} = y_{m'+l'}^{(l'+1)})$ is Gaussian. Suppose that two linear regression models

$$y_{k'+1}^{\dagger(1)} = \alpha^\dagger + (y_{k'}^{\dagger(l')} \oplus x_k^{\dagger(l)}) \cdot A^\dagger + \epsilon^\dagger, \tag{27}$$

$$y_{k'+1}^{\dagger(1)} = \alpha'^{\dagger} + (y_{k'}^{\dagger(l')}) \cdot A'^{\dagger} + \epsilon^{\dagger}, \tag{28}$$

where $\alpha^{\dagger}$ ($\alpha'^{\dagger}$) is a constant term, $A^{\dagger}$ ($A'^{\dagger}$) is the vector of regression coefficients and $\epsilon^{\dagger}$ ($\epsilon'^{\dagger}$) is an error term. These linear regression models give a prediction of the past state of $\mathcal{Y}$ using the future time series of $\mathcal{X}$ and $\mathcal{Y}$. Intuitively speaking, we consider GC of $\mathcal{X}$ to $\mathcal{Y}$ for the rewind playback video of composite dynamics $\mathcal{X}$ and $\mathcal{Y}$. We call this causality test the Granger *anti*-causality of $\mathcal{X}$ to $\mathcal{Y}$. Performing ordinary mean squares to find $A^{\dagger}$ ($A'^{\dagger}$) and $\alpha^{\dagger}$ ($\alpha'^{\dagger}$) that minimize $\mathrm{var}(\epsilon^{\dagger})$ ($\mathrm{var}(\epsilon'^{\dagger})$), we define a measure of the Granger *anti*-causality of $\mathcal{X}$ to $\mathcal{Y}$ as $\mathcal{F}_{X_k^{\dagger(l)} \to Y_{k'+1}^{\dagger(l'+1)}} := \ln\ [\mathrm{var}(\epsilon^{\dagger})/\mathrm{var}(\epsilon^{\dagger})]$. The backward transfer entropy is equivalent to the Granger *anti*-causality up to factor 2,

$$2T_{X_k^{\dagger(l)} \to Y_{k'+1}^{\dagger(l'+1)}} = \mathcal{F}_{X_k^{\dagger(l)} \to Y_{k'+1}^{\dagger(l'+1)}}. \tag{29}$$

This fact implies that BTE can be interpreted as a kind of *anti*-causality test. We stress that composite dynamics of $\mathcal{X}$ and $\mathcal{Y}$ are not necessarily driven with *anti*-causality even if a measure of the Granger *anti*-causality $\mathcal{F}_{X_k^{\dagger(l)} \to Y_{k'+1}^{\dagger(l'+1)}}$ has nonzero value. As GC just finds only the predictive causality[23,24], the Granger *anti*-causality also finds only the predictive causality for the backward time series.

Discussion

We proposed that directed measure of information called BTE, which is possibly useful to detect a hidden Markov model (7) and predictive anti-causality (29). In the both setting of thermodynamics and the gambling, the measurement of BTE has a profitable meaning; the detection of a loss of a possible benefit in the inequalities (16) and (22).

The concept of BTE can provide a clear perspective in the studies of the biochemical sensor and thermodynamics of information, because the difference between TE and DIF has attracted attention recently in these fields[14,35]. In ref. 14, Hartich *et al.* have proposed the novel informational measure for the biochemical sensor called *sensory capacity*. The sensory capacity is defined as the ratio between TE and DIF $C := -I_{\mathrm{flow}}^{\mathrm{k}}/T_{X_k^{(1)} \to Y_{k+1}^{(2)}}$. Because DIF can be rewritten by TE and BTE [Eq. (18)] for Markovian interacting dynamics, we have the following expression for the sensory capacity in a stationary state,

$$C = 1 - \frac{T_{X_{N-k}^{\dagger(1)} \to Y_{N-k+1}^{\dagger(2)}}}{T_{X_k^{(1)} \to Y_{k+1}^{(2)}}}, \tag{30}$$

where we used $I(X_{k+1}; Y_{k+1}) = I(X_k; Y_k)$ in a stationary state. This fact indicates that the ratio between TE and BTE could be useful to quantify the performance of the biochemical sensor. By using this expression (29), we show that the maximum value of the sensory capacity $C = 1$ can be achieved if a Markov chain of a hidden Markov model $Y_k \to Y_{k+1} \to X_{k+1}$ exists. In ref. 35, Horowitz and Sandberg have shown a comparison between two thermodynamic bound by TE and DIF for two dimensional Langevin dynamics. For the Kalman-Bucy filter which is the optimal controller, they have found the fact that DIF is equivalent to TE in a stationary state. This idea can be clarified by the concept of BTE. Because the Kalman-Bucy filter can be interpreted as a hidden Markov model, BTE should be zero, and DIF is equivalent to TE in a stationary state.

Our results can be interpreted as a generalization of previous works in thermodynamics of information[46–48]. In refs 46 and 47, S. Still *et al.* discuss the prediction in thermodynamics for Markovian interacting dynamics. In our results, we show the connection between thermodynamics of information and the predictive causality from the view point of GC. Thus, our results give a new insight into these works of the prediction in thermodynamics. In ref. 48, G. Diana and M. Esposito have introduced the time-reversed mutual information for Markovian interacting dynamics. In our results, we introduce BTE, which is TE in the time-reversed way. Thus, our result provides a similar description of thermodynamics by introducing BTE, even for non-Markovian interacting dynamics.

We point out the time symmetry in the generalized second law (12). For Markovian interacting dynamics, the equality in Eq. (12) holds if dynamics of $\mathcal{X}$ has a local reversibility (see SI). Here we consider a time reversed transformation $\mathcal{T}$: $k \to N - k + 1$, and assume a local reversibility such that the backward probability $p_{\mathrm{B}}(A = a|B = b)$ equals to the original probability $p(A = a|B = b)$ for any random variables A and B. In a time reversed transformation, we have $\mathcal{T}$: $\Delta S_{\mathcal{XB}} \to -\Delta S_{\mathcal{XB}}$, $\mathcal{T}$: $I^n(X_N^{(N)} \to Y_N^{(N)}) \to I^n(X_N^{\dagger(N)} \to Y_N^{\dagger(N)})$ and $\mathcal{T}$: $I^n(X_N^{\dagger(N)} \to Y_N^{\dagger(N)}) \to I^n(X_N^{(N)} \to Y_N^{(N)})$. The generalized second law Eq. (12) changes the sign in a time reversed transformation, $\mathcal{T}$: $[0 \le \Delta S_{\mathcal{XB}} + I^n(X_N^{(N)} \to Y_N^{(N)}) - I^n(X_N^{\dagger(N)} \to Y_N^{\dagger(N)})] \to [0 \le -(\Delta S_{\mathcal{XB}} + I^n(X_N^{(N)} \to Y_N^{(N)}) - I^n(X_N^{\dagger(N)} \to Y_N^{\dagger(N)}))]$. Thus, the generalized second law (12) has the same time symmetry in the conventional second law, i.e., $\mathcal{T}$:$[0 \le \Delta S_{\mathrm{tot}}] \to [0 \le -\Delta S_{\mathrm{tot}}]$ even for non-Markovian interacting dynamics, where ΔS_{tot} is the entropy change in total systems. In other words, the generalized second law (12) provides the arrow of time as the conventional second law. This fact may indicate that BTE is useful as well as TE in physical situations where the time symmetry plays a crucial role in physical laws.

We also point out that this paper clarifies the analogy between thermodynamics of information and the gambling. The analogy between the gambling and thermodynamics has been proposed in ref. 49, however, the analogy between Eqs (16) and (22) are different from one in ref. 49. In ref. 49, D. A. Vinkler *et al.* discuss the particular case of the work extraction in Szilard engine, and consider the work extraction in Szilard engine as the gambling. On the other hand, our result provides the analogy between the general law of thermodynamics of information and the gambling. To clarify this analogy, we may apply the theory of gambling, for example the portfolio theory[50,51], to thermodynamic situations in general. We also stress that the gambling with side information directly connects with the data compression in information theory[4]. Therefore, the generalized second law of thermodynamics may directly connect with the data compression in information theory. To consider such applications, BTE would play a tricky role in the theory of the gambling where the odds should be decided with anti-causality.

Finally, we discuss the usage of BTE in time series analysis. In principle, we prepare the backward time series data from the original time series data, and do a calculation of BTE as TE. To calculate BTE, we can estimate how far it is from dynamics of two time series to a hidden Markov model, or detect the predictive causality for the backward time series. In physical situations, we also can detect thermodynamic performance by comparing BTE with TE. If the sum of BTE from the target system to the other systems is larger than the sum of TE from the target system to the other systems, the target system could seem to violate the second law of thermodynamics because of the inequality (16), where the other systems play a similar role of Maxwell's demon. Therefore, BTE could be useful to detect phenomena of Maxwell's demon in several settings such as Brownian particles[52,53], electric devices[54,55], and biochemical networks[13,56–60].

Method

The outline of the derivation of inequality (12). We here show the outline of the derivation of the generalized second law (12) [see also SI for details]. In SI, we show that the quantity $\Delta S_{\mathcal{X}\mathcal{B}} + I^n(X_N^{(N)} \to Y_N^{(N)}) - I^n(X_N^{\dagger(N)} \to Y_N^{\dagger(N)})$, can be rewritten as the Kullbuck-Leiber divergence $D_{\mathrm{KL}}(\rho\|\tilde{\rho}) := \sum_{x_N^{(N)}, y_N^{(N)}} \rho(x_N^{(N)}, y_N^{(N)}) \ln[\rho(x_N^{(N)}, y_N^{(N)})/\tilde{\rho}(x_N^{(N)}, y_N^{(N)})]$[4], where $\rho(x_N^{(N)}, y_N^{(N)}) := p(X_N^{(N)} = x_N^{(N)}, Y_N^{(N)} = y_N^{(N)})$ and $\tilde{\rho}(x_N^{(N)}, y_N^{(N)}) := p(X_N = x_N, Y_N = y_N) \prod_{k'=n+1}^{N-1} p_{\mathrm{B}}(X_{k'} = x_{k'} | X_{k'+1} = x_{k'+1}, Y_{k'-n} = y_{k'-n}) \prod_{m'=N-n}^{N-1} p(Y_{m'} = y_{m'} | Y_{m'+1} = y_{m'+1}, X_N = x_N) \prod_{k=1}^{n} p_{\mathrm{B}}(X_k = x_k | X_{k+1} = x_{k+1}, Y_1 = y_1) \prod_{m=1}^{N-n-1} p(Y_m = y_m | Y_{m+1} = y_{m+1}, X_{m+n+1} = x_{m+n+1})$, are nonnegative functions that satisfy the normalizations $\sum_{x_N^{(N)}, y_N^{(N)}} \rho(x_N^{(N)}, y_N^{(N)}) = 1$ and $\sum_{x_N^{(N)}, y_N^{(N)}} \tilde{\rho}(x_N^{(N)}, y_N^{(N)}) = 1$. Due to the nonnegativity of the Kullbuck-Leiber divergence, we obtain the inequality (12), i.e., $\Delta S_{\mathcal{X}\mathcal{B}} + I^n(X_N^{(N)} \to Y_N^{(N)}) - I^n(X_N^{\dagger(N)} \to Y_N^{\dagger(N)}) \geq 0$. We add that the integrated fluctuation theorem corresponding to the inequality (12) is also valid, i.e., $\sum \rho(x_N^{(N)}, y_N^{(N)}) \exp(-\ln[\rho(x_N^{(N)}, y_N^{(N)})/\tilde{\rho}(x_N^{(N)}, y_N^{(N)})]) = 1$.

The outline of the derivation of inequality (20). We here show the outline of the derivation of the gambling inequality (20) [see also SI for details]. The quantity $-G + \sum_{k=1}^{N} \langle \ln o_k \rangle - S(Y_N^{(N)}) + I^0(X_N^{(N)} \to Y_N^{(N)})$ can be rewritten as the Kullbuck-Leiber divergence $D_{\mathrm{KL}}(\rho\|\pi)$, where $\rho(x_N^{(N)}, y_N^{(N)}) := p(X_N^{(N)} = x_N^{(N)}, Y_N^{(N)} = y_N^{(N)})$ and $\pi(x_N^{(N)}, y_N^{(N)}) := f_1(y_1|x_1) p(X_1 = x_1) \prod_{k=1}^{N-1} p(X_{k+1} = x_{k+1} | Y_k^{(k)} = y_k^{(k)}, X_k = x_k) f_{k+1}(y_{k+1} | y_k^{(k)}, x_k)$, are nonnegative functions that satisfy the normalizations $\sum_{x_N^{(N)}, y_N^{(N)}} \rho(x_N^{(N)}, y_N^{(N)}) = 1$ and $\sum_{x_N^{(N)}, y_N^{(N)}} \pi(x_N^{(N)}, y_N^{(N)}) = 1$. Due to the nonnegativity of the Kullbuck-Leiber divergence, we have the inequality (20), i.e., $-G + \sum_{k=1}^{N} \langle \ln o_k \rangle - S(Y_N^{(N)}) + I^0(X_N^{(N)} \to Y_N^{(N)}) \geq 0$.

References

1. Hamilton, J. D. "*Time series analysis*" (Princeton: Princeton university press, 1994).
2. Gao, Z. K. & Jin, N. D. A directed weighted complex network for characterizing chaotic dynamics from time series. *Nonlinear Analysis: Real World Applications*, **13,** 947–952 (2012).
3. Ahmed, M. U. & Mandic, D. P. Multivariate multiscale entropy analysis. *IEEE Signal Processing Letters*, **19,** 91–94 (2012).
4. Cover, T. M. & Thomas, J. A. "*Elements of Information Theory*" (John Wiley and Sons, New York, 1991).
5. Schreiber, T. Measuring information transfer. *Phys. Rev. Lett.* **85,** 461 (2000).
6. Kaiser, A. & Schreiber, T. Information transfer in continuous processes. *Physica D* **166,** 43–62 (2002).
7. Hlaváčková-Schindler, K., Paluš, M., Vejmelka, M. & Bhattacharya, J. Causality detection based on information-theoretic approaches in time series analysis. *Phys. Rep.* **441,** 1–46 (2007).
8. Marschinski, R. & Kantz, H. Analyzing the information flow between financial time series. *Eur. Phys. J. B* **30,** 275 (2002).
9. Lungarella, M. & Sporns, O. Mapping information flow in sensorimotor networks. *PLoS Comput. Biol*, **2,** e144 (2006).
10. Vicente, R., Wibral, M., Lindner, M. & Pipa, G. Transfer entropya model-free measure of effective connectivity for the neurosciences. *J. Comput. Neurosci*. **30,** 45–67 (2011).
11. Wibral, M. *et al.* Measuring information-transfer delays. *PloS one*, **8(2),** e55809 (2013).
12. Bauer, M., Cox, J. W., Caveness, M. H., Downs, J. J. & Thornhill, N. F. Finding the direction of disturbance propagation in a chemical process using transfer entropy. *IEEE Trans. Control Syst. Techn.* **15,** 12–21 (2007).
13. Ito, S. & Sagawa, T. Maxwell's demon in biochemical signal transduction with feedback loop. *Nat. Commun.* **6,** 7498 (2015).
14. Hartich, D., Barato, A. C. & Seifert, U. Sensory capacity: An information theoretical measure of the performance of a sensor. *Phys. Rev. E* **93,** 022116 (2016).
15. Ito, S. & Sagawa, T. Information thermodynamics on causal networks. *Phys. Rev. Lett.* **111,** 180603 (2013).
16. Prokopenko, M., Lizier, J. T. & Price, D. C. On thermodynamic interpretation of transfer entropy. *Entropy* **15,** 524–543 (2013).
17. Barnett, L., Lizier, J. T., Harr, M., Seth, A. K. & Bossomaier, T. Information flow in a kinetic Ising model peaks in the disordered phase. *Phys. Rev. Lett.* **111,** 177203 (2013).
18. Hartich, D., Barato, A. C. & Seifert, U. Stochastic thermodynamics of bipartite systems: transfer entropy inequalities and a Maxwell's demon interpretation. *J. Stat. Mech.* P02016 (2014).
19. Prokopenko, M. & Einav, I. Information thermodynamics of near-equilibrium computation. *Phys. Rev. E* **91,** 062143 (2015).
20. Wibral, M., Vicente, R. & Lizier, J. T. (Eds.). *Directed information measures in neuroscience* (Heidelberg: Springer. 2014).
21. Staniek, M. & Lehnertz, K. Symbolic transfer entropy. *Phys. Rev. Lett.* **100,** 158101 (2008).
22. Williams, P. L. & Randall, D. B. "Generalized measures of information transfer." arXiv preprint arXiv:1102.1507 (2011).
23. Granger, C. W., Ghysels, E., Swanson, N. R. & Watson, M. W. "*Essays in econometrics: collected papers of Clive WJ Granger*" (Cambridge University Press 2001).
24. Granger, C. W. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37,** 424438 (1969).
25. Barnett, L., Barrett, A. B. & Seth, A. K. Granger causality and transfer entropy are equivalent for Gaussian variables. *Phys. Rev. Lett.* **103,** 238701 (2009).

26. Marko, H. The bidirectional communication theory-a generalization of information theory. *IEEE Trans. Infom. Theory* **21,** 1345 (1973).
27. Massey, J. Causality, feedback and directed information. *In Proc. Int. Symp. Inf. Theory Applic.* 303–305 (1990).
28. Hirono, Y. & Hidaka, Y. Jarzynski-type equalities in gambling: role of information in capital growth. *J. Stat. Phys.* **161,** 721 (2015).
29. Permuter, H. H., Kim, Y. H. & Weissman, T. On directed information and gambling. In Proc. International Symposium on Information Theory (ISIT), 1403 (2008).
30. Kelly, J. A new interpretation of information rate. Bell Syst. Tech. J. **35,** 917926 (1956).
31. Sagawa, T. & Ueda, M. Generalized Jarzynski equality under nonequilibrium feedback control. *Phys. Rev. Lett.* **104,** 090602 (2010).
32. Parrondo, J. M., Horowitz, J. M. & Sagawa, T. Thermodynamics of information. *Nat. Phys.* **11,** 131–139 (2015).
33. Allahverdyan, A. E., Janzing, D. & Mahler, G. Thermodynamic efficiency of information and heat flow. *J. Stat. Mech.* P09011 (2009).
34. Horowitz, J. M. & Esposito, M. Thermodynamics with continuous information flow. *Phys. Rev. X*, **4,** 031015 (2014).
35. Horowitz, J. M. & Sandberg, H. Second-law-like inequalities with information and their interpretations. *New J. Phys.* **16,** 125007 (2014).
36. Horowitz, J. M. Multipartite information flow for multiple Maxwell demons. *J. Stat. Mech.* P03006 (2015).
37. Ito, S. & Sagawa, T. *"Information flow and entropy production on Bayesian networks"* arXiv: 1506.08519 (2015); a book chapter in M. Dehmer, F. Emmert-Streib, Z. Chen & Y. Shi (Eds.) *"Mathematical Foundations and Applications of Graph Entropy"* (Wiley-VCH Verlag, Weinheim, 2016).
38. Ito, S., *Information thermodynamics on causal networks and its application to biochemical signal transduction*. (Springer Japan, 2016).
39. Shiraishi, N. & Sagawa, T. Fluctuation theorem for partially masked nonequilibrium dynamics. *Phys. Rev. E.* **91,** 012130 (2015).
40. Shiraishi, N., Ito, S., Kawaguchi, K. & Sagawa, T. Role of measurement-feedback separation in autonomous Maxwell's demons. *New J. Phys.* **17,** 045012 (2015).
41. Rosinberg, M. L., Munakata, T. & Tarjus, G. Stochastic thermodynamics of Langevin systems under time-delayed feedback control: Second-law-like inequalities. *Phys. Rev. E* **91,** 042114 (2015).
42. Cafaro, C., Ali, S. A. & Giffin, A. Thermodynamic aspects of information transfer in complex dynamical systems. *Phys. Rev. E*, **93,** 022114 (2016).
43. Yamamoto, S., Ito, S., Shiraishi, N. & Sagawa, T. Linear Irreversible Thermodynamics and Onsager Reciprocity for Information-driven Engines. arXiv:1604.07988 (2016).
44. Sekimoto, K. *Stochastic Energetics* (Springer, 2010).
45. Seifert, U. Stochastic thermodynamics, fluctuation theorems and molecular machines. *Rep. Prog. Phys.* **75,** 126001 (2012).
46. Still, S., Sivak, D. A., Bell, A. J. & Crooks, G. E. Thermodynamics of prediction. *Phys. Rev. Lett.* **109,** 120604 (2012).
47. Still, S. Information bottleneck approach to predictive inference. *Entropy*, **16,** 968–989 (2014).
48. Diana, G. & Esposito, M. Mutual entropy production in bipartite systems. *Journal of Statistical Mechanics: Theory and Experiment*, P04010 (2014).
49. Vinkler, D. A., Permuter, H. H. & Merhav, N. Analogy between gambling and measurement-based work extraction. *J. Stat. Mech.* P043403 (2016).
50. Cover, T. M. & Ordentlich, E. Universal portfolios with side information. *IEEE Trans. Infom. Theory* **42,** 348–363 (1996).
51. Permuter, H. H., Kim, Y. H. & Weissman, T. Interpretations of directed information in portfolio theory, data compression, and hypothesis testing. *IEEE Trans. Inform. Theory* **57,** 3248–3259 (2011).
52. Toyabe, S., Sagawa, T., Ueda, M., Muneyuki, E. & Sano, M. Experimental demonstration of information-to-energy conversion and validation of the generalized Jarzynski equality. *Nat. Phys.* **6,** 988 (2010).
53. Bérut, A., Arakelyan, A., Petrosyan, A., Ciliberto, S., Dillenschneider, R. & Lutz, E. Experimental verification of Landauer's principle linking information and thermodynamics. *Nature* **483,** 187 (2012).
54. Koski, J. V., Maisi, V. F., Sagawa, T. & Pekola, J. P. Experimental observation of the role of mutual information in the nonequilibrium dynamics of a Maxwell demon. *Phys. Rev. Lett.* **113,** 030601 (2014).
55. Kutvonen, A., Koski, J. & Ala-Nissila, T. Thermodynamics and efficiency of an autonomous on-chip Maxwell's demon. *Sci. Rep.* **6,** 21126 (2016).
56. Sartori, P., Granger, L., Lee, C. F. & Horowitz, J. M. Thermodynamic costs of information processing in sensory adaptation. *PLoS Comput Biol*, **10,** e1003974 (2014).
57. Barato, A. C., Hartich, D. & Seifert, U. Efficiency of cellular information processing. *New J. Phys.* **16,** 103024 (2014).
58. Bo, S., Del Giudice, M. & Celani, A. Thermodynamic limits to information harvesting by sensory systems. *J. Stat. Mech.* P01014 (2015).
59. Ouldridge, T. E., Govern, C. C. & Wolde, P. R. T. The thermodynamics of computational copying in biochemical systems. arXiv:1503.00909 (2015).
60. McGrath, T., Jones, N. S., Wolde, P. R. T. & Ouldridge, T. E., A biochemical machine for the interconversion of mutual information and work. arXiv:1604.05474 (2016).

Acknowledgements

We are grateful to Takumi Matsumoto, Takahiro Sagawa, Naoto Shiraishi and Shumpei Yamamoto for the fruitful discussion on a range of issues of thermodynamics. This work was supported by JSPS KAKENHI Grant Numbers JP16K17780 and JP15J07404.

Author Contributions

S.I. constructed the theory, and carried out the analytical calculations, and wrote the paper.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The author declares no competing financial interests.

How to cite this article: Ito, S. Backward transfer entropy: Informational measure for detecting hidden Markov models and its interpretations in thermodynamics, gambling and causality. *Sci. Rep.* **6**, 36831; doi: 10.1038/srep36831 (2016).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.