Minimal entropy production rate of interacting systems

David H Wolpert

1 Santa Fe Institute, Santa Fe, New Mexico, United States of America
2 Complexity Science Hub, Vienna, Austria
3 Arizona State University, Tempe, Arizona, United States of America

∗ Author to whom any correspondence should be addressed.
E-mail: david.h.wolpert@gmail.com

Keywords: Landauer bound, stochastic thermodynamics, multipartite processes, entropy production rate, rate matrix constraints

Abstract

Many systems are composed of multiple, interacting subsystems, where the dynamics of each subsystem only depends on the states of a subset of the other subsystems, rather than on all of them. I analyze how such constraints on the dependencies of each subsystem’s dynamics affects the thermodynamics of the overall, composite system. Specifically, I derive a strictly nonzero lower bound on the minimal achievable entropy production rate of the overall system in terms of these constraints. The bound is based on constructing counterfactual rate matrices, in which some subsystems are held fixed while the others are allowed to evolve. This bound is related to the ‘learning rate’ of stationary bipartite systems, and more generally to the ‘information flow’ in bipartite systems. It can be viewed as a strengthened form of the second law, applicable whenever there are constraints on which subsystem within an overall system can directly affect which other subsystem.

1. Introduction

Many systems are naturally modeled as composite systems, with two or more interacting subsystems. For example, a biological cell is naturally modeled as a composite system, composed of many separate organelles and biomolecule species. As another example, digital devices are naturally modeled as a set of separate, interacting logical gates. Recent research in stochastic thermodynamics [7, 26, 32, 38] has started to investigate such composite systems [1, 9, 11, 13, 20–22]. Most of this research has considered the special case of bipartite processes [1, 2, 9, 11, 13, 20–22, 27, 28], i.e., systems composed of two co-evolving subsystems, whose states fluctuate according to independent noise processes (e.g., since they are physically separated and so are connected to different parts of any shared thermodynamic reservoirs). However, given that many systems have more than just two interacting subsystems, research is starting to extend to fully multipartite processes [10, 13, 40].

The definition of any composite system specifies which subsystems directly affect the dynamics of which other subsystems. It is now known that just by itself, such a specification of which subsystem affects which other one can cause a strictly positive lower bound on the entropy production rate (EP) of the overall composite system [4, 36, 38]. In contrast, if all subsystems were allowed to interact with all other subsystems, the minimal EP would be zero. Accordingly this minimal EP due to constraints on which subsystem can affect the dynamics of which other subsystems has sometimes been called ‘Landauer loss’ [36–38]. Landauer loss can be viewed as a strengthened form of the second law, applicable whenever there are constraints on which subsystem within an overall system can directly affect which other subsystem.

Previous analyses of this strengthened second law focused on scenarios where at every moment, each subsystem evolves in isolation, in a ‘modular’ fashion, without any direct coupling to the other subsystems. This is a severe limitation of those analyses. As an illustration of a simple scenario not covered by such analyses, consider a composite system with three subsystems A, B and C. B evolves independently of A and C. However, B is continually observed by C as well as A. Moreover, suppose that A is really two subsystems,
context) has been called the ‘information flow’ between a pair of subsystems [10, 11]. That expression is also related to what (in a different context) has been called the ‘learning rate’ in the special case of stationary bipartite systems [1, 5, 9].

1 and 2. Only subsystem 2 directly observes B, whereas subsystem 1 observes subsystem 2, e.g., to record a running average of the values of subsystem 2 (see figure 1).

Physically, such a scenario arises whenever any of the many stochastic thermodynamics models of one classical system observing another classical system without any back-action [12, 20, 23, 27, 33, 34] are ‘chained together’. As an example, [3, 9] considers a tripartite system where receptors in the wall of a cell observe the concentration level of a ligand in a surrounding medium, with no back-action on that concentration level, while a memory observes the state of those receptors, again with no back-action. This is exactly the scenario considered in figure 1, just without subsystem 4; subsystem 3 is the concentration level in the medium; subsystem 2 is the set of receptors in a cell observing that concentration level, and subsystem 1 is the memory within the cell observing the state of the ligand receptors. To extend this scenario to the precise scenario presented in figure 1 we just need to introduce a second cell, which is observing the same medium as the first cell; subsystem 4 is the state of the receptors of that second cell.

To investigate the second law in such composite systems, here I model them as multipartite processes, in which each subsystem evolves according to its own rate matrix [10]. So restrictions on the direct coupling of any subsystem to the other subsystems are modeled as restrictions on the rate matrix of subsystem i, to only involve a limited set of other subsystems, called the ‘unit’ of i.

In this paper I derive a lower bound on the EP of composite systems, by deriving an exact equation for that minimal EP rate as a sum of non-negative expressions. One of those expressions is related to quantities that were earlier considered in the literature. It reduces to what has been called the ‘learning rate’ in the special case of stationary bipartite systems [1, 5, 9]. That expression is also related to what (in a different context) has been called the ‘information flow’ between a pair of subsystems [10, 11].

2. Rate matrix units

I write \( \mathcal{N} \) for a particular set of \( N \) subsystems, with finite state spaces \( \{ X_i : i = 1, \ldots, N \} \). x and \( x' \) both indicate a vector in \( X \), the joint space of \( \mathcal{N} \). For any \( A \subset \mathcal{N} \), I write \( -A := \mathcal{N} \setminus A \). So for example \( x \mid A \) is the vector of all components of x other than those in A. A distribution over a set of values x at time t is written as \( p^x(t) \), with its value for \( x \in X \) written as \( p^x_x(t) \), or just \( p_x(t) \) for short. Similarly, \( p^{X|Y}_x(t) \) is the conditional distribution of X given Y at time t, evaluated for the event \( X = x, Y = y \) (which I sometimes shorten to \( p_{x|y}(t) \)). I write Shannon entropy as \( S(p_x(t)) \), \( S(X) \), or \( S(t) \), as convenient. I also write the conditional entropy of X given Y at t as \( S^{XY}(t) \). I write the Kronecker delta as both \( \delta(a, b) \) or \( \delta^a_b \).

Since the joint system evolves as a multi-partite process, there is a set of time-varying stochastic rate matrices, \( \{ K^X_x(i; t) : i = 1, \ldots, N \} \), where \( x' \) refers to the current state of the system, while x refers to the next state; for all i, \( K^X_x(i; t) = 0 \) if \( x'_i \neq x_i \); and the joint dynamics over X is governed by the master equation.
\[
\frac{dp_x(t)}{dt} = \sum_{x'} K^{x'}_x(t)p_{x'}(t) \\
= \sum_{x'} \sum_{i \in N} K^{x'}_x(i; t)p_{i'}(t).
\]

Note that each subsystem can be driven by its own external work reservoir, according to a time-varying protocol. For any \( A \subseteq N \) I define
\[
K^A_x(A; t) := \sum_{i \in A} K^{x'}_x(i; t).
\]
Each subsystem \( i \)'s marginal distribution evolves as
\[
\frac{dp_x(t)}{dt} = \sum_{x', i} \sum_{j} K^{x'}_x(j; t)p_{x'}(t) \\
= \sum_{x'} \sum_{i \in A} K^{x'}_x(i; t)p_{x'}(t)
\]
due to the multipartite nature of the process \([15]\). Equation (5) shows that in general the marginal distribution \( p_x \) will not evolve according to a continuous-time Markov chain (CTMC) over \( \Delta X_x \).

For any subsystem \( i \), I write \( r(i; t) \) for any set of subsystems at time \( t \) that includes \( i \) where we can write
\[
K^{x'}_x(i; t) = K^{x'}_{\delta r(i)}(i; t)\delta(x'_{\delta r(i)}, x_{\delta r(i)}).
\]
for an appropriate set of functions \( K^{x'}_{\delta r(i)}(i; t) \). In general, \( r(i; t) \) is not uniquely defined, since I make no requirement that it be minimal. (A minimal such \( r(i; t) \) is called a ‘neighborhood’ in \([10]\).) I refer to the elements of \( r(i; t) \) as the leaders of \( i \) at time \( t \). Note that the leader relation need not be symmetric. A unit \( \omega \) at time \( t \) is a set of subsystems such that \( i \in \omega \) implies that \( r(i; t) \subseteq \omega \).

Any intersection of two units is a unit, as is any union of two units. Intuitively, a unit is any set of subsystems whose evolution is independent of the states of the subsystems outside the unit (although in general, the evolution of those external subsystems may depend on the states of subsystems in the unit). A specific set of units that covers \( N \) and is closed under intersections is a unit structure. Unless explicitly stated otherwise, any unit structure being discussed does not have \( N \) itself as a member.

As an example of these definitions, \([1, 8, 9]\) investigate a special type of bipartite system, where the ‘internal’ subsystem \( B \) observes the ‘external’ subsystem \( A \), but cannot affect the dynamics of that external subsystem. So \( A \) is its own unit, evolving independently of \( B \), while \( B \) is not its own unit; its dynamics depend on the state of \( A \) as well as its own state. Another example of these definitions is illustrated in figure 1.

For simplicity, from now on I assume that the set of units does not change with \( t \). Accordingly I shorten \( r(i; t) \) to \( r(i) \). For any unit \( \omega \) I write
\[
K^{x'}_x(\omega; t) := \sum_{i \in \omega} K^{x'}_x(i; t).
\]
So \( K^{x'}_x(\omega; t) = K^{x'}_{\Delta x}(\omega; t)\delta(x'_{\Delta x}, x_{\Delta x}) \), by equations (3) and (6).

At any time \( t \), for any unit \( \omega \), \( p_{x\omega}(t) \) evolves as a CTMC with rate matrix \( K^{x'}_x(\omega; t) \):
\[
\frac{dp_{x\omega}(t)}{dt} = \sum_{x'} K^{x'}_x(\omega; t)p_{x\omega}(t).
\]
(See appendix C.) So a unit evolves according to a self-contained CTMC, in contrast to the general case of a single subsystem (cf. equation (5)).

I assume that each subsystem is attached to at most one thermal reservoir, and that all such reservoirs have the same temperature (or equivalently, that all subsystems are attached to separate, statistically independent parts of the same thermal reservoir \([10]\)). Accordingly, the expected entropy flow (EF) rate of any unit \( \omega \subseteq N \) at time \( t \) is the sum of the EFs of the subsystems in \( \omega \):
\[
\langle \dot{Q}^{x\omega}(t) \rangle = \sum_{x', \omega} K^{x'}_x(\omega; t)p_{x\omega}(t)\ln \left[ \frac{K^{x'}_x(\omega; t)}{K^{x'}_x(\omega; t)} \right]
\]
(9)
which I often short to \(\langle \dot{Q}^w(t) \rangle \) [7, 32]. (Note that this is EF from \(\omega\) into the environment.) Make the associated definition that the expected EP rate of \(\omega\) at time \(t\) is

\[
\langle \dot{\sigma}^{\omega;K}(t) \rangle = \frac{d\hbar\omega(t)}{dt} + \langle \dot{Q}^w(t) \rangle \\
= \sum_{x',x} K_x^{x'}(\omega; t)p_{x'}(t) \ln \left( \frac{K_x^{x'}(\omega; t)p_{x'}(t)}{K_x^{x'}(\omega; t)p_x(t)} \right)
\]

which I often short to \(\langle \dot{\sigma}^{\omega}(t) \rangle \).

I refer to \(\langle \dot{\sigma}^{\omega}(t) \rangle\) as a local EP rate, and define the global EP rate as \(\langle \dot{\sigma}(t) \rangle := \langle \dot{\sigma}^{\omega}(t) \rangle\). For any unit \(\omega\), \(\langle \dot{\sigma}^{\omega}(t) \rangle \geq 0\), since \(\langle \dot{\sigma}^{\omega}(t) \rangle\) has the usual form of an EP rate of a single system. In addition, that lower bound of 0 is achievable, e.g., if \(K_x^{x'}(\omega; t)p_{x'}(t) = K_x^{x'}(\omega; t)p_x(t)\) at time \(t\) for all \(x, x'\). Note also that while the EP of a unit is the sum of the EFs of the subsystems within the unit, in general that is not true for the EP.

It is worth comparing the local EP rate to similar quantities that have been investigated in the literature. In contrast to \(\langle \dot{\sigma}^{\omega}(t) \rangle\), the quantity \(\langle \sigma_X \rangle\) introduced in the analysis of (autonomous) bipartite systems in [28] is the EP of a single trajectory, integrated over time. More importantly, its expectation can be negative, unlike (the time-integration of) \(\langle \dot{\sigma}^{\omega}(t) \rangle\). On the other hand, the quantity \('S^X_t\)' considered in the analysis of bipartite systems in [11] is a proper expected EP rate, and so is non-negative. However, it (and its extension considered in [10]) is one term in a decomposition of the expected EP rate generated by a single unit. It does not concern the EP rate of an entire unit in a system with multiple units. In addition, the quantity \(\langle \sigma_{\Omega_1} \rangle\) considered in [28] is also non-negative. However, it gives the total EP rate generated by a subset of all possible global state transitions, rather than the EP rate of a unit [16]. Finally, papers on the ‘thermodynamics of information’ considers scenarios in which two subsystems interact but the second one does not change its state [20, 23, 25]. Most of those papers use a variant of the usual definition of ‘entropy production’ whose expectation can be strictly negative, in contrast to \(\langle \dot{\sigma}^{\omega}(t) \rangle\) [17].

3. EP bounds from counterfactual rate matrices

To analyze the minimal EP rate in multipartite processes, I need to introduce two more definitions. First, given any function \(f : \Delta_X \to \mathbb{R}\) and any \(A \subset \mathcal{N}\) (not necessarily a unit), define the \(A\)-windowed derivative of \(f(p(t))\) under rate matrix \(K(t)\) as

\[
\frac{d^A;K(t)}{dt}f(p(t)) = \sum_{x,x'} K_x^{x'}(A; t)p_{x'}(t) \frac{\partial f(p(t))}{\partial p_{x}(t)}.
\]

(See equation (3).) Intuitively, this is what the derivative of \(f(p(t))\) would be if (counterfactually) only the subsystems in \(A\) were allowed to change their states.

In particular, the \(A\)-derivative of the conditional entropy of \(X\) given \(X_A\) is

\[
\frac{d^A;K(t)}{dt}S^{\mathcal{X}_A}(p(t)) = -\sum_{x,x'} K_x^{x'}(A; t)p_{x'}(t) \ln p_{x|x_A}(t)
\]

which I sometimes write as just \(d^A;S^{\mathcal{X}_A}(p(t))/dt\). \(d^A;S^{\mathcal{X}_A}(p(t))/dt\) measures how quickly the statistical coupling between \(X_A\) and \(X_{-A}\) changes with time, if rather than evolving under the actual rate matrix, the system evolved under a counterfactual rate matrix, in which \(X_{-A}\) is not allowed to change. In the special case of two subsystems, both leading each other, with \(A\) one of those subsystems, \(-d^A;S^{\mathcal{X}_A}(p(t))/dt\) is the same as the ‘information flow’ analyzed in [11]. (See equation (4) in [10] for a generalization of information flow to multiple subsystems that is similar to the general expression \(-d^A;S^{\mathcal{X}_A}(p(t))/dt\).)

\(S^{\mathcal{X},-A}(t)\) does not change under the counterfactual rate matrix \(K(A; t)\). Accordingly, \(\frac{d^A;S^{\mathcal{X}_A}(p(t))}{dt}\) is the derivative of the negative mutual information between \(X_A\) and \(X_{-A}\), under that counterfactual rate matrix. In appendix C it is shown that in the special case that \(A\) is a unit, this expression is non-negative [18].

The second definition we need is a variant of \(\langle \dot{\sigma}^{\omega;K}(t) \rangle\), which will be indicated by using subscripts rather than superscripts. For any \(A \subseteq B \subseteq \mathcal{N}\) where \(B\) is a unit (but \(A\) need not be),

\[
\langle \dot{\sigma}_{K(A;B)} \rangle := \sum_{x,x'} K_x^{x'}(A; t)p_{x'}(t) \ln \left( \frac{K_x^{x'}(A; t)p_{x'}(t)}{K_x^{x'}(A; t)p_x(t)} \right)
\]
which I abbreviate as \( \langle \hat{\sigma}_{K(A,t)}(t) \rangle \) when \( B = N \). \( \langle \hat{\sigma}_{K(A,t)}(t) \rangle \) is a global EP rate, only evaluated under the counterfactual rate matrix \( K(A,t) \). Therefore it is non-negative. In contrast, \( \langle \hat{\sigma}^{\omega,K}(t) \rangle \) is a local EP rate. In the special case that \( A = \omega \) is a unit, these two EP rates are related by \( \langle \hat{\sigma}_{K(A,t)}(t) \rangle = \langle \hat{\sigma}^{\omega,K}(t) \rangle + \frac{d^\omega}{dt} S^{\omega,X_4}(t) \) (see equation (B3) in appendix B).

In appendix D it is shown that for any pair of units, \( \omega \) and \( \omega' \subset \omega \),

\[
\langle \hat{\sigma}^{\omega,K(\omega,t)}(t) \rangle = \langle \hat{\sigma}^{\omega',K(\omega,t)}(t) \rangle + \langle \hat{\sigma}_{K(\omega',\omega;\omega,t)}(t) \rangle + \frac{d^\omega}{dt} S^{\omega,X_4}(t).\tag{15}
\]

(See figure 1 for an illustration of such a pair of units \( \omega, \omega' \subset \omega \).) The first term on the rhs is the EP rate arising from the subsystems within unit \( \omega' \), and the second term is the 'left over' EP rate from the subsystems that are in \( \omega \) but not in \( \omega' \). The third term is a time-derivative of the conditional entropy between those two sets of subsystems. All three of these terms are non-negative, so each of them provides a lower bound on the EP rate.

Equation (15) is the major result of this paper. It holds no matter what the scale of the full system, so long as that system can be modeled as a multipartite process. In particular, setting \( \omega = N \) and then consolidating notation by rewriting \( \omega' \) as \( \omega \), equation (15) shows that for any unit \( \omega \),

\[
\langle \hat{\sigma}(t) \rangle = \langle \hat{\sigma}^\omega(t) \rangle + \langle \hat{\sigma}_{K(\omega',\omega;\omega,t)}(t) \rangle + \frac{d^\omega}{dt} S^{\omega,X_4}(t)\tag{16}
\]

\[
\geq \frac{d^\omega}{dt} S^{\omega,X_4}(t)\tag{17}
\]

(where the shorthand notation has been used). This lower bound on the expected EP rate can be evaluated without knowing any of the detailed physics occurring within unit \( \omega \), only knowing how the statistical coupling between \( \omega \) and the rest of the subsystems evolves with time.

As an example of equation (17), consider again the type of bipartite process analyzed in [1, 8, 9]. Suppose we set \( \omega \) to contain only what in [8] is called the 'external' subsystem. Then if we also make the assumption of those papers that the full system is in a stationary state, \( dS^X/dt = dS^\omega/dt = dS^{\omega'}/dt = 0 \). Moreover,

\[
\frac{dS^X}{dt} = -\sum_{x,x'} K^x_{x'}(\omega;t)p_x(t) \ln p_x(t) - \sum_{x,x'} K^x_{x'}(-\omega;t)p_x(t) \ln p_x(t).\tag{18}
\]

Therefore by equation (13),

\[
\frac{d^\omega}{dt} S^{\omega,X_4}(t) = -\frac{d^\omega}{dt} S^{\omega,X_4}(t).\tag{19}
\]

(The rhs of equation (19) is called the 'learning rate' of the internal subsystem about the external subsystem—see equation (8) in [5], noting that the rate matrix is normalized.)

So in this scenario, equation (17) above reduces to equation 7 of [1], which lower-bounds the global EP rate by the learning rate. However, equation (17) lower-bounds the global EP rate even if the system is not in a stationary state, which need be the case with the learning rate [19]. More generally, equation (16) applies to arbitrary multipartite processes, not just those with two subsystems, and is an exact equality rather than just a bound.

4. Extensions

In some situations we can get an even more refined decomposition of EP rate by substituting equation (15) into equation (16) to expand the first EP rate on the rhs of equation (16). This gives a larger lower bound on \( \langle \hat{\sigma}(t) \rangle \) than the one in equation (17). For example, if \( \omega \) and \( \omega' \subset \omega \) are both units under \( K(t) \), then

\[
\langle \hat{\sigma}(t) \rangle = \langle \hat{\sigma}^\omega(t) \rangle + \langle \hat{\sigma}_{K(\omega',\omega;\omega,t)}(t) \rangle + \frac{d^\omega}{dt} S^{\omega,X_4}(t) + \langle \hat{\sigma}_{K(\omega',\omega;\omega,t)}(t) \rangle + \frac{d^\omega}{dt} S^{\omega,X_4}(t)\tag{20}
\]

\[
\geq \frac{d^\omega}{dt} S^{\omega,X_4}(t) + \frac{d^\omega}{dt} S^{\omega,X_4}(t).\tag{21}
\]

Both of the terms on the rhs in equation (21) are non-negative. In addition, both can be evaluated without knowing the detailed physics occurring within units \( \omega \) or \( \omega' \), only knowing how the statistical coupling between units evolves with time.

This can be illustrated with the scenario depicted in figure 1. Using the units \( \omega \) and \( \omega' \) specified there, equation (21) says that the global EP rate is lower-bounded by the sum of two terms. The first is the derivative of the negative mutual information between subsystem 4 and the first three subsystems, if
subsystem 4 were held fixed. The second is the derivative of the negative mutual information between subsystem 3 and the first two subsystems, if those two subsystems were held fixed.

Alternatively, suppose that \( \omega \) is a unit under \( K \), and that some set of subsystems \( \alpha \) is a unit under \( K(\mathcal{N} \setminus \omega; t) \). Then since the term \( \langle \dot{\sigma}_{K(\mathcal{N} \setminus \omega; t)}(t) \rangle \) in equation (16) is a global EP rate over \( \mathcal{N} \) under rate matrix \( K(\mathcal{N} \setminus \omega; t) \), we can again feed equation (15) into equation (16), [this time to expand the second rather than first term on the rhs of equation (16)] to get

\[
\langle \dot{\sigma}(t) \rangle = \langle \dot{\sigma}^{\omega}_{K(\mathcal{N} \setminus \omega)}(t) \rangle + \frac{d}{dt} S_{X(\omega)}(t) + \frac{d}{dt} S_{X_{\mathcal{N} \setminus \omega}}(t) + \langle \dot{\sigma}^{\alpha,K(\mathcal{N} \setminus \omega; t)}(t) \rangle + \frac{d}{dt} S_{X_{\mathcal{N} \setminus \omega}}(t) \]

\[
\geq \frac{d}{dt} S_{X(\omega)}(t) + \frac{d}{dt} S_{X_{\mathcal{N} \setminus \omega}}(t) + \langle \dot{\sigma}^{\alpha,K(\mathcal{N} \setminus \omega; t)}(t) \rangle \]

(22)

The rhs of equation (23) also exceeds the bound in equation (17), by the negative \( \alpha \)-derivative of the mutual information between \( X_{\mathcal{N} \setminus \omega} \) and \( X_{\omega} \), under the rate matrix \( K(\mathcal{N} \setminus \omega; t) \).

Note that depending on the full unit structure, we may be able to combine equations (15) and (22) into an even larger lower bound on the global EP rate than equation (23). An example of this is illustrated below, in section 5. (Indeed, the more subsystems the overall system contains, the more times one might be able to iterate this process, getting progressively larger and larger lower bounds.)

As a final comment, \( \langle \dot{\sigma} \rangle \) is just the conventional expected EP rate of stochastic thermodynamics [32]. So it obeys all of the usual fluctuation theorems, thermodynamic uncertainty relations, speed-limit theorems, etc. directly, without any modification. (As mentioned at the end of section 2, this is not true for some of the other quantities referred to as ‘entropy production’ in the literature.) So for example,

As a final comment, \( \langle \dot{\sigma} \rangle \) is the standard expected EP considered in the literature, i.e., the rate of increase of entropy in the combination of the full system and all baths that are connected to it. As a result, the powerful results of stochastic thermodynamics concerning entropy production—fluctuation theorems, thermodynamic uncertainty relations, speed-limit theorems, etc.—all hold when the EP is set to \( \langle \dot{\sigma} \rangle \).

(Those results do not all hold for some of the variants of entropy production discussed at the end of section 2.) As an illustration, write the EP of unit \( \omega \) during some time interval as \( \Delta \sigma^\omega = \int \dot{\sigma}^\omega \, dt \), and define \( \Delta \sigma_{K(\mathcal{N} \setminus \omega)} \) similarly. Then plugging equation (17) into the usual integral fluctuation theorem gives

\[
\left\langle \exp \left( - \Delta \sigma^\omega - \Delta \sigma_{K(\mathcal{N} \setminus \omega)} - \int dt \frac{d}{dt} S_{X(\omega)}(t) \right) \right\rangle = 1
\]

(24)

(recall that the windowed derivative in the integrand in the exponential is not a proper time-derivative).

Using the non-negativity of EP during the process, equation (25) gives

\[
\left\langle \exp \left( - \int dt \frac{d}{dt} S_{X(\omega)}(t) \right) \right\rangle \geq 1.
\]

This inequality characterizes how the mutual information between \( \omega \) and the rest of the system can vary during the process, independent of all details of the physical process besides the fact that \( \omega \) is a unit.

5. Example

To illustrate some of the results above, return to the physical scenario depicted in figure 1, in which there are two distinct cells in a medium both observing the concentration level of a ligand in that medium. Recall that subsystem 3 is the concentration level, subsystem 2 is the set of receptors in the first cell observing that concentration level, and subsystem 1 is the memory within that first cell which observing the state of that cell’s ligand receptors. Subsystem 4 is the set of receptors in the second cell that are observing the same concentration level. (A simple variant would involve only a single cell, where subsystem 4 is a different set of receptors observing the concentration level.)

Take \( \omega = \{1, 2, 3\} \) and \( \alpha = \{3, 4\} \), as indicated in figure 1. So \( \omega \) is the joint state of the concentration level, the receptors in the first cell, and the memory in the first cell. \( \alpha \) instead is the joint state of the same concentration level, along with the state of the receptors in the second cell.

Note that the four sets \( \{1, 2\}, \{3\}, \{3, 4\} \) form a unit structure of

\[
K_{\omega}^{\mathcal{N}}(\mathcal{N} \setminus \omega; t) = K_{\omega}^{\mathcal{N}}(\{4\}; t)
\]

(26)
since in fact, under the rate matrix $K^X_t(\{4\};t)$, neither subsystem 1, 2 nor 3 changes its state. So $\alpha$ is a member of a unit structure of $K(\mathcal{N} \setminus \omega; t)$, and we can apply equation (22).

The first term in equation (22), $\langle \dot{\sigma}(t) \rangle$, is the local EP rate that would be jointly generated by the set of three subsystems $\{1, 2, 3\}$, if they evolved in isolation from the other subsystem, under the self-contained rate matrix

$$K^X_t(\{1, 2, 3\}; t) = K_{\{1,2,3\}}^X(1,t) + K_{\{1,2,3\}}^X(2,t) + K_{\{1,2,3\}}^X(3,t).$$  

(27)

The third term in equation (22) is the local EP rate that would be jointly generated by the two subsystems $\{3, 4\}$, if they evolved in isolation from the other two subsystems, but rather than do so under the rate matrix $K(\alpha; t) = K(\{3, 4\}; t)$, they did so under the rate matrix $K^X_t(\mathcal{N} \setminus \omega; t)$ given in equation (26). (Note that $K^X_t(\mathcal{N} \setminus \alpha; t) = 0$ if $x'_3 \neq x_3$, unlike $K^X_t(\{3, 4\}; t)$.) The fourth term in equation (22) is the global EP rate that would be generated by evolving all four subsystems under the rate matrix for the subsystems in $\mathcal{N} \setminus \omega \setminus \alpha$. But there are no subsystems in that set. So this fourth term is zero.

Those first, third and fourth terms in equation (22) are all non-negative. The remaining two terms, the second and the fifth, are also non-negative. However, in contrast to the terms just discussed, these two depend only on derivatives of mutual informations. Specifically, the second term in equation (22) is the negative derivative of the mutual information between the joint random variable $X_{1,2,3}$ and $X_4$, under the rate matrix $K^X_t(\{1, 2, 3\}; t)$. Next, since $\mathcal{N} \setminus \alpha = \{1, 2\}$, the fifth term is the negative derivative of the mutual information between $X_{1,2}$ and $X_{3,4}$, under the rate matrix given by windowing $\alpha$ onto $K(\mathcal{N} \setminus \omega; t)$, i.e., under the rate matrix $K^X_t(\{4\}; t)$.

Recalling that $\omega := \{1, 2, 3\}, \alpha := \{3, 4\}$ and defining $\gamma := \{4\}$, we can combine these results to express the global EP rate of the system illustrated in figure 1 in terms of the rate matrices of the four subsystems:

$$\langle \dot{\sigma}(t) \rangle = \sum_{x,\omega} K^X_t(\omega; t)p_{x\omega}(t) \ln \frac{K^X_t(\omega; t)p_{x\omega}(t)}{K^X_t(\omega; t)p_{x\omega}(t)} + \sum_{x,\alpha,\gamma} K^X_t(\gamma; t)p_{\omega\gamma}(t) \ln \frac{K^X_t(\gamma; t)p_{\omega\gamma}(t)}{K^X_t(\gamma; t)p_{\omega\gamma}(t)}$$

$$- \sum_{x,x'} \left[ K^X_t(\omega; t)p_{x\omega}(t) \ln p_{x\omega}(t) + K^X_t(\gamma; t)p_{x\gamma}(t) \ln p_{x\gamma}(t) \right].$$

(28)

All four terms on the rhs of equation (28) are non-negative. Translated to this scenario, previous results concerning learning rates consider the special case of a stationary state $p_{x}(t)$, and only tell us that the global EP rate is bounded by the fourth term on the rhs of equation (28):

$$\langle \dot{\sigma}(t) \rangle \geq - \sum_{x,x'} K^X_t(\omega; t)p_{x\omega}(t) \ln p_{x\omega}(t).$$

(29)

Finally note that we also have a unit $\omega' = \{3\}$ which is a proper subset of both $\omega$ and $\alpha$. So, for example, we can plug this $\omega'$ into equation (15) to expand the first term in equation (22), $\langle \dot{\sigma}(K(\omega'; t))(t) \rangle$, replacing it with the sum of three terms. The first of these three new terms, $\langle \dot{\sigma}(K(\omega'; t))(t) \rangle$, is the local EP rate generated by subsystem $\{3\}$ evolving in isolation from all the other subsystems. The second of these new terms, $\langle \dot{\sigma}_{K(\omega'; t)}(\omega(t))(t) \rangle$, is the EP rate that would be generated if the set of three subsystems $\{1, 2, 3\}$ evolved in isolation from the remaining subsystem, 4, but under the rate matrix

$$K(\omega' \setminus \omega; t) = K_{\{1,2,3\}}^X(1,t) + K_{\{1,2,3\}}^X(2,t) + K_{\{1,2,3\}}^X(3,t).$$

(30)

The third new term is the negative derivative of the mutual information between $X_{1,2}$ and $X_{3,4}$, under rate matrix $K(\omega'; t)$. All three of these new terms are non-negative.

6. **Discussion**

There are other decompositions of the global EP rate which are of interest, but do not always provide non-negative lower-bounds on the EP rate. One of them, discussed in appendix E, generalizes the results in [38] which relate ‘subsystem Landauer loss’ to multi-information. Future work involves combining these (and other) decompositions, to get even larger lower bounds.
Acknowledgments

I would like to thank Sosuke Ito, Artemy Kolchinsky, and Alec Boyd for stimulating discussion. This work was supported by the Santa Fe Institute, Grant No. CHE-1648973 from the US National Science Foundation and Grant No. FQXi-RFP-IPW-1912 from the FQXi foundation. The opinions expressed in this paper are those of the author and do not necessarily reflect the view of the National Science Foundation.

Appendix A. Proof of equation (8)

Write

$$\frac{dp_{x}(t)}{dt} = \sum_{x' \in \omega} \sum_{j} K^{x'}_{x}(j; t)p_{x'}(t)$$

$$= \sum_{x'} p_{x'}(t) \left[ \sum_{j \in \omega \setminus x'} K^{x'}_{x}(j; t) + \sum_{j \in \omega \setminus x_{\omega}} K^{x'}_{x}(j; t) \right]. \quad (A1)$$

If $j \notin \omega$, then a sum over all $x_{\omega}$ in particular runs over all $x_{j}$. Therefore we get

$$\frac{dp_{x}(t)}{dt} = \sum_{x'} p_{x'}(t) \sum_{x_{\omega}} K^{x'}_{x}(j; t). \quad (A2)$$

Using the fact that we have a multipartite process and then the fact that $\omega$ is a unit, we can expand this remaining expression as

$$\sum_{x'} \sum_{j \in \omega} K^{x'}_{x}(j; t)p_{x'}(t) = \sum_{x'} \sum_{j \in \omega} K^{x'}_{x}(j; t)p_{x'}(t)$$

$$= \sum_{x_{\omega}} \sum_{j \in \omega} K^{x'}_{x}(j; t)p_{x'}(t). \quad (A3)$$

To complete the proof plug in the definition of $K^{x'}_{x}(\omega; t)$.

Appendix B. Expansions of EP rates in multipartite processes

This appendix derives an expansion of EP rates that is used in the main text and the other appendices.

Lemma 1. Suppose we have a multipartite process over a set of systems $\mathcal{N}$ defined by a set of rate matrices $\{K^{x'}_{x}(i; t)\}$ and a subset $A \in \mathcal{N}$. Then

$$\sum_{x, x'} K^{x'}_{x}(A; t)p_{x'}(t) \ln \left( \frac{K^{x'}_{x}(t)}{K^{x'}_{x}(t)} \right) = \sum_{x, x'} K^{x'}_{x}(A; t)p_{x'}(t) \ln \left( \frac{K^{x'}_{x}(A; t)}{K^{x'}_{x}(A; t)} \right)$$

$$= \sum_{i \in A, x, x'} K^{x'}_{x}(i; t)p_{x'}(t) \ln \left( \frac{K^{x'}_{x}(i; t)}{K^{x'}_{x}(i; t)} \right). \quad (B1)$$

If in addition $A$ is a unit under $K$, then we can also write the quantity in equation (B1) as

$$\sum_{x, x'} K^{x'}_{x}(A; t)p_{x'}(t) \ln \left( \frac{K^{x'}_{x}(A; t)}{K^{x'}_{x}(A; t)} \right). \quad (B3)$$

Proof. Invoking the multipartite nature of the process allows us to write
\[
\sum_{x,x'} K'_{\Delta}^x(A; t) \delta_{x',a} p_{\Delta}(t) \ln \left[ \frac{K'_{\Delta}^x(t)}{K'_{\Delta}(t)} \right]
\]
\[
= \sum_{i \in A \cap x'} K'_{\Delta}^x(i; t) p_{\Delta}(t) \ln \left[ \frac{K'_{\Delta}^x(t)}{K'_{\Delta}(t)} \right]
\]
\[
= \sum_{i \in A \cap x', x \neq x'} K'_{\Delta}^{x',i-}(i; t) p_{\Delta}(t) \ln \left[ \frac{\sum_{j \in A} K'_{\Delta}^{x',i-}(j; t)}{K'_{\Delta}^{x',i-}(i; t)} \right]
\]
\[
+ \sum_{i \in A \cap x_i} K'_{\Delta}^{x',i-}(i; t) p_{\Delta}(t) \ln \left[ \frac{\sum_{j \in A} K'_{\Delta}^{x',i-}(j; t)}{K'_{\Delta}^{x',i-}(i; t)} \right]
\]
\[
= \sum_{i \in A \cap x_i} K'_{\Delta}^{x',i-}(i; t) p_{\Delta}(t) \ln \left[ \frac{K'_{\Delta}^{x',i-}(i; t)}{\sum_{j \in A} K'_{\Delta}^{x',i-}(j; t)} \right]. \quad \text{(B4)}
\]
\[
= \sum_{i \in A \cap x_i} K'_{\Delta}^{x',i-}(i; t) p_{\Delta}(t) \ln \left[ \frac{K'_{\Delta}^{x',i-}(i; t)}{K'_{\Delta}^{x',i-}(i; t)} \right]. \quad \text{(B5)}
\]

Equation (B4) establishes equation (B1) and equation (B5) establishes equation (B2).

To establish equation (B3), use the hypothesis that \( A \) is a unit to expand

\[
\sum_{x,x'} K'_{\Delta}^x(A; t) p_{\Delta}(t) \ln \left[ \frac{K'_{\Delta}^x(A; t)}{K'_{\Delta}(A; t)} \right] = \sum_{x,x'} K'_{\Delta}^x(A; t) \delta_{x',a} p_{\Delta}(t) \ln \left[ \frac{K'_{\Delta}^x(A; t) \delta_{x',a}}{K'_{\Delta}(A; t) \delta_{x',a}} \right] \quad \text{(B6)}
\]
\[
= \sum_{x,x'} K'_{\Delta}^x(A; t) \delta_{x',a} p_{\Delta}(t) \ln \left[ \frac{K'_{\Delta}^x(A; t)}{K'_{\Delta}(A; t)} \right] \quad \text{(B7)}
\]
\[
= \sum_{x_i, x \neq x_i} K'_{\Delta}^{x,i-}(A; t) p_{\Delta}(t) \ln \left[ \frac{K'_{\Delta}^{x,i-}(A; t)}{K'_{\Delta}^{x,i-}(A; t)} \right] \quad \text{(B8)}
\]
\[
= \sum_{x_i} K'_{\Delta}^{x,i-}(A; t) p_{\Delta}(t) \ln \left[ \frac{K'_{\Delta}^{x,i-}(A; t)}{K'_{\Delta}^{x,i-}(A; t)} \right]. \quad \text{(B9)}
\]

\[\square\]

Appendix C. Proof that if \( A \) is a unit, then \( \frac{\partial}{\partial t} \delta_{x_i} x(t) \geq 0 \)

First simplify notation by using \( P \) rather than \( p \) to indicate joint distributions that would evolve if \( K(t) \) were replaced by the counterfactual rate matrix \( K(A; t) \), starting from \( p_{\Delta}(t) \). By definition,

\[
K_{x_i(t), x \neq x_i(t)}^{x_i(t+t+\delta t)}(A; t) = \lim_{\delta t \to 0} \frac{P \left( x_i(t + \delta t), x \neq x_i(t + \delta t) | x_i(t), x \neq x_i(t) \right) - \delta_{x_i(t), x \neq x_i(t+t+\delta t)}(A; t)}{\delta t}. \quad \text{(C1)}
\]

However, since by hypothesis \( A \) is a unit,

\[
K_{x_i(t), x \neq x_i(t)}^{x_i(t)}(A; t) = K_{x_i(t+t+\delta t), x \neq x_i(t+t+\delta t)}^{x_i(t+t+\delta t)}(A; t), \quad \text{(C2)}
\]
Plugging this into equation (C1), summing both sides over \( x_A(t + \delta t) \), and using the normalization of \( K(A; t) \) shows that to leading order in \( \delta t \),

\[
    P\left( x_A(t + \delta t) | x_A(t), x_{-A}(t) \right) = \delta_{x_A(t + \delta t)}^{x_A(t)},
\]

(C3)

Equation (C3) in turn implies that to leading order in \( \delta t \),

\[
    P\left( x(t + \delta t) | x(t) \right) = P\left( x_A(t + \delta t) | x_A(t), x_{-A}(t) \right) P\left( x_{-A}(t + \delta t) | x_{-A}(t) \right)
\]

(C4)

\[
    = P\left( x_A(t + \delta t) | x_A(t), x_{-A}(t), x_{-A}(t + \delta t) = x_{-A}(t) \right) \delta_{x_A(t + \delta t)}^{x_A(t)}.
\]

(C5)

This formalizes the statement in the text that under the rate matrix \( K(A) \), \( x_{-A} \) does not change its state.

Next, since \( A \) is a unit under \( K(A; t) \), we can expand further to get

\[
    P( x_A(t + \delta t), x_{-A}(t + \delta t) | x_A(t), x_{-A}(t) ) = P\left( x_A(t + \delta t) | x_A(t) \right) \delta_{x_{-A}(t + \delta t)}^{x_{-A}(t)}.
\]

(C6)

So the full joint distribution is

\[
    P( x_A(t + \delta t), x_{-A}(t + \delta t), x_A(t), x_{-A}(t) ) = P\left( x_A(t + \delta t) | x_A(t) \right) \delta_{x_{-A}(t + \delta t)}^{x_{-A}(t)} P( x_A(t), x_{-A}(t) ).
\]

(C7)

I can use this form of the joint distribution to establish the following two equations

\[
    S_P( X_{-A}(t + \delta t) | X_{-A}(t), X_A(t)) = 0 \quad \text{(C8)}
\]

\[
    S_P( X_{-A}(t) | X_{-A}(t + \delta t), X_A(t + \delta t) ) = 0. \quad \text{(C9)}
\]

Applying the chain rule for entropy to decompose \( S_P( X_{-A}(t), X_{-A}(t + \delta t) | X_A(t + \delta t) ) \) in two different ways, and plugging equations (C8) and (C9), respectively, into those two decompositions, we see that

\[
    S_P( X_{-A}(t + \delta t) | X_A(t + \delta t) ) = S_P( X_{-A}(t) | X_A(t) ). \quad \text{(C10)}
\]

Next, use equation (C10) to expand

\[
    \frac{d^{A,K}}{dt} S_X(t) = - \lim_{\delta t \to 0} \left[ S_P( X_{-A}(t + \delta t) | X_A(t) ) + S_P( X_A(t + \delta t) | X_{-A}(t) ) \right] \frac{\delta t}{\delta t}
\]

(C11)

\[
    = - \lim_{\delta t \to 0} \left[ S_P( X_{-A}(t + \delta t) | X_A(t) ) + S_P( X_A(t + \delta t) | X_{-A}(t) ) \right] \frac{\delta t}{\delta t}.
\]

(C12)

Add and subtract \( S_P( X_{-A}(t) ) \) in the numerator on the rhs to get

\[
    \frac{d^{A,K}}{dt} S_X(t) = - \lim_{\delta t \to 0} \left[ I_P( X_{-A}(t + \delta t) | X_A(t) ) + I_P( X_A(t + \delta t) | X_{-A}(t) ) \right] \frac{\delta t}{\delta t}.
\]

(C13)

Since \( X_{-A}(t) \) and \( X_A(t + \delta t) \) are conditionally independent given \( X_A(t) \), we have a Markov chain

\[
    X_{-A}(t) \leftrightarrow X_A(t) \leftrightarrow X_A(t + \delta t)
\]

and we can apply the data-processing inequality [6] to establish that the difference of mutual informations in the numerator on the rhs of equation (C13) is non-positive.

This completes the proof.

**Appendix D. Proof of equation (15)**

For simplicity of the exposition, treat \( \omega \) as though it were all of \( \mathcal{X} \), i.e., suppress the \( \omega \) index in \( x_{\omega} \) and \( x'_{\omega} \), suppress the \( \omega \) argument of \( K(\omega; t) \), and implicitly restrict sums over subsystems \( i \) to elements of \( \omega \). Then using the definition of \( K(\omega'; t) \), we can expand

\[
    \dot{\gamma}(t) = \sum_{x,x'} K^x_{x'}(t) p_{x'}(t) \ln \left[ \frac{K^x_{x'}(t)}{K^x_{x'}(t) p_x(t)} \right] \quad \text{(D1)}
\]

\[
    = \sum_{x,x'} K^x_{x'}(\omega'; t) p_{x'}(t) \ln \left[ \frac{K^x_{x'}(t)}{K^x_{x'}(t) p_x(t)} \right] + \sum_{i \in \omega} \sum_{x,x'} K^x_{x'}(i; t) p_{x'}(t) \ln \left[ \frac{K^x_{x'}(t)}{K^x_{x'}(t) p_x(t)} \right]. \quad \text{(D2)}
\]
Since $\omega'$ is a unit, by equation (B3) we can rewrite the first sum on the rhs of equation (D2) as

$$\sum_{x \in \omega'} K_{x'}(\omega'; t)p_{x}(t) \ln \left[ \frac{K_{x'}(\omega'; t)}{K_{x'}(\omega'; t)p_{x}(t)} \right] - \sum_{x \neq x'} K_{x'}(\omega'; t)p_{x}(t) \ln \left[ \frac{p_{x}(t)}{p_{x'}(t)} \right]$$

$$= \langle \sigma_{\omega'}(t) \rangle + \frac{d^{\omega'/K}}{dt} S^{\omega'_{\omega'}}(t). \quad (D3)$$

Moreover, by equation (B2), even though $\omega \setminus \omega'$ need not be a unit, the second sum in equation (D2) can be rewritten as

$$\sum_{i \neq \omega'} \sum_{x \in \omega'/x} K_{x}(i; t)p_{x}(t) \ln \left[ \frac{K_{x}(i; t)}{K_{x}(i; t)p_{x}(t)} \right] = \sum_{x \neq x'} K_{x}(\omega \setminus \omega'; t)p_{x}(t) \ln \left[ \frac{K_{x}(\omega \setminus \omega'; t)}{K_{x}(\omega \setminus \omega'; t)p_{x}(t)} \right]. \quad (D4)$$

Combining completes the proof. In order to express that proof as in the main text, with the implicit $\omega$ once again made explicit, use the fact that windowing $K(\omega; t)$ to $\omega' \subset \omega$ is the same as windowing $K(t)$ to $\omega'$.

### Appendix E. EP bounds from the inclusion-exclusion principle

For all $n > 1$, write $\mathcal{N}^n$ for the multiset of all intersections of $n$ of the sets of subsystems $\omega_i$:

$$\mathcal{N}^2 = \{ \omega_i \cap \omega_j : 1 \leq i < j < |\mathcal{N}^1| \} \quad (E1)$$

$$\mathcal{N}^3 = \{ \omega_i \cap \omega_j \cap \omega_k : 1 \leq i < j < k < |\mathcal{N}^1| \} \quad (E2)$$

and so on, up to $\mathcal{N}^{|\mathcal{N}^1|}$. Any unit structure $\mathcal{N}$ specifies an associated set of sets,

$$\mathcal{N}^* := \mathcal{N}^1 \cup \mathcal{N}^2 \cup \ldots \cup \mathcal{N}^{|\mathcal{N}^1|}. \quad (E3)$$

Note that every element of $\mathcal{N}^*$ is itself a unit, since intersections of units are units.

Suppose we also have a real-valued vector indexed by the sets of $\mathcal{N}^*$, written as $f^\omega$. The associated inclusion–exclusion sum (or just ‘in–ex sum’) is defined as

$$\sum_{\omega \in \mathcal{N}^*} f^\omega := \sum_{j=1}^{n} f^{\omega_j} - \sum_{1 \leq j < r \leq n} f^{\omega_j \cap \omega_r} + \sum_{1 \leq j < k < n} f^{\omega_j \cap \omega_k \cap \omega_r} - \ldots \quad (E4)$$

(Note that the precise assignment of integer indices to the units in $\mathcal{N}^*$ is irrelevant.)

The time-$t$ in–ex information is defined in terms of this notation, as

$$I_{\mathcal{N}^*} := \sum_{\omega \in \mathcal{N}^*} S^\omega - S^V$$

$$= -S^V + \sum_{j=1}^{n} S^{\omega_j} - \sum_{1 \leq j < r \leq n} S^{\omega_j \cap \omega_r} + \ldots \quad (E5)$$

As an example, if $\mathcal{N}^*$ consists of two units, $\omega_1, \omega_2$, with no intersection, then the expected in-ex information at time $t$ is just the mutual information between those units at that time. More generally, if there an arbitrary number of units in $\mathcal{N}^*$ but none of them overlap, then the expected in–ex information is what is called the ‘multi-information’, or ‘total correlation’, among those units [31]. Mutual information is the special case where there are exactly two random variables [14, 31].

Given any function $f : \mathcal{N}^* \rightarrow \mathbb{R}$, the associated inclusion–exclusion sum (or just ‘in–ex sum’) is

$$\Sigma := \sum_{\omega \in \mathcal{N}^*} f(\omega) - \sum_{\omega \in \mathcal{N}^2} f(\omega) + \sum_{\omega \in \mathcal{N}^3} f(\omega) - \ldots \quad (E6)$$
In particular, given any distribution \( p_x \), there is an associated real-valued function mapping any \( \omega \in \mathcal{N}^* \) to the marginal entropy of (the subsystems in) \( \omega \). So using \( S^{\mathcal{N}^*} \) to indicate that function,

\[
\Sigma^S := \sum_{\omega \in \mathcal{N}^*} S^\omega - \sum_{\omega \in \mathcal{N}^0} S^\omega + \sum_{\omega \in \mathcal{N}^3} S^\omega - \ldots \tag{E7}
\]

where \( S^\omega \) is shorthand for \( S^{\mathcal{N}^*} \). I refer to \( \Sigma^S - S^{\mathcal{N}^3} \) as the in–ex information. As an example, if \( \mathcal{N}^i \) consists of two subsets, \( \omega_1, \omega_2 \), with no intersection, then the in–ex information is just the mutual information \( I(X_{\omega_1}; X_{\omega_2}) \). As another example, if \( \mathcal{N}^i \) consists of all singletons \( i \in \mathcal{N} \), then the in–ex information is the multi-information of the \( N \) separate random variables.

The global EP rate is the negative derivative of the in–ex information, plus the in–ex sum of local EP rates:

\[
\langle \dot{\sigma}(t) \rangle = \frac{dS^\mathcal{N}(t)}{dt} + \langle \dot{Q}^\mathcal{N}(t) \rangle \tag{E8}
\]

\[
= -\frac{d}{dt} S^\mathcal{N}(t) + \sum_{\omega \in \mathcal{N}^*} \langle \dot{\sigma}^\omega(t) \rangle. \tag{E9}
\]

**Proof.** To establish equation (E8), first plug in to the result in appendix B and use the normalization of the rate matrices to see that the EP rate of the full set of \( N \) coupled subsystems is

\[
\langle \dot{\sigma}(t) \rangle = \sum_{i \neq x} K_x^\prime(t) p_x(t) \ln \left[ \frac{K_x^\prime(t) p_x(t)}{K_x^\prime(t) p_x(t)} \right] \\
= \sum_{i \neq x} K_x^\prime(t; i) p_x(t) \ln \left[ \frac{K_x^\prime(t) p_x(t)}{K_x^\prime(t) p_x(t)} \right] \\
= \sum_{i \neq x} K_x^\prime(t; i) p_x(t) \ln \left[ \frac{K_x^\prime(t) p_x(t)}{K_x^\prime(t) p_x(t)} \right]. \tag{E10}
\]

Now introduce the shorthand

\[
G(A \subseteq \mathcal{N}) := \sum_{i \in A, x \neq x} K_x^\prime(i; t) p_x(t) \ln \left[ \frac{K_x^\prime(t) p_x(t)}{K_x^\prime(t) p_x(t)} \right]. \tag{E11}
\]

Note that \( \mathcal{N} \) itself is a unit; \( G \) is an additive function over subsets of \( \mathcal{N} \); and \( \langle \dot{\sigma}(t) \rangle = G(\mathcal{N}) \). Accordingly, we can apply the inclusion–exclusion principle to equation (E10) for the set of subsets \( \mathcal{N}(t) \) to get

\[
\langle \dot{\sigma}(t) \rangle = \sum_{\omega \in \mathcal{N}(t)} G(\omega) - \sum_{\omega \in \mathcal{N}^0(t)} G(\omega) + \sum_{\omega \in \mathcal{N}^3(t)} G(\omega) - \ldots
\]

\[
= \sum_{\omega \in \mathcal{N}(t)} \sum_{i \neq x} K_x^\prime(i; t) p_x(t) \ln \left[ \frac{K_x^\prime(t) p_x(t)}{K_x^\prime(t) p_x(t)} \right]
\]

\[
- \sum_{\omega \in \mathcal{N}^0(t)} \sum_{i \neq x} K_x^\prime(i; t) p_x(t) \ln \left[ \frac{K_x^\prime(t) p_x(t)}{K_x^\prime(t) p_x(t)} \right]
\]

\[
+ \sum_{\omega \in \mathcal{N}^3(t)} \sum_{i \neq x} K_x^\prime(i; t) p_x(t) \ln \left[ \frac{K_x^\prime(t) p_x(t)}{K_x^\prime(t) p_x(t)} \right] - \ldots
\]

\[
= \sum_{\omega \in \mathcal{N}(t)} \sum_{i \neq x} K_x^\prime(\omega; t) p_x(t) \ln \left[ \frac{K_x^\prime(t) p_x(t)}{K_x^\prime(t) p_x(t)} \right]
\]

\[
- \sum_{\omega \in \mathcal{N}^0(t)} \sum_{i \neq x} K_x^\prime(\omega; t) p_x(t) \ln \left[ \frac{K_x^\prime(t) p_x(t)}{K_x^\prime(t) p_x(t)} \right] + \ldots \tag{E12}
\]

Now use equation (B3) in lemma 1 to rewrite equation (E12) as
\[
\langle \dot{\sigma}(t) \rangle = \sum_{\omega \in \mathcal{N}^s(t) \times \mathcal{K}_{\omega}^{\mathcal{N}}(t)} K_{\omega}^{\mathcal{N}}(\omega; t) \delta_{x_{\omega}, x_{\omega}'}(t) \ln \left[ \frac{K_{\omega}^{\mathcal{N}}(\omega; t)}{K_{x_{\omega}'}^{\mathcal{N}}(\omega; t)p_{x_{\omega}'}(t)} \right] \\
- \sum_{\omega \in \mathcal{N}^s(t) \times \mathcal{K}_{\omega}^{\mathcal{N}}(t)} K_{\omega}^{\mathcal{N}}(\omega; t) \delta_{x_{\omega}, x_{\omega}'}(t) \ln \left[ \frac{K_{\omega}^{\mathcal{N}}(\omega; t)}{K_{x_{\omega}'}^{\mathcal{N}}(\omega; t)p_{x_{\omega}'}(t)} \right] + \ldots \\
= \sum_{\omega \in \mathcal{N}^s(t) \times \mathcal{K}_{\omega}^{\mathcal{N}}(t)} K_{\omega}^{\mathcal{N}}(\omega; t) \ln \left[ \frac{K_{\omega}^{\mathcal{N}}(\omega; t)}{K_{\omega}^{\mathcal{N}}(\omega; t)p_{x_{\omega}}(t)} \right] - \sum_{\omega \in \mathcal{N}^s(t) \times \mathcal{K}_{\omega}^{\mathcal{N}}(t)} K_{\omega}^{\mathcal{N}}(\omega; t) \ln \left[ \frac{K_{\omega}^{\mathcal{N}}(\omega; t)}{K_{x_{\omega}'}^{\mathcal{N}}(\omega; t)p_{x_{\omega}'}(t)} \right] + \ldots \tag{E13}
\]

Next, use the same kind of reasoning that resulted in equation (E13) to show that the sum

\[
\sum_{\omega \in \mathcal{N}^s(t) \times \mathcal{K}_{\omega}^{\mathcal{N}}(t)} K_{\omega}^{\mathcal{N}}(\omega; t) \ln \left[ \frac{K_{\omega}^{\mathcal{N}}(\omega; t)}{K_{x_{\omega}'}^{\mathcal{N}}(\omega; t)p_{x_{\omega}'}(t)} \right] - \sum_{\omega \in \mathcal{N}^s(t) \times \mathcal{K}_{\omega}^{\mathcal{N}}(t)} K_{\omega}^{\mathcal{N}}(\omega; t) \ln \left[ \frac{K_{\omega}^{\mathcal{N}}(\omega; t)}{K_{x_{\omega}'}^{\mathcal{N}}(\omega; t)p_{x_{\omega}'}(t)} \right] + \ldots \tag{E14}
\]

can be written as \(\sum_{\omega \in \mathcal{N}^s(t) \times \mathcal{K}_{\omega}^{\mathcal{N}}(t)} K_{\omega}^{\mathcal{N}}(t)p_{x_{\omega}}(t)\). I can use this to rewrite equation (E13) as

\[
\langle \dot{\sigma}(t) \rangle = -\frac{d}{dt} T_{\mathcal{N}^s}(t) + \sum_{\omega \in \mathcal{N}^s(t) \times \mathcal{K}_{\omega}^{\mathcal{N}}(t)} K_{\omega}^{\mathcal{N}}(\omega; t) \ln \left[ \frac{K_{\omega}^{\mathcal{N}}(\omega; t)}{K_{x_{\omega}'}^{\mathcal{N}}(\omega; t)p_{x_{\omega}'}(t)} \right] \\
- \sum_{\omega \in \mathcal{N}^s(t) \times \mathcal{K}_{\omega}^{\mathcal{N}}(t)} K_{\omega}^{\mathcal{N}}(\omega; t) \ln \left[ \frac{K_{\omega}^{\mathcal{N}}(\omega; t)}{K_{x_{\omega}'}^{\mathcal{N}}(\omega; t)p_{x_{\omega}'}(t)} \right] + \ldots \tag{E15}
\]

This establishes the claim.

If we use equation (10) to expand each local EP term in equation (E9) and then compare to equation (E8), we see that the global expected EF rate equals the \(\text{in–ex}\) sum of the local expected EF rates:

\[
\langle \dot{Q}_{\mathcal{N}}(t) \rangle = \sum_{\omega \in \mathcal{N}^s} \langle \dot{Q}_{\omega}(t) \rangle. \tag{E16}
\]

Note as well that we can apply equation (E9) to itself, by using it to expand any of the local EP terms \(\sigma_{\omega}(t)\) that occur in the \(\text{in–ex}\) sum \(\sum_{\omega \in \mathcal{N}^s} \langle \dot{Q}_{\omega}(t) \rangle\) on its own rhs.

**Example 1.** A particularly simple application of equation (E9) is to extend the results in [38] on ‘subsystem Landauer loss’. Suppose that all units that are intersections of two or more other units don’t ever change their state. That means that the unit EPs of such intersection units all equal 0, and that their drops in entropy during the process all equal 0. Accordingly, equation (E9) gives

\[
\langle \sigma \rangle = \sum_{A} (\sigma_{\omega}) + \Delta S(p) - \sum_{\omega} \Delta S(p_{\omega}). \tag{E17}
\]

So by non-negativity of expected unit EPs:

\[
\langle \sigma \rangle \geq \Delta S(p) - \sum_{\omega} \Delta S(p_{\omega}). \tag{E18}
\]

The analysis in [38] on ‘subsystem Landauer loss’ derives these results for the special case where there are in fact no intersection units; in that special case, the unit structure is just a partition of \(\mathcal{N}\), so the rhs of equation (E18) is just the change in multi-information (sometimes called ‘total correlation’ [30, 31, 35]) of the system during the process.
Example 2. Equation (E9) can be particularly useful when combined with the fact that for any two units \( \omega, \omega' \subset \omega, (\dot{\sigma}^\omega) \leq (\dot{\sigma}^{\omega'}) \) (see equation (15)). To illustrate this, return to the scenario of figure 1. There are three units in \( \mathcal{N}^1 \) (namely, \{1, 2, 3\}, \{3\}, \{3, 4\}), three in \( \mathcal{N}^2 \) (namely, three copies of \{3\}), and one in \( \mathcal{N}^3 \) (namely, \{3\}). Therefore,

\[
\langle \dot{\sigma}(t) \rangle = \frac{dS^{1, 2, 3}(t)}{dt} - \frac{dS^{1, 2}(t)}{dt} - \frac{dS^{3}(t)}{dt} + \frac{dS^{4}(t)}{dt} + \langle \dot{\sigma}^{1, 2, 3} \rangle + \langle \dot{\sigma}^{3, 4} \rangle - \langle \dot{\sigma}^3 \rangle \tag{E19}
\]

Note that the derivatives in equation (E20) are conventional, non-windowed derivatives; none of the terms in equation (E20) involve counterfactual rate matrices.

All of the local EPs in equation (E19) can equal 0, in a quasi-static process. In addition, \( \langle \dot{\sigma}(t) \rangle \geq 0 \).

Therefore \( dS^{1, 2, 3}(t)/dt - dS^{3}(t)/dt \geq 0 \), i.e.,

\[
\langle \dot{\sigma}(t) \rangle \geq \frac{dS^{1, 2, 3}(t)}{dt} - \frac{dS^{3}(t)}{dt} \geq 0. \tag{E21}
\]

As a final comment, it is worth noting that in contrast to multi-information, in some situations the in–ex information can be negative. (In this it is just like some other extensions of mutual information to more than two variables [14, 31].) As an example, suppose \( N = 6 \), and label the subsystems as \( \mathcal{N} = \{12, 13, 14, 23, 24, 34\} \). Then take \( \mathcal{N}^1 \) to have four elements, \{12, 13, 14\}, \{23, 24, 12\}, \{34, 13, 23\} and \{34, 24, 14\}. (So the first element consists of all subsystems whose label involves a 1, the second consists of all subsystems whose label involves a 2, etc.). Also suppose that with probability 1, the state of every subsystem is the same. Then if the probability distribution of that identical state is \( p \), the in-ex information is \( -S(p) + 4S(p) = 6S(p) = -3S(p) \leq 0 \).

ORCID iDs

David H Wolpert © https://orcid.org/0000-0003-3105-2869

References

[1] Barato A C, Hartich D and Seifert U 2014 Efficiency of cellular information processing New J. Phys. 16 103024
[2] Bisker G, Polettini M, Gingrich T R and Horowitz J M 2017 Hierarchical bounds on entropy production inferred from partial information J. Stat. Mech. 2017 093210
[3] Bo S, Giudice M D and Celani A 2015 Thermodynamic limits to information harvesting by sensory systems J. Stat. Mech. 2015 P01014
[4] Boyd A B, Mandal D and Crutchfield J P 2018 Thermodynamics of modularity: structural costs beyond the Landauer bound Phys. Rev. X 8 031036
[5] Brittain R A, Jones N S and Ouldridge T E 2017 What we learn from the learning rate J. Stat. Mech. 2017 063502
[6] Cover T M and Thomas J A 2012 Elements of Information Theory (New York: Wiley)
[7] Esposito M and van den Broeck C 2010 Three faces of the second law. I. Master equation formulation Phys. Rev. E 82 011143
[8] Hartich D, Barato A C and Seifert U 2014 Stochastic thermodynamics of bipartite systems: transfer entropy inequalities and a Maxwell’s demon interpretation J. Stat. Mech. 2014 P02016
[9] Hartich D, Barato A C and Seifert U 2016 Sensory capacity: an information theoretical measure of the performance of a sensor Phys. Rev. E 93 022116
[10] Horowitz J M 2015 Multiplicative information flow for multiple Maxwell demons J. Stat. Mech. 2015 P03006
[11] Horowitz J M and Esposito M 2014 Thermodynamics with continuous information flow Phys. Rev. X 4 031015
[12] Horowitz J M and Parrondo J M R 2011 Designing optimal discrete-feedback thermodynamic engines New J. Phys. 13 123019
[13] Ito S and Sagawa T 2013 Information thermodynamics on causal networks Phys. Rev. Lett. 111 180603
[14] McGill W 1954 Multivariate information transmission Trans. IEEE Prof. Group Inf. Theory 4 93–111
[15] To see this, note that if \( x'_i = x_i \), then the only way for \( S^{i} (\| \sigma \| p) \) to be nonzero is if \( x'_{i'} = x_{i'} \) and \( i' = i \). If instead \( x'_i \neq x_i \) can differ from \( i \). However, if \( i' \neq i \) then the sum over \( x_{i'} \) in equation (4) runs over all values of \( x_i \). By normalization of the rate matrix \( K^{i}_{i'}(\| \sigma \| p) \) that sum must equal zero
[16] It is possible to choose that subset of state transitions so that \( \sigma \Omega \) concerns all transitions in which one particular subsystem changes state while all others do not. In this case, like the quantity \( S^{i}_{\Omega} \) considered in [11], \( \sigma \Omega \) is a single term in the decomposition of the EP rate generated by a single unit
[17] Not only can this variant of entropy production be negative; since it differs from the standard stochastic thermodynamics definition of entropy production, e.g., as in [32], in general this variant does not obey the usual fluctuation theorems, thermodynamic uncertainty relations, speed-limit theorems, etc. that all hold for the standard definition. (Indeed, one of the main results of [23] was precisely how to modify the standard fluctuation theorem for this modified definition of the standard entropy production.) As discussed below in section 4, all of those theorems are obeyed by \( (\dot{\sigma}^\omega(t)) \). See appendix C in [39] for discussion of the relation between standard entropy production and the variant considered in the thermodynamics of information literature
[18] In [10, 11], the A-derivative of the mutual information between A and N\( \setminus A \), \( d\mathcal{I}(X_{\omega}X_{N\setminus A})/dt \), is interpreted as the ‘information flow’ from \( N\setminus A \) to A. (See equation (12) in [10].) However, in the scenarios considered in this paper A will always be a unit. Therefore none of the subsystems in A will evolve in a way directly dependent on the state of any subsystem in \( N\setminus A \) (nor vice versa). So the fact that \( d\mathcal{I}(X_{\omega}X_{N\setminus A})/dt < 0 \) will not indicate that information in A concerning \( N\setminus A \) ‘flows’ between A and \( N\setminus A \) in any sense. In the current context, it would be more accurate to refer to \( -\frac{dS(X_{\omega})}{dt} \) as the ‘forgetting rate’ of A concerning \( N\setminus A \), than as ‘information flow’.

[20] Recall from the discussion above of the scenario considered in [1] that while the external subsystem \( \omega \) is its own unit, the internal subsystem is not. This means that in general, if the full system is not in a stationary state, then the learning rate of the internal subsystem about the external subsystem (as defined in [1]) cannot be expressed as \( -\frac{dS(X_{\omega})}{dt} \) with \( \omega \) being a unit.