

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/356195059>

# Thermodynamics of Encoding and Encoders

Preprint · November 2021

DOI: 10.13140/RG.2.2.30387.35368

CITATIONS

0

READS

48

2 authors:



Yang Tian

Tsinghua University

6 PUBLICATIONS 17 CITATIONS

SEE PROFILE



Pei Sun

Tsinghua University

68 PUBLICATIONS 805 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Simple Action Alters Attention Towards Visual Features [View project](#)



Theoretical neuroscience and statistical physics project [View project](#)

---

# THERMODYNAMICS OF ENCODING AND ENCODERS

---

A PREPRINT

**Yang Tian**

Department of Psychology & Tsinghua Laboratory of Brain and Intelligence  
Tsinghua University  
Beijing, 100084, China.  
tiany20@mails.tsinghua.edu.cn

**Pei Sun**

Department of Psychology & Tsinghua Laboratory of Brain and Intelligence  
Tsinghua University  
Beijing, 100084, China.  
peisun@tsinghua.edu.cn

November 15, 2021

## ABSTRACT

Non-isolated systems have diverse coupling relations with the external environment. These relations generate complex thermodynamics and information transmission between the system and its environment. The framework depicted in the current research attempts to glance at the critical role of the internal orders inside the non-isolated system in shaping the information thermodynamics coupling. We characterize the coupling as a generalized encoding process, where the system acts as an information thermodynamics encoder to encode the external information based on thermodynamics. We formalize the encoding process in the context of the nonequilibrium second law of thermodynamics, revealing an intrinsic difference in information thermodynamics characteristics between information thermodynamics encoders with and without internal correlations. During the information encoding process of an external source  $Y$ , specific sub-systems in an encoder  $X$  with internal correlations can exceed the information thermodynamics bound on  $(X, Y)$  and encode more information than system  $X$  works as a whole. We computationally verify this theoretical finding in an Ising model with a random external field and a neural data set of the human brain during visual perception and recognition. Our analysis demonstrates that the stronger internal correlation inside these systems implies a higher possibility for specific sub-systems to encode more information than the global one. These findings may suggest a new perspective in studying information thermodynamics in diverse physical and biological systems.

Perhaps the connection between thermodynamics and information is one of the most intriguing relations discovered in physics. It is impressive to see the ubiquitous and critical roles of information thermodynamics links in diverse physical, chemical, and biological systems. Nowadays, emerging interests have arisen in studying information thermodynamics in complex systems (e.g., the human brain), whose complexity originates from heterogeneous elements and intricate internal correlations. An open challenge relates to the elusive roles of internal orders inside a complex system in influencing information thermodynamics. This research attempts to lay a cornerstone of this rising direction. We present a theory of the thermodynamics of encoding, where a non-isolated system  $X$  acts as an information thermodynamics encoder to encode the information of an external source  $Y$  through thermodynamics coupling. This theory reveals an intrinsic difference between the non-isolated system with intra-system coupling and the non-isolated system of independent elements in information thermodynamics characteristics. It demonstrates that a stronger internal correlation inside system  $X$  implies a higher possibility for specific sub-systems of  $X$  to encode more information of  $Y$  than  $X$  itself. We

**further verify the ubiquity of these findings in representative non-isolated systems in physics and biology, such as the Ising model with a random external field and the human brain, an ultra-complex system of neurons, during perception. These validated theoretical results may deepen our understanding of physical and biological information processing systems and inspire advanced computing architecture, storage infrastructure, and artificial neural network designs.**

## 1 Introduction

As you are reading this sentence, synergistic neural dynamics emerges in your brain (a non-isolated system of neurons) to encode the text information [19, 23], changing the thermodynamic state of your brain continuously [21]. This is an instance of the coupling between a non-isolated system and its external environment. Similar instances is common in classic [85, 55, 97], quantum [97, 15, 12], biological [52, 73, 39], ecological [50, 94, 75], and social [75, 26, 82] systems. The pursuit to understand this coupling plays a pivotal role in statistical physics. Distinguished from the macroscopic systems studied by classic thermodynamics [14], the microscopic and mesoscopic systems in stochastic thermodynamics [86, 88] usually feature non-negligible system-environment coupling [49]. Over decades, the study of the thermodynamic attributes of these systems has become a rapidly emerging direction [47, 32, 27, 71, 87, 70, 96].

Apart from the above progress, a frequently neglected perspective is that the coupling can also be a process for the system to encode the external environment information. At first glance, this information-theoretical aspect seems to focus on quantifying the information processed by a population of elements about external sources (e.g., with multiple mutual information [64], total correlation [101], or connected information [83]), irrelevant to the thermodynamics analysis. However, as suggested by Szilard’s [95] and Landauer’s [57, 56] classical works, information is physical. Historically, the information-thermodynamics connection is first discovered in the paradox of the Maxwell demon [59]. Recently, more fundamental connections have been identified in reset [59, 77, 76], measurement [42, 77, 33, 80, 36, 28], and feedback [42, 80, 77, 2] processes and have been verified experimentally [7, 8, 51, 99]. The idea to consider thermodynamics from the information aspect (or vice versa) has been proven effective in identifying potential links between physics and information quantities [53, 78, 90, 45, 44, 69, 36, 28]. Meanwhile, these intriguing connections have inspired widespread explorations on the possibility for the information-carrying capacity of a memory device to function as a thermodynamic fuel [11, 10, 91, 5, 41, 100, 65], leading to the remarkable development of designing nanoscale autonomous machines that are fueled by information (referred to as information engines) [24, 63, 62, 60, 17]. These recent advances in physics have further stimulated growing attention from other fields such as biology [16, 21, 4, 48, 6, 81] and computer science [43].

Till now, it is unclear how internal orders inside a non-isolated system affect information thermodynamics. The necessity to study this problem emerges especially for the complex systems (e.g., the brain), whose complexity originates from heterogeneous elements and intricate internal correlations. Our primary objective is to seek a systematic framework that formalizes the information thermodynamics coupling between a non-isolated system  $X$  and an external source  $Y$  and helps understand the critical roles of the existing order inside the system  $X$  on this coupling relation. Another objective is to explore potential diversities of information thermodynamics characteristics determined by intra-system factors and verify these diversities in concrete non-isolated systems.

The rest of our paper is organized as follows. Sec. 2 depicts the theoretical framework of our research. We suggest an original perspective to characterize the non-isolated system  $X$  as an information thermodynamics encoder, which encodes the information of a coupled external source  $Y$  through thermodynamics. After contextualizing this perspective with existing theories, we formalize the definition of encoding in the context of the nonequilibrium second law of thermodynamics. Based on this formalization, Sec. 3 goes deep into the information thermodynamics of different non-isolated systems. We unveil how the nature of order inside the non-isolated system determines the information thermodynamics characteristics. An intrinsic difference in information thermodynamics is revealed between the encoders with and without intra-system coupling—unlike those with independent elements, an encoder  $X$  with intra-system coupling allows the encoded information of  $Y$  in its sub-systems to exceed the information thermodynamics bound on the joint system  $(X, Y)$ . In other words, every time a particular amount of irreversible work is extracted or dispensed from the joint system  $(X, Y)$ , specific sub-systems of  $X$  may be able to encode more information of  $Y$  than system  $X$  itself. This possibility originates from the internal correlation inside encoder  $X$ . In Sec. 4, we have computationally verified our theory in an Ising model with random external field and a real data set of the human brain during the perception process.

## 2 Information thermodynamics encoder and encoding

### 2.1 Definition of information thermodynamics encoder

Let us consider a non-isolated system  $X = (\{X_i\}, \mathcal{R})$ . Here we represent it as a graph. Set  $\{X_i\}$  includes  $n$  elements (vertices), where each  $X_i$  has  $m$  states  $\Omega(X_i) = \{X_i^j\}$  (here  $\Omega(\cdot)$  defines the sample space). Set  $\mathcal{R} \subseteq \{X_i\} \times \{X_i\}$  describes the intra-system coupling relations (edges). Note that the coupling is a kind of equivalence relation, making every connected component of the graph be a clique (see Fig. 1). To function as an information processor, system  $X$  must have multiple distinguishable states for information storage [69]. Therefore, we always consider the cases with  $\sum_i |\Omega(X_i)| > n$  to guarantee  $|\Omega(X)| > 1$ .

We refer to  $X$  as an information thermodynamics encoder when it couples with an external source  $Y = \{Y_i\}$  (see Fig. 1 for an instance). Following the idea in [33, 28], the joint system  $(X, Y)$  is in contact with a heat bath HB and the total system  $[(X, Y), \text{HB}]$  is assumed as isolated.

On the one hand, the coupling between  $X$  and  $Y$  allows the transformation of heat, energy or substances. On the other hand, this coupling ensures the existence of  $\mathcal{P}(X|Y)$  (here  $\mathcal{P}(\cdot)$  denotes the probability), establishing an information channel between  $X$  and  $Y$ . Such a channel supports the information-theoretical analysis of the coupling relation in the context of that  $X$  encodes the information of  $Y$ . Taken together, the dual attributes of this coupling allow us implement a unified analysis of information and thermodynamics.

### 2.2 The second law of thermodynamics

To embed our analysis with solid physics backgrounds, we begin with a discussion on the second law of thermodynamics.

Let us consider a general case where  $X$  will change if it interacts with  $Y$  during an interval  $[0, t]$ . The joint system  $(X, Y)$  varies due to this interaction and the heat bath HB accounts for supplying necessary heat. In [80, 79], the time evolution of  $X$  under the influence of  $Y$  has been previously used for deriving the nonequilibrium second law of thermodynamics from the information perspective.

Here we offer a more fundamental derivation from the thermodynamics perspective. System  $(X, Y)$  varies during  $[0, t]$ , accompanied by a nonequilibrium entropy change  $\Delta S$  and a nonequilibrium free energy change  $\Delta \mathcal{F}$ . By decomposing  $\Delta S$  as  $\Delta S = \Delta_e S + \Delta_i S$ , one can distinguish between the reversible contribution  $\Delta_e S = Q/T$  due to the heat flow (heat  $Q$  comes from the heat bath HB and  $T$  is the temperature) and the irreversible contribution  $\Delta_i S \geq 0$ . As demonstrated in [28], an equivalent formulation of the nonequilibrium second law can be obtained by  $T\Delta_i S = \mathcal{W} - \Delta \mathcal{F} \geq 0$ , where  $\mathcal{W}$  is the work performed on system  $(X, Y)$ .

Based on the knowledge of free energy [25, 72], the actual free energy  $\mathcal{F}_{(\tau)}$  at moment  $\tau \in [0, t]$  is associated with the equilibrium free energy  $\hat{\mathcal{F}}_{(\tau)}$  as well as the relative entropy between the actual state  $\mathcal{P}(X_{(\tau)}, Y_{(\tau)})$  and the

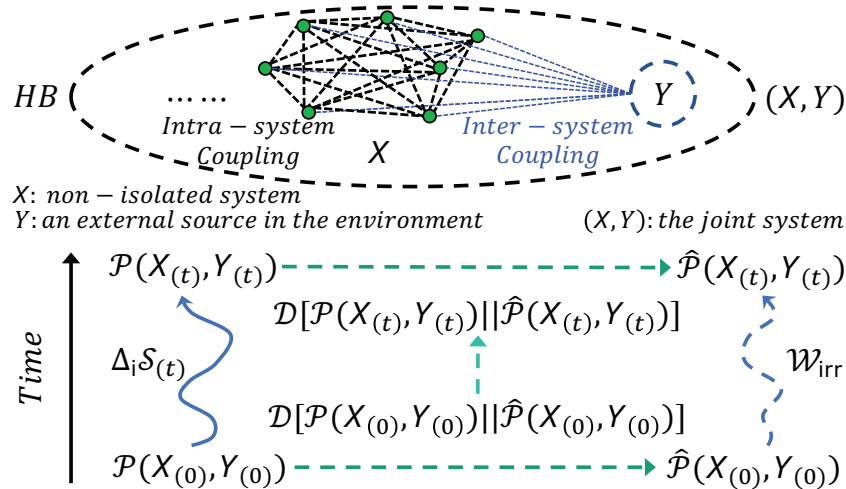


Figure 1: A schematic diagram of the coupled system  $(X, Y)$  and the illustration of equality (1).

equilibrium state  $\hat{\mathcal{P}}(X_{(\tau)}, Y_{(\tau)})$  (here notion  $X_{(\tau)}$  represents the state of  $X$  at moment  $\tau$ ). More specifically, one can know  $\mathcal{F}_{(\tau)} = \hat{\mathcal{F}}_{(\tau)} + T\mathcal{D}[\mathcal{P}(X_{(\tau)}, Y_{(\tau)}) \parallel \hat{\mathcal{P}}(X_{(\tau)}, Y_{(\tau)})]$ , where  $\mathcal{D}[\cdot]$  denotes the relative entropy [22].

Therefore, the nonequilibrium second law during  $[0, t]$  can be specified as  $T\Delta_i\mathcal{S}_{(t)} = \mathcal{W} - (\mathcal{F}_{(t)} - \mathcal{F}_{(0)}) = \mathcal{W} - (\hat{\mathcal{F}}_{(t)} - \hat{\mathcal{F}}_{(0)}) - T\mathcal{D}[\mathcal{P}(X_{(t)}, Y_{(t)}) \parallel \hat{\mathcal{P}}(X_{(t)}, Y_{(t)})] + T\mathcal{D}[\mathcal{P}(X_{(0)}, Y_{(0)}) \parallel \hat{\mathcal{P}}(X_{(0)}, Y_{(0)})]$ , where the term  $\mathcal{W} - (\hat{\mathcal{F}}_{(t)} - \hat{\mathcal{F}}_{(0)}) = \mathcal{W} - \Delta\hat{\mathcal{F}} = \mathcal{W}_{\text{irr}}$  is referred to as the irreversible work [28]. Taking these derivations together, the nonequilibrium second law can be reformulated as following [28] (please see Fig. 1 for an illustration)

$$\frac{\mathcal{W}_{\text{irr}}}{T} = \mathcal{D}[\mathcal{P}(X_{(t)}, Y_{(t)}) \parallel \hat{\mathcal{P}}(X_{(t)}, Y_{(t)})] - \mathcal{D}[\mathcal{P}(X_{(0)}, Y_{(0)}) \parallel \hat{\mathcal{P}}(X_{(0)}, Y_{(0)})] + \Delta_i\mathcal{S}_{(t)}. \quad (1)$$

Here the term  $\mathcal{D}[\mathcal{P}(X_{(t)}, Y_{(t)}) \parallel \hat{\mathcal{P}}(X_{(t)}, Y_{(t)})] - \mathcal{D}[\mathcal{P}(X_{(0)}, Y_{(0)}) \parallel \hat{\mathcal{P}}(X_{(0)}, Y_{(0)})]$  can be treated as information gain if it is positive, otherwise it is information reduction [28]. Accordingly, the irreversible work  $\mathcal{W}_{\text{irr}}$  in (1) can be either negative when the information reduction is greater than the entropy production, or non-negative in other cases [28].

### 2.3 Thermodynamics and information

Given the second law of thermodynamics in (1), one might be curious about its connection to the above mentioned information-theoretical perspective. Let us consider the information amount of  $Y$  can be encoded in  $X$  [22]. This aspect is studied in the term of measurement in statistical physics [69, 42, 77, 33, 80, 36, 28] and substantial progress has been made in its application in neuroscience [21, 23, 102].

To organize our derivation, we divide the timeline into 3 sections, respectively corresponding to the period before, during, and after the measurement process.

- Before the measurement (before moment 0), there is no restriction on  $X$  and  $Y$ . They can be either independent or coupled. Specially, we refer to the case where  $X$  and  $Y$  are originally independent until the measurement begins as the **UCI Condition** (Uncoupled initialization condition), namely that  $\hat{\mathcal{P}}(X_{(0)}, Y_{(0)}) = \hat{\mathcal{P}}(X_{(0)}) \hat{\mathcal{P}}(Y_{(0)})$  in (1). Moreover, for the special case where the joint system  $(X, Y)$  is at equilibrium when the measurement begins, we refer to it as the **EI Condition** (Equilibrium initialization condition), namely that  $\mathcal{P}(X_{(0)}, Y_{(0)}) = \hat{\mathcal{P}}(X_{(0)}, Y_{(0)})$ ;
- During the measurement (during  $[0, t]$ ), it is necessary for  $X$  and  $Y$  to be coupled with each other. Specially, for the case where the total system  $[(X_{(\tau)}, Y_{(\tau)}), HB_{(\tau)}]$  (here  $\tau \in [0, t]$ ) always features a equilibrium state where system  $(X, Y)$  and  $HB$  are independent and  $HB$  is at equilibrium (there is no restriction on the state of system  $(X, Y)$ ), we refer to it as the **PE Condition** (Partial equilibrium condition), namely that  $\hat{\mathcal{P}}[(X_{(\tau)}, Y_{(\tau)}), HB_{(\tau)}] = \mathcal{P}(X_{(\tau)}, Y_{(\tau)}) \hat{\mathcal{P}}(HB_{(\tau)})$  holds for any  $\tau \in [0, t]$ ;

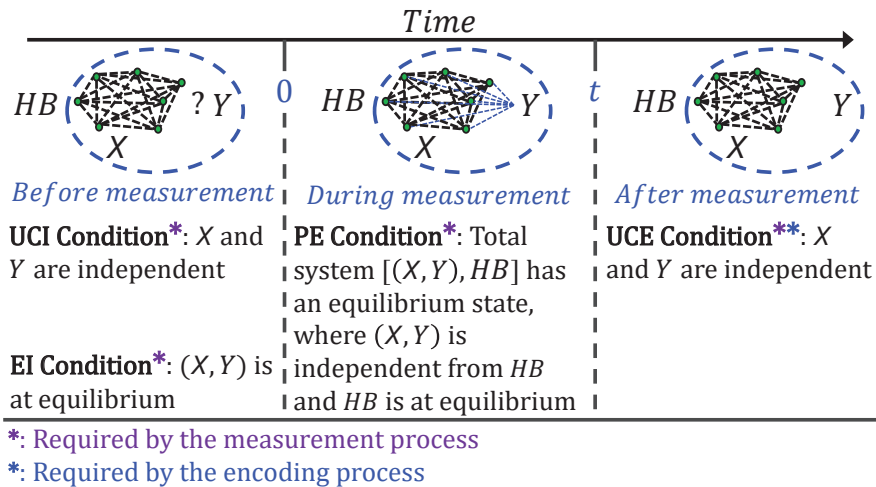


Figure 2: The difference between measurement and encoding across time sections.

- After the measurement (after moment  $t$ ), there is no restriction on  $X$  and  $Y$  as well. For the case where the coupling between  $X$  and  $Y$  is turned off after the measurement ends, we refer to it as the **UCE Condition** (Uncoupled ending condition). Under this condition, we have  $\hat{\mathcal{P}}(X_{(t)}, Y_{(t)}) = \hat{\mathcal{P}}(X_{(t)}) \hat{\mathcal{P}}(Y_{(t)})$  in (1).

As suggested by [28, 27, 46], the **PE Condition** is necessary for defining the entropy production in (1)

$$\forall \tau, \Delta_i \mathcal{S}_{(\tau)} = \mathcal{D}\{\mathcal{P}(X_{(\tau)}, Y_{(\tau)}) \parallel \text{HB}_{(\tau)} \parallel \mathcal{P}(X_{(\tau)}, Y_{(\tau)}) \hat{\mathcal{P}}(\text{HB}_{(\tau)})\}, \quad (2)$$

where the entropy production is defined as the irreversible contribution to the entropy change of system  $(X, Y)$  when it is coupled with HB. For the **PE Condition** and (2), one can see their detailed derivation in [27] and a summary in [28]. Moreover, an equivalent vision of the **PE Condition** can be seen in [33].

When the **UCE Condition** is satisfied in (1), we can know that  $\mathcal{D}[\mathcal{P}(X_{(t)}, Y_{(t)}) \parallel \hat{\mathcal{P}}(X_{(t)}, Y_{(t)})] = \mathcal{D}[\mathcal{P}(X_{(t)}, Y_{(t)}) \parallel \hat{\mathcal{P}}(X_{(t)}) \hat{\mathcal{P}}(Y_{(t)})]$ . In previous studies, the term  $\mathcal{D}[\mathcal{P}(X_{(t)}, Y_{(t)}) \parallel \hat{\mathcal{P}}(X_{(t)}) \hat{\mathcal{P}}(Y_{(t)})]$  is treated as the mutual information  $\mathcal{I}(X_{(t)}, Y_{(t)})$  between systems  $X$  and  $Y$  (e.g., see [28]). Considering the non-negative entropy production  $\Delta_i \mathcal{S}_{(t)} \geq 0$ , one can derive the following inequality

$$\frac{\mathcal{W}_{\text{irr}}}{T} + \mathcal{D}[\mathcal{P}(X_{(0)}, Y_{(0)}) \parallel \hat{\mathcal{P}}(X_{(0)}, Y_{(0)})] \geq \mathcal{I}(X_{(t)}; Y_{(t)}). \quad (3)$$

In the meanwhile, one can obtain a special case

$$\frac{\mathcal{W}_{\text{irr}}}{T} \geq \mathcal{I}(X_{(t)}, Y_{(t)}) \quad (4)$$

if the **EI Condition** is satisfied as well. (4) is the usual form of the thermodynamics cost of the measurement process [33, 36, 69]. Although the **UCI Condition** is not necessarily met in the above derivation, the energy definition of  $(X, Y)$  can be calculated in a simple form under this condition [28]. In [33], this condition is set for the measurement process as well. In a summary, the definition of measurement essentially requires the above four conditions to be met (see Fig. 2).

## 2.4 Thermodynamics of encoding

Given the summary above, the connection between the second law in (1) and the information has been revealed from the aspect of measurement. Then, we show how to generalize the concept of measurement to encoding.

For encoding, we only require the **UCE Condition** to be satisfied. Therefore, the measurement process studied by previous researches [69, 42, 77, 33, 80, 36, 28] is a particular case of encoding. In Fig. 2, we summarize the difference between the measurement process and encoding.

To formalize the thermodynamics of encoding, we need to note that  $\mathcal{D}[\mathcal{P}(X_{(\tau)}, Y_{(\tau)}) \parallel \hat{\mathcal{P}}(X_{(\tau)}) \hat{\mathcal{P}}(Y_{(\tau)})]$  (here  $\tau \in [0, t]$ ) is relevant with rather than equivalent to the mutual information  $\mathcal{I}(X_{(\tau)}, Y_{(\tau)})$  (please see Fig. 3). Although this nominal equivalence has been used to derive (3-4) in previous studies [28, 33], one can still verify that

$$\begin{aligned} \mathcal{D}[\mathcal{P}(X_{(\tau)}, Y_{(\tau)}) \parallel \hat{\mathcal{P}}(X_{(\tau)}) \hat{\mathcal{P}}(Y_{(\tau)})] &= \mathcal{I}(X_{(\tau)}, Y_{(\tau)}) + \\ &\sum_{\Omega(X) \times \Omega(Y)} \mathcal{P}(X_{(\tau)}, Y_{(\tau)}) \log \left( \frac{\mathcal{P}(X_{(\tau)}) \mathcal{P}(Y_{(\tau)})}{\hat{\mathcal{P}}(X_{(\tau)}) \hat{\mathcal{P}}(Y_{(\tau)})} \right), \end{aligned} \quad (5)$$

where  $\mathcal{D}[\mathcal{P}(X_{(\tau)}, Y_{(\tau)}) \parallel \hat{\mathcal{P}}(X_{(\tau)}) \hat{\mathcal{P}}(Y_{(\tau)})]$  consists of not only the mutual information  $\mathcal{I}(X_{(\tau)}, Y_{(\tau)})$  but also a compound term. One can find the detailed derivation of (5) in (23-25) in appendix A. Luckily, the non-equivalence suggested in (5) does not contradict with (3) or (4), since

$$\sum_{\Omega(X) \times \Omega(Y)} \mathcal{P}(X_{(\tau)}, Y_{(\tau)}) \log \left( \frac{\mathcal{P}(X_{(\tau)}) \mathcal{P}(Y_{(\tau)})}{\hat{\mathcal{P}}(X_{(\tau)}) \hat{\mathcal{P}}(Y_{(\tau)})} \right) \geq 0 \quad (6)$$

can be proven (see (25-37) in appendix A).

Given (6), we can formalize the thermodynamics of encoding in the form of (3). Under the **UCE Condition**, the irreversible work  $\mathcal{W}_{\text{irr}}$  that can be extracted or dispensed from system  $(X, Y)$  satisfies

$$\begin{aligned} &\frac{\mathcal{W}_{\text{irr}}}{T} + \mathcal{D}[\mathcal{P}(X_{(0)}, Y_{(0)}) \parallel \hat{\mathcal{P}}(X_{(0)}, Y_{(0)})] - \Delta_i \mathcal{S}_{(t)} \\ &= \mathcal{D}[\mathcal{P}(X_{(t)}, Y_{(t)}) \parallel \hat{\mathcal{P}}(X_{(t)}) \hat{\mathcal{P}}(Y_{(t)})] \end{aligned} \quad (7)$$

$$\geq \mathcal{I}(X_{(t)}; Y_{(t)}), \quad (8)$$

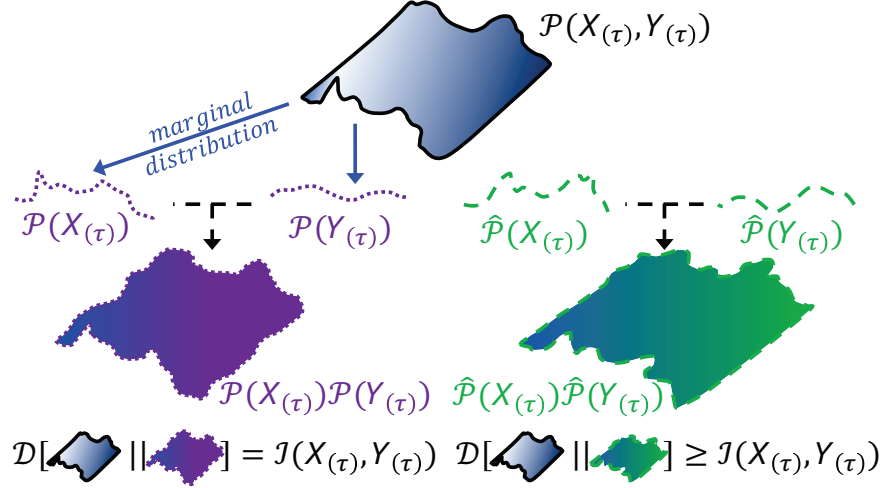


Figure 3: The relation between  $\mathcal{I}(X_{(\tau)}, Y_{(\tau)})$  and  $\mathcal{D}[\mathcal{P}(X_{(\tau)}, Y_{(\tau)}) || \hat{\mathcal{P}}(X_{(\tau)}) \hat{\mathcal{P}}(Y_{(\tau)})]$ .

where the equality is satisfied only when  $\mathcal{P}(X_{(t)}) \mathcal{P}(Y_{(t)}) = \hat{\mathcal{P}}(X_{(t)}) \hat{\mathcal{P}}(Y_{(t)})$ . Moreover, we prove that the **UCE Condition** is necessary for deriving (6) in appendix B. Therefore, the thermodynamics of encoding described in (7-8) can not be independent of the **UCE Condition**.

Before we move on, let us rethink about the **UCE Condition**. A possible question might be about the physical ground underlying its definition:  $\hat{\mathcal{P}}(X_{(t)}, Y_{(t)}) = \hat{\mathcal{P}}(X_{(t)}) \hat{\mathcal{P}}(Y_{(t)})$ . Here we emphasize that any thermodynamics variation costs time. Although the coupling between  $X$  and  $Y$  is closed at moment  $t$ , one can not expect that  $\mathcal{P}(X_{(t)}, Y_{(t)})$  jump to the independent state  $\mathcal{P}(X_{(t)}) \mathcal{P}(Y_{(t)})$  without time delay. Therefore, the approaching of  $(X_{(t)}, Y_{(t)})$  to independence can only be treated as an internal tendency at moment  $t$ .

### 3 Information thermodynamics during encoding

#### 3.1 Problem setup

Different from the classic ideas, our research concentrates on a new aspect of encoding. We notice that  $X$  consists of multiple sub-systems, and every sub-system involves encoding as well. An exploration of the relative role of each sub-system (or individual element) in information encoding (system coupling) helps reveal the intrinsic order of the system [83]. The significance of such an analysis has been discussed in information theory [101], statistical physics [83], and neuroscience [68, 1, 84]. Combine this new aspect with the classic ideas, an important question is whether (8) can be generalized to these sub-systems. Specifically, we wonder if the encoded information of  $Y$  in an arbitrary sub-system of  $X$  is bounded by the irreversible work  $\mathcal{W}_{\text{irr}}$  from the joint system  $(X, Y)$  following (8) as well.

Consider any sub-system  $X' = (\{X_O\}, \mathcal{R}')$  of  $X$  (here  $O \subseteq \{1, \dots, n\}$  is the index set), we mark a corresponding sub-system  $\overline{X'} = (\{X_i\} - \{X_O\}, \overline{\mathcal{R}'})$ . We denote the relations between  $X'$  and  $\overline{X'}$  utilizing  $\mathcal{R}(X'_{(t)}, \overline{X'}_{(t)})$ . Then, we can rewrite the whole system as a tuple  $X = ((X'_{(t)}, \overline{X'}_{(t)}), \mathcal{R}(X'_{(t)}, \overline{X'}_{(t)}))$ . Depending on the nature of  $\mathcal{R}(X'_{(t)}, \overline{X'}_{(t)})$ , we can distinguish between two cases of system  $X$  where  $\mathcal{R}(X'_{(t)}, \overline{X'}_{(t)}) \equiv \emptyset$  (one can see the case 3 in Fig. 4) or  $\mathcal{R}(X'_{(t)}, \overline{X'}_{(t)}) \neq \emptyset$  (cases 1, 2, and 4 in Fig. 4). These definitions allow reformulating (1) under the **UCE Condition** as

$$\frac{\mathcal{W}_{\text{irr}}}{T} \geq \mathcal{I}[(X'_{(t)}, \overline{X'}_{(t)}); Y_{(t)}] - \mathcal{D}[\mathcal{P}(X_{(0)}, Y_{(0)}) || \hat{\mathcal{P}}(X_{(0)}, Y_{(0)})] + \Delta_i \mathcal{S}_{(t)}. \quad (9)$$

Given (9), a natural thought for solving our question on the generalization of (8) is to compare  $\mathcal{I}[(X'_{(t)}, \overline{X'}_{(t)}); Y_{(t)}]$  and  $\mathcal{I}(X'_{(t)}; Y_{(t)})$ . Below, we demonstrate that the solution to this question can be developed based on fundamental information theory tools and further leads to an interesting view regarding non-isolated systems.

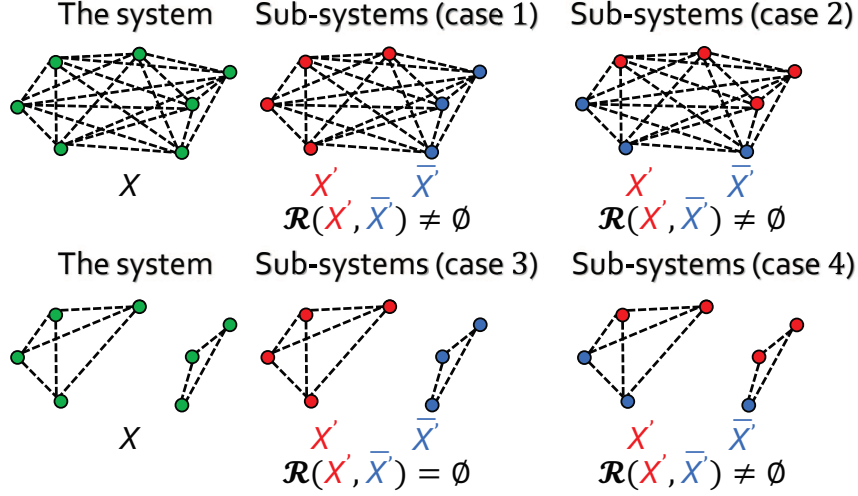


Figure 4: Examples of the sub-system selection with different attributes of  $\mathcal{R}(X'_{(t)}, \overline{X}'_{(t)})$ .

### 3.2 Encoding without intra-system coupling

We start the comparison in a simple case where  $\mathcal{R}(X'_{(t)}, \overline{X}'_{(t)}) \equiv \emptyset$ , namely that  $X$  can always be divided into two uncorrelated sub-systems (see the left panel of Fig. 6). In this case, we know that  $\mathcal{I}(X'_{(t)}; \overline{X}'_{(t)}) \equiv 0$  is ensured by the independence between  $X'_{(t)}$  and  $\overline{X}'_{(t)}$ . Moreover, we can find the item  $\mathcal{I}((X'_{(t)}, \overline{X}'_{(t)}); Y_{(t)})$  in the 3-order multiple mutual information [64, 83]

$$\begin{aligned} \mathcal{I}(X'_{(t)}; \overline{X}'_{(t)}; Y_{(t)}) &= \mathcal{I}(X'_{(t)}; Y_{(t)}) + \mathcal{I}(\overline{X}'_{(t)}; Y_{(t)}) - \\ &\quad \mathcal{I}[(X'_{(t)}, \overline{X}'_{(t)}); Y_{(t)}]. \end{aligned} \quad (10)$$

Such an equality helps connect between  $\mathcal{I}(X'_{(t)}; Y_{(t)})$  and  $\mathcal{I}((X'_{(t)}, \overline{X}'_{(t)}); Y_{(t)})$  (see Fig. 5). We need to remind that equality (4) (or an equivalent formulation of it proposed by [98]) can not be calculated directly with our general settings. An estimation for it is necessary. By applying Yeung's inequality [103] on (10), we can derive

$$\begin{aligned} &\mathcal{I}(X'_{(t)}; Y_{(t)}) + \mathcal{I}(\overline{X}'_{(t)}; Y_{(t)}) - \mathcal{I}[(X'_{(t)}, \overline{X}'_{(t)}); Y_{(t)}] \\ &\leq \min \{ \mathcal{I}(X'_{(t)}; Y_{(t)}), \mathcal{I}(\overline{X}'_{(t)}; Y_{(t)}), \mathcal{I}(X'_{(t)}; \overline{X}'_{(t)}) \} \\ &\leq 0. \end{aligned} \quad (11)$$

Based on (11), we know that  $\mathcal{I}[(X'_{(t)}, \overline{X}'_{(t)}); Y_{(t)}] \geq \max \{ \mathcal{I}(X'_{(t)}; Y_{(t)}), \mathcal{I}(\overline{X}'_{(t)}; Y_{(t)}) \}$  (see Fig. 5 and the left panel of Fig. 6). Therefore, one can immediately obtain

$$\frac{\mathcal{W}_{\text{irr}}}{T} + \mathcal{D}[\mathcal{P}(X_{(0)}, Y_{(0)}) \parallel \widehat{\mathcal{P}}(X_{(0)}, Y_{(0)})] \geq \max \{ \mathcal{I}(X'_{(t)}; Y_{(t)}), \mathcal{I}(\overline{X}'_{(t)}; Y_{(t)}) \}, \quad (12)$$

proving that (8) holds for sub-systems in this case.

Let us suspend the analysis and review the above derivation. For the encoding without intra-system coupling, the item  $\mathcal{I}[(X'_{(t)}, \overline{X}'_{(t)}); Y_{(t)}]$  in (3) can be connected with  $\mathcal{I}(X'_{(t)}; Y_{(t)})$  by the three order multiple mutual information [64, 83] in (10). Utilizing Yeung's inequality [103] in (11), we can implement the comparison in (6) and demonstrate that (8) can be generalized to sub-systems in this case. Considering the arbitrary selection of  $X'$ , the condition  $\mathcal{R}(X'_{(t)}, \overline{X}'_{(t)}) \equiv \emptyset$  means that each element in  $X$  is independent. An intuitive discovery can be identified in such a case, suggesting that the information that can be encoded by any sub-system is bounded by the encoded information base on the whole system.



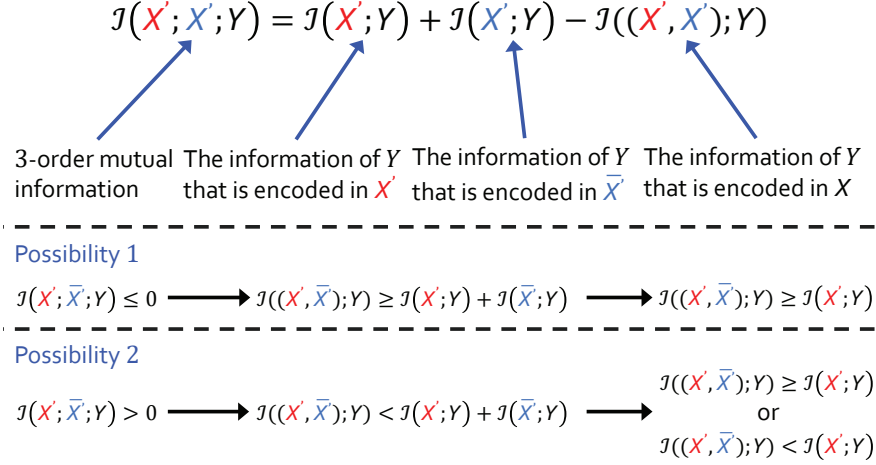


Figure 5: A schematic diagram of the 3-order mutual information defined in (10) and our idea underlying inequality (11), suggesting the necessity to verify if the 3-order mutual information can be positive.

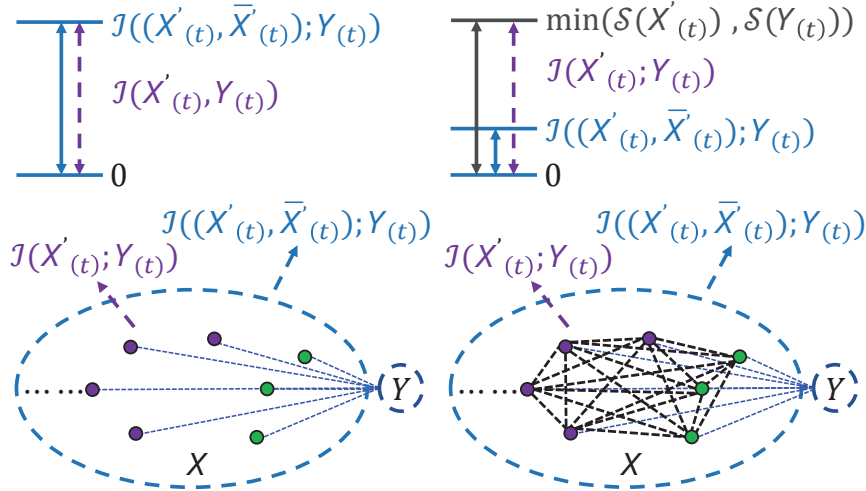


Figure 6: The encoding processes with (right panel) or without (left panel) intra-system coupling. An arbitrary sub-system  $X'$  is marked as purple and the corresponding sub-system  $\bar{X}'$  is marked as green.

Furthermore, one can obtain more knowledge about the above discovery after applying the theory of information synergy and redundancy [64, 83]. More specifically, there is synergy between  $X'_{(t)}$  and  $\bar{X}'_{(t)}$  if (10) is negative, namely that  $X'_{(t)}$  and  $\bar{X}'_{(t)}$  taken together (denoted by  $(X'_{(t)}, \bar{X}'_{(t)})$  in (10)) can encode more information of  $Y_{(t)}$  than the case when they are taken separately. If (10) is positive, then the pair  $(X'_{(t)}, \bar{X}'_{(t)})$  is redundant in the information encoding about  $Y_{(t)}$ . Based on (11), one can find that there is no information redundancy in the system of independent elements. Moreover, one can easily prove that the equality in (11) can be satisfied if and only if  $\mathcal{I}(X'_{(t)}; \bar{X}'_{(t)} | Y_{(t)}) = 0$ , because  $\mathcal{I}(X'_{(t)}; \bar{X}'_{(t)}; Y_{(t)}) = \mathcal{I}(X'_{(t)}; \bar{X}'_{(t)}) - \mathcal{I}(X'_{(t)}; \bar{X}'_{(t)} | Y_{(t)})$  [22] and  $\mathcal{I}(X'_{(t)}; \bar{X}'_{(t)}) = 0$ . Applying the definition of conditional mutual information [22], it can be known that  $\mathcal{I}(X'_{(t)}; \bar{X}'_{(t)} | Y_{(t)}) = 0$  is equivalent to the conditionally independence between  $X'_{(t)}$  and  $\bar{X}'_{(t)}$  given  $Y_{(t)}$

$$\mathcal{P}(X'_{(t)}; \bar{X}'_{(t)} | Y_{(t)}) = \mathcal{P}(X'_{(t)} | Y_{(t)}) \mathcal{P}(\bar{X}'_{(t)} | Y_{(t)}). \quad (13)$$

There exists information synergy in system  $X$  unless (13) is satisfied. Considering that (13) is a strong condition, one can generally treat it as a rare case for a system of independent elements to have no information synergy (see the top panel of Fig. 7a).

### 3.3 Encoding with intra-system coupling

We turn to the comparison in a more complex case where  $\mathcal{R}(X'_{(t)}, \overline{X}'_{(t)}) \neq \emptyset$ , namely that  $X$  is a system with non-negligible intra-system coupling (the right panel of Fig. 6). In this case, we can see  $\mathcal{I}(X'_{(t)}; \overline{X}'_{(t)}) \geq 0$ . Following the same idea that has been shown in (10-12), we know that  $\mathcal{I}(X'_{(t)}; Y_{(t)}) + \mathcal{I}(\overline{X}'_{(t)}; Y_{(t)}) - \mathcal{I}[(X'_{(t)}, \overline{X}'_{(t)}); Y_{(t)}] \leq \min \{ \mathcal{I}(X'_{(t)}; Y_{(t)}), \mathcal{I}(\overline{X}'_{(t)}; Y_{(t)}), \mathcal{I}(X'_{(t)}; \overline{X}'_{(t)}) \}$  is ensured by Yeung's inequality [103]. Note that one can easily find that the right side of the inequality is either zero or positive in this case. Therefore, the upper bound in Yeung's inequality fails in determining the sign of item  $\mathcal{I}(X'_{(t)}; Y_{(t)}) + \mathcal{I}(\overline{X}'_{(t)}; Y_{(t)}) - \mathcal{I}[(X'_{(t)}, \overline{X}'_{(t)}); Y_{(t)}]$ . Combining this finding with (9), one can prove that the bound in (8) does not always hold for sub-systems  $X'$  and  $\overline{X}'$ . Although the encoded information of  $Y$  in system  $X$  is bounded by the irreversible work  $\mathcal{W}_{\text{irr}}$  from the joint system  $(X, Y)$  following (8), an arbitrary sub-system  $X'$  (or  $\overline{X}'$ ) might disobey this bound by encoding more information than  $X$  (see the right panel of Fig. 6). We should emphasize that this does not disobey the second law of thermodynamics. The encoded information in sub-system  $X'$  is still bound by the irreversible work  $\mathcal{W}'_{\text{irr}}$  from  $(X', Y)$  following  $\frac{\mathcal{W}'_{\text{irr}}}{kT} + \mathcal{D}[\mathcal{P}(X'_{(0)}, Y_{(0)}) \| \hat{\mathcal{P}}(X'_{(0)}, Y_{(0)})] \geq \mathcal{I}(X'_{(t)}; Y_{(t)})$ . Moreover, one can measure the maximum extent that an arbitrary sub-system  $X'$  can disobey the bound offered by the irreversible work  $\mathcal{W}_{\text{irr}}$ . An elementary bound

$$\mathcal{I}(X'_{(t)}; Y_{(t)}) - \mathcal{I}[(X'_{(t)}, \overline{X}'_{(t)}); Y_{(t)}] \leq \min \{ \mathcal{S}(X'_{(t)}), \mathcal{S}(Y_{(t)}) \} \quad (14)$$

can be derived from Shannon entropy  $\mathcal{S}$  (the right panel of Fig. 6). Nevertheless, the non-trivial bound remains elusive and requires more explorations in future.

One can see that the encoding by a system with non-negligible intra-system coupling might involve both information synergy and redundancy. The synergy case is in accord with our common sense of the relation between a system and its sub-systems. The redundancy case (where the bound offered by the irreversible work  $\mathcal{W}_{\text{irr}}$  is disobeyed by sub-systems) corresponds to the situation where an arbitrary sub-system encodes more information than the whole system. Such a case results from the in-harmony between sub-systems in encoding (see the bottom panel of Fig. 7a).

To this point, we have proposed our answer to the question on the generalization of (8). We demonstrate that the irreversible work  $\mathcal{W}_{\text{irr}}$  from  $(X, Y)$  can bound the encoded information of  $Y$  in any sub-system of  $X$  if there is no intra-system coupling. Otherwise, this bound might be disobeyed by the sub-systems of  $X$  (see Fig. 7). We have discussed these findings from the aspect of information synergy and redundancy [64, 83]. These discussions lead to an interesting view that a non-isolated system without intra-system coupling frequently involves information synergy, in comparison, the system with intra-system coupling can involve either information synergy or redundancy (see Fig. 7).

Rather than limit ourselves in treating the redundancy as specific information inefficiency, we suggest a new perspective that the redundancy means a possibility to hide information in a non-isolated system. By observing the irreversible work  $\mathcal{W}_{\text{irr}}$  from a joint system  $(X, Y)$ , one might underestimate the information of  $Y$  encoded in  $X$  (underestimate the inter-system coupling strength) if  $X$  consists of correlated elements. In such a case, more information of  $Y$  can be hidden in the sub-systems of  $X$  (see an illustration in Fig. 7b).

Below, we implement our analysis of this perspective by applying our theoretical framework to concrete non-isolated systems. These systems come from statistical physics and neuroscience, helping demonstrate the ubiquity of the discussed phenomenon above.

## 4 Computational verification in non-isolated systems

### 4.1 Ising model with random external field

We begin our demonstration with the Ising model [20], one of the significant themes in physics. Specifically, we consider a 1-dimensional Ising model with random external field, whose Hamiltonian is

$$\mathbf{H}(\sigma, h) = -J \sum_{\langle i, j \rangle} \sigma_i \sigma_j - h \sum_i \sigma_i. \quad (15)$$

Here  $\sigma$  is a shorthand for a specification of the spin  $\sigma_i$  on each site. The notion  $J$  denotes the coupling strength between spins (the intra-system coupling). The external field  $h \in (0, m] \cap \mathbb{Z}$  is defined as a discrete random variable (the inter-system coupling)

$$\mathcal{P}(h) = \frac{1}{m}. \quad (16)$$

To this point, a spin system  $\sigma$  and an external source  $h$  have been characterized, functioning as the systems  $X$  and  $Y$  in our previous discussion, respectively. Here we define a random external field  $h$  rather than a fixed one since it is more natural to consider a random variable  $h$  in our subsequent encoding analysis. One can find other similar models in [3, 104].

Given a  $h$ , we know that system  $\sigma$  always has a unique equilibrium Boltzman distribution at a constant temperature  $T$  [74, 58]

$$\hat{\mathcal{P}}(\sigma | h) = \frac{1}{Z(h)} \exp \left[ -\frac{1}{kT} \mathbf{H}(\sigma, h) \right], \quad (17)$$

where  $k$  is the Boltzmann constant, and  $Z(h)$  denotes the normalization term (partition function)

$$\begin{aligned} Z(h) &= \sum_{\sigma} \exp \left[ -\frac{1}{kT} \mathbf{H}(\sigma, h) \right] \\ &= \exp \left( \frac{nJ}{kT} \right) \left[ \cosh \frac{h}{kT} + \sqrt{\sinh^2 \frac{h}{kT} + \exp(-4 \frac{J}{kT})} \right]^n \\ &\quad + \exp \left( \frac{nJ}{kT} \right) \left[ \cosh \frac{h}{kT} - \sqrt{\sinh^2 \frac{h}{kT} + \exp(-4 \frac{J}{kT})} \right]^n, \end{aligned} \quad (18)$$

where  $n$  is the number of spins. Assuming (16) is the equilibrium distribution  $\hat{\mathcal{P}}(h)$  of  $h$ , we can calculate the equilibrium distribution of joint system  $(\sigma, h)$

$$\hat{\mathcal{P}}(\sigma, h) = \frac{\hat{\mathcal{P}}(h)}{Z(h)} \exp \left[ -\frac{1}{kT} \mathbf{H}(\sigma, h) \right]. \quad (19)$$

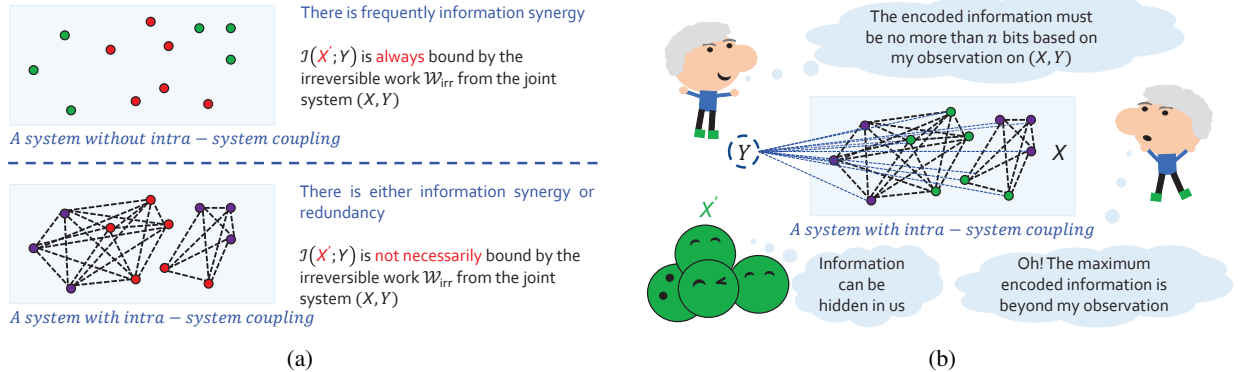


Figure 7: Summary of theoretical findings. (a) The differences between the encoding processes in systems with/without intra-system coupling. (b) The illustration of hiding information in non-isolated system without intra-system coupling.

Now, let us consider the thermodynamics of encoding in the joint system  $(\sigma, h)$  during  $[0, t]$  (e.g., try to calculate inequality (1)). We treat system  $\sigma$  as an information thermodynamics encoder to encode the information of  $h$ . Given the above review, we can calculate the equilibrium distribution  $\hat{\mathcal{P}}(\sigma_{(\tau)}, h_{(\tau)})$  at any moment  $\tau \in [0, t]$ . When  $\sigma$  is coupled with an external field  $h \in (0, m]$ , the equilibrium distribution of  $(\sigma, h)$  can be worked out following (19). When the **UCE Condition** is met at moment  $t$ , one can see that  $\hat{\mathcal{P}}(\sigma_{(t)}) = \hat{\mathcal{P}}(\sigma_{(t)} | 0)$ , where  $\hat{\mathcal{P}}(\sigma_{(t)} | 0)$  measures the equilibrium distribution of  $\sigma$  independent from the external field  $h$ . Thus,

$$\hat{\mathcal{P}}(\sigma_{(t)}, h_{(t)}) = \hat{\mathcal{P}}(\sigma_{(t)} | 0) \hat{\mathcal{P}}(h_{(t)}). \quad (20)$$

As for the work  $\mathcal{W}$  performed on system  $(\sigma, h)$ , it can be calculated under the framework of stochastic thermodynamics [86]

$$\dot{\mathcal{W}} = \sum_h \sum_{\sigma} \mathcal{P}(\sigma | h) \mathcal{P}(h) \dot{\mathbf{H}}(\sigma, h). \quad (21)$$

For convenience, we implement the experiment in a special case where  $h_{(t)} = h_{(0)}$  and the **UCI Condition** is satisfied. One can know that  $\hat{\mathcal{P}}(\sigma_{(0)}, h_{(0)}) = \hat{\mathcal{P}}(\sigma_{(t)}, h_{(t)})$  and therefore the change in equilibrium free energy  $\Delta\hat{\mathcal{F}} = 0$ . Under this condition, one can derive  $\mathcal{W}_{irr} = \mathcal{W}$ . Taken together, the most part of (1) has been calculated to this point.

However, one can immediately find it non-trivial to define the time-dependent distribution  $\mathcal{P}(\sigma_{(\tau)}, h_{(\tau)})$  analytically. To overcome this challenge, we turn to the computational implementation of (1).

Computationally, we generate  $m \times k_J \times k_T \times k_{\theta}$  random sequences of the external field, where each random sequence  $(h_{(0)}, \dots, h_{(t)})$  corresponds to an experiment condition. The condition includes a possible variation process of external field that begins and ends with a unique  $h_{(0)} = h_{(t)} \in (0, m] \cap \mathbb{Z}$ . This sequence features a certain transition rate  $\theta^{-1}$  (there are  $k_{\theta}$  transition rates that quantify the variation speed of the external field). Meanwhile, each condition corresponds to a coupling strength  $J$  and temperature  $T$  (there are  $k_J$  and  $k_T$  kinds of coupling strength and temperature quantities, respectively). Under each experiment condition, the simulation experiment is implemented as following

- We randomly initialize a specification  $\sigma_{(0)}$  at moment 0;
- We perform the continuous-time Monte Carlo simulation utilizing the Metropolis algorithm  $(\sigma_{(0)}, \dots, \sigma_{(t)}) = \text{Metropolis}[\sigma_{(0)}, (h_{(0)}, \dots, h_{(t)})]$  [9]. The simulation lasts for a duration  $[0, t]$ . During the simulation, we use the update of the external field  $h$  to drive the Metropolis algorithm. The work is initialized as  $\mathcal{W} = 0$ . Every time when the spin specification changes from  $\sigma_{(\tau)}$  to  $\sigma'_{(\tau+1)}$ , we add  $\frac{\mathbf{H}(\sigma_{(\tau)}, h_{(\tau)}) - \mathbf{H}(\sigma'_{(\tau+1)}, h_{(\tau+1)})}{T}$  to  $\frac{\mathcal{W}}{T}$ ;
- The Metropolis algorithm  $(\sigma_{(0)}, \dots, \sigma_{(t)}) = \text{Metropolis}[\sigma_{(0)}, (h_{(0)}, \dots, h_{(t)})]$  runs  $\xi$  times in total. The  $i$ -th simulation corresponds to a final state  $\sigma^i_{(t)}$ . Then, we estimate  $\mathcal{P}(\sigma_{(t)} | h_{(t)})$  based on the frequency distribution on  $\{\sigma^1_{(t)}, \dots, \sigma^{\xi}_{(t)}\}$ .

Based on the above procedure, we can computationally obtain the probability distribution  $\mathcal{P}(\sigma_{(t)} | h_{(t)})$  and the work  $\mathcal{W}$  (or the irreversible work  $\mathcal{W}_{irr}$ ). Moreover, we can further calculate the term  $\mathcal{D}[\mathcal{P}(\sigma_{(t)}, h_{(t)}) || \hat{\mathcal{P}}(\sigma_{(t)}, h_{(t)})]$  following (20). As for  $\mathcal{P}(\sigma_{(0)}, h_{(0)})$ , we assume it is uniformly distributed on the sample space  $\Omega(\sigma) \times \Omega(h)$ . Given this assumption, we can calculate the term  $\mathcal{D}[\mathcal{P}(\sigma_{(0)}, h_{(0)}) || \hat{\mathcal{P}}(\sigma_{(0)}, h_{(0)})]$  under the **UCI Condition**. Finally, we can directly obtain the entropy production  $\Delta_i \mathcal{S}_{(t)}$  since other parts of (1) have been given. We implement the computational experiment using the parameter settings in appendix C. The spin system  $\sigma$  is set with a small size, allowing us to do a relatively exhaustive sampling.

Fig. 8a shows the variation trends of every term of (1) along with the increasing  $J$ ,  $T$ , or  $\theta$  conditions. The corresponding mutual information calculated based on (5) is illustrated as the function of  $J$ ,  $T$ , or  $\theta$  in Fig. 8b. One can see that the mutual information (mean value) features a positive correlation with the inverse transition rate  $\theta$  (logarithmic value). The correlation is strong with a Pearson coefficient  $R = 0.9429$  and significant at  $p < 10^{-5}$ , suggesting that the spin system  $\sigma$  encodes more information about the external field  $h$  when it has more relaxation time between any two times of external field variations (larger  $\theta$ ). Apart of that, more complex variation trends of the mutual information can be found along with the coupling strength  $J$  and the temperature  $T$ .

Featuring intra-system coupling relations (the interactions between different spins), the spin system  $\sigma$  is expected to allow the encoded information by its sub-systems to exceed the encoded information by itself. To verify this speculation, we randomly select sub-systems  $\sigma'$  (and the corresponding  $\overline{\sigma'}$ ) 100 times under each experiment conditions. As

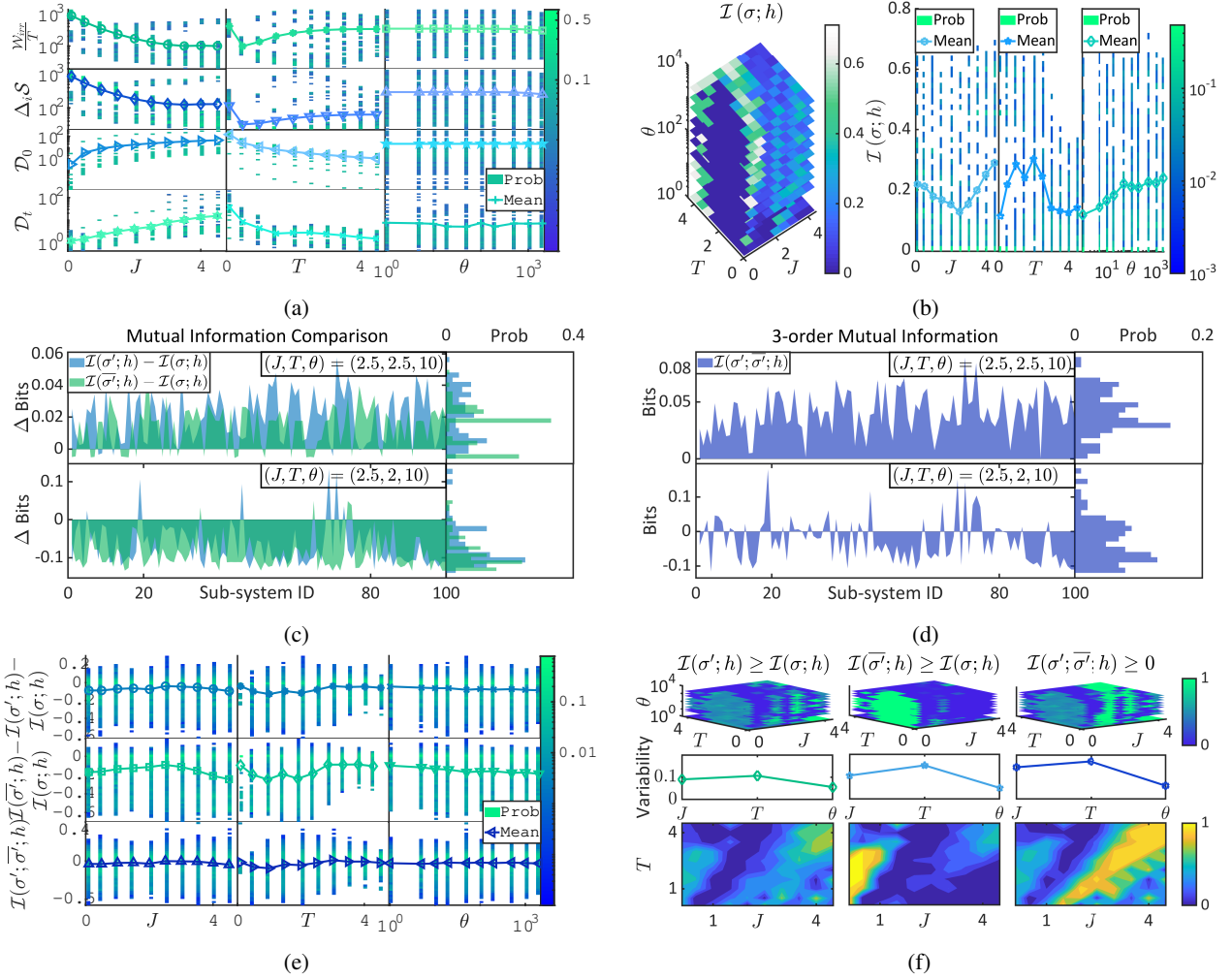


Figure 8: Information thermodynamics in the Ising model. (a) The calculated terms (mean value and probability distributions) of (1) are shown as the functions of  $J$ ,  $T$ , and  $\theta$ . While analyzing each parameter direction among  $J$ ,  $T$ , and  $\theta$ , the other two directions are averaged. (b) The calculated mutual information quantities (raw data distributions, mean value, and probability distributions) act as the functions of  $J$ ,  $T$ , and  $\theta$ . (c-d) Two representative examples of the mutual information comparison and their corresponding 3-order mutual information. (e) The mutual information comparison result and the corresponding 3-order mutual information are analyzed as the functions of  $J$ ,  $T$ , and  $\theta$ . (f) The probability distributions of the cases where  $\mathcal{I}(\sigma'_t; h_t) \geq \mathcal{I}(\sigma_t; h_t)$ ,  $\mathcal{I}(\sigma'_t; \sigma'_t; h_t) \geq \mathcal{I}(\sigma_t; h_t)$ , or  $\mathcal{I}(\sigma'_t; \sigma'_t; h_t) \geq 0$  are shown as the functions of  $J$ ,  $T$ , and  $\theta$  (upper line). We also compare the probability variability of these cases in each parameter direction (middle line). Then, we analyze the probability distributions of these cases in the directions of  $J$  and  $T$  due to the high variability (bottom line).

shown by the examples in Fig. 8c, one can see that the information quantity differences  $\mathcal{I}(\sigma'_t; h_t) - \mathcal{I}(\sigma_t; h_t)$  and  $\mathcal{I}(\sigma'_t; \sigma'_t; h_t) - \mathcal{I}(\sigma_t; h_t)$  can be either non-positive (meaning that the bound (8) is followed by sub-systems) or positive (the bound (8) may be disobeyed by sub-systems). Moreover, one can find the corresponding positive 3-order mutual information  $\mathcal{I}(\sigma'_t; \sigma'_t; h_t)$  in Fig. 8d, which suggests the existence of information redundancy or synergy.

Although the experiment condition  $(J, T, \theta)$  has significant effects on  $\mathcal{I}(\sigma'_t; h_t) - \mathcal{I}(\sigma_t; h_t)$ ,  $\mathcal{I}(\sigma'_t; \sigma'_t; h_t) - \mathcal{I}(\sigma_t; h_t)$ , and  $\mathcal{I}(\sigma'_t; \sigma'_t; h_t)$  in Fig. 8c-8d, we suggest that these effects do not show clear trends if they

are attributed to a single parameter among  $J$ ,  $T$ , and  $\theta$  in Fig. 8e. Therefore, we turn to search for underlying patterns from the probabilistic perspective and implement a unified analysis. In Fig. 8f, we investigate the probability distributions of  $\mathcal{I}(\sigma'_{(t)}; h_{(t)}) \geq \mathcal{I}(\sigma_{(t)}; h_{(t)})$ ,  $\mathcal{I}(\bar{\sigma}'_{(t)}; h_{(t)}) \geq \mathcal{I}(\sigma_{(t)}; h_{(t)})$ , and  $\mathcal{I}(\sigma'_{(t)}; \bar{\sigma}'_{(t)}; h_{(t)}) \geq 0$  as the functions of  $(J, T, \theta)$ . After measuring the mean variability of these probability distributions in each parameter direction (e.g., the probability variability in the direction of  $J$  at  $(J, T, \theta)$  is defined as the variance  $\text{Var}_J[\mathcal{P}(J, T, \theta)]$ ). Then the mean variability is quantified as  $\mathbb{E}_{T, \theta}\{\text{Var}_J[\mathcal{P}(J, T, \theta)]\}$ , we confirm that the variability of these probability distributions can be better explained by  $(J, T)$  rather than  $\theta$ . This finding inspires us to concentrate on the effects of  $(J, T)$  on the probability distributions. The bottom panel of Fig. 8f demonstrates that the probability densities of  $\mathcal{I}(\sigma'_{(t)}; h_{(t)}) \geq \mathcal{I}(\sigma_{(t)}; h_{(t)})$ ,  $\mathcal{I}(\bar{\sigma}'_{(t)}; h_{(t)}) \geq \mathcal{I}(\sigma_{(t)}; h_{(t)})$ , and  $\mathcal{I}(\sigma'_{(t)}; \bar{\sigma}'_{(t)}; h_{(t)}) \geq 0$  relatively concentrate on the sub-region where  $J \sim T$  (or equivalently,  $\min(\frac{J}{T}, \frac{T}{J}) \sim 1$ ). Because the temperature  $T$  is not an inherent characteristic of the spin system  $\sigma$ , our research focuses on the coupling strength  $J$ . As the temperature  $T$  increases, the coupling strength  $J$  where these probability densities concentrate will also increase. In other words, the cases where the encoded information of  $h$  by a sub-system  $\sigma'$  (or  $\bar{\sigma}'$ ) exceeds  $\mathcal{I}(\sigma; h)$  and the cases with information redundancy frequently emerge when the coupling strength  $J$  (the intra-system coupling) is relatively proportional to the temperature  $T$ .

Assuming there is an observer who pursues to estimate the maximum encoded information of  $h$  in  $\sigma$ . Given our findings above, the encoded information might be underestimated if this observer implements the observation outside the system  $(\sigma, h)$  (e.g., the observer estimates the encoded information utilizing the irreversible work  $\mathcal{W}_{\text{irr}}$  from system  $(\sigma, h)$  and (8)). When the encoded information  $\mathcal{I}(\sigma'; h)$  by a sub-system  $\sigma'$  (or  $\bar{\sigma}'$ ) can exceed  $\mathcal{I}(\sigma; h)$ , a certain amount of encoded information (measured as  $\mathcal{I}(\sigma'; h) - \mathcal{I}(\sigma; h)$  or  $\mathcal{I}(\bar{\sigma}'; h) - \mathcal{I}(\sigma; h)$ ) is hidden from the observer.

## 4.2 Real data of the human brain during the perception process

Perhaps the brain is the most natural example of information thermodynamics encoder with intra-system coupling. We choose the brain as our second demonstration since neuroscience studies can offer vast amounts of data for  $X$  (neural data) and a clear characterization for  $Y$  (stimulus data). Nevertheless, we need to note that the temperature  $T$  and the irreversible work  $\mathcal{W}_{\text{irr}}$  are not well-defined concepts for the brain and can not be calculated directly. Therefore, the analysis can only be performed indirectly.

Specifically, we implement our analysis on an open-source functional magnetic resonance imaging (fMRI) data set obtained from the visual object recognition experiment, where the random stimulus sequence consists of 8 kinds of objects [38]. The fMRI technique measures neural activities through the blood-oxygen-level-dependent (BOLD) contrast in the magnetic field [40]. This data set includes the fMRI signals of 6 subjects from a 3 Tesla scanner, covering a high-resolution whole-brain region of 163840 voxels (small neural clusters) for each subject. Each voxel corresponds to a time series of the activities of a neural cluster. In the experiment, 12 pairs of (neural data, stimulus data) are obtained from each subject, and we arrange these pairs into a single matrix. Thus, every subject corresponds to a joint system  $(X, Y)$ . The validity of the data has been verified previously [37, 35, 67], ensuring the reproducibility of our analysis. Please see appendix D for more details.

To measure information quantities in real data set, we estimate the mutual information  $\hat{\mathcal{I}}$  following the computational approach proposed by Gao et al. [30]

$$\hat{\mathcal{I}}(X, Y) = \mathbb{E}_{\tau} \left[ \psi(\hat{k}_{(\tau)}) - \log(\mathcal{C}_X(\tau) + 1)(\mathcal{C}_Y(\tau) + 1) \right] + \log t, \quad (22)$$

where  $\psi(\cdot)$  denotes the digamma function. The estimation is implemented based on  $k$ -nearest neighbor (KNN) method, where each  $\tau \in [0, t]$  corresponds to a sample (here  $t$  is the length of data). Let  $\epsilon_{(\tau)}$  be the distance from  $(X_{(\tau)}, Y_{(\tau)})$  to its  $k$ -th neighbor with a given parameter  $k$  (in our research, we set  $k = 5$ ). One have  $\hat{k}_{(\tau)} = k$  if  $\epsilon_{(\tau)} > 0$ . Otherwise, one need to measure  $\hat{k}_{(\tau)} = |\{(X_{(\tau')}, Y_{(\tau')}) \mid (X_{(\tau')}, Y_{(\tau')}) = (X_{(\tau)}, Y_{(\tau)})\}|$ . In the sub-space of  $X$ , one can count the number  $\mathcal{C}_X(\tau)$  of the sample whose distance from  $X_{(\tau)}$  is no more than  $\epsilon_{(\tau)}$ . Similarly, one can measure  $\mathcal{C}_Y(\tau)$  in the sub-space of  $Y$ . In [30], the distance is defined utilizing the maximum norm. Unlike the classic estimator designed with  $3\mathcal{H}$ -principle that  $\hat{\mathcal{I}}(X, Y) = \hat{\mathcal{H}}(X) + \hat{\mathcal{H}}(Y) - \hat{\mathcal{H}}(X, Y)$  (e.g., see [54]), this estimation is more efficient in controlling systematic errors and dealing with mixture spaces [30].

It is computationally costly to calculate information quantities on the whole brain regarding a large number of voxels. Therefore, we implement our analysis utilizing a reverse sequence. Specifically, we randomly initialize a system of  $z$  voxels in each brain ( $z \in [400, 600]$ ). Then, we randomly add  $r$  voxels into this system in each iteration ( $r \in [50, 150]$  and every sequence lasts for 100 iterations). We repeat this random reverse sequence generation in each subject 10 times. Given these settings, the previous system and the newly added voxel set are complementary sub-systems of the

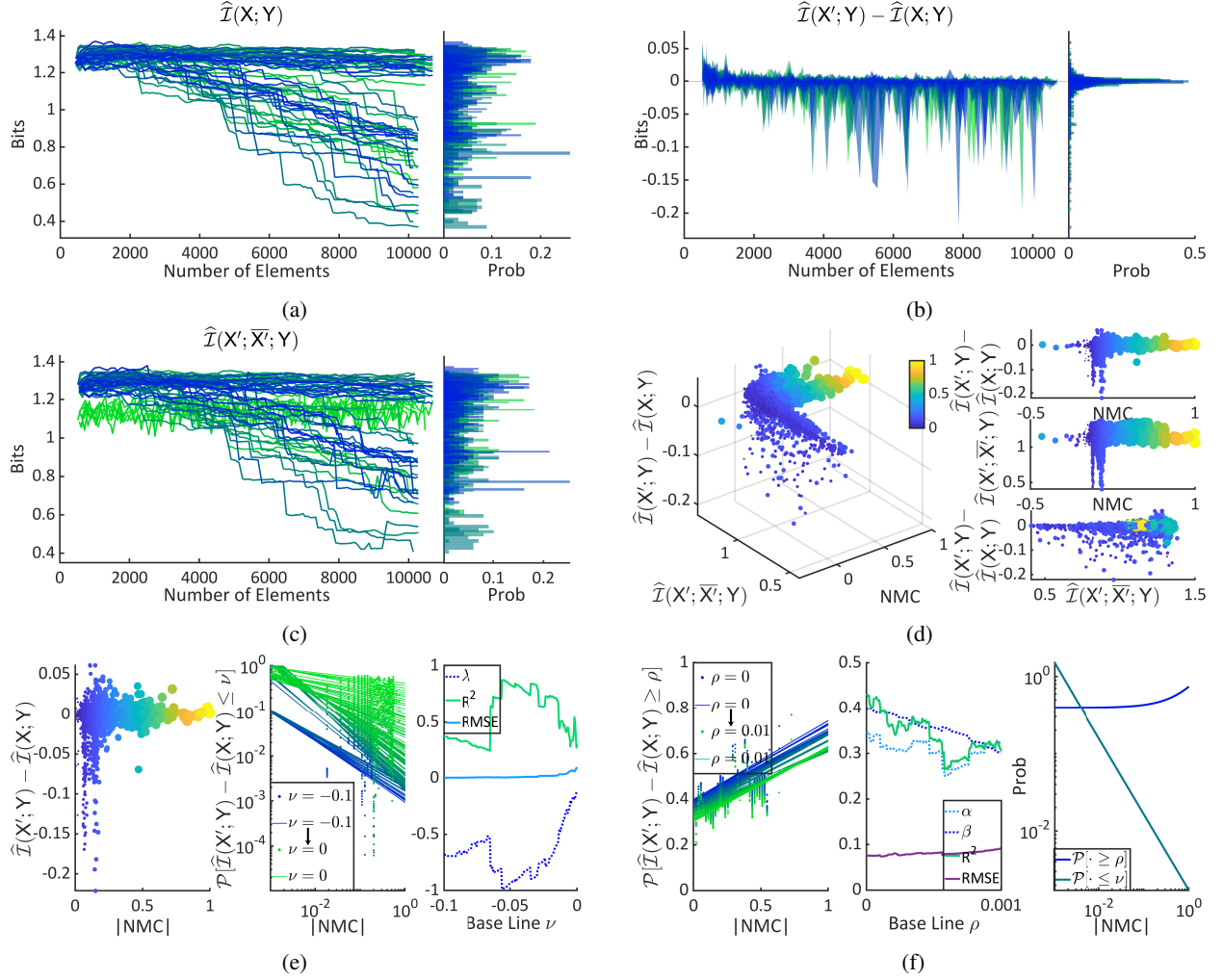


Figure 9: Information thermodynamics in the human brain. (a-c) The estimated information quantities and their probability distributions. (d) The relations between information quantities and the normalized mean correlation (NMC) are shown, where the sizes and colors of data points scale according to NMC. (e) We analyze the information quantity  $\hat{I}(X'_t; Y_t) - \hat{I}(X_t, Y_t)$  as a function of the absolute value of NMC (left). By adjusting the base line  $\epsilon$ , we can quantify the probability distribution of the case where  $\hat{I}(X'_t; Y_t) - \hat{I}(X_t, Y_t) \leq \nu$  under different standards. Our results suggest that this kind of case, irrespective of the selected base line, the probability  $\mathcal{P}[\cdot \leq \nu]$  follows a power regression model of  $|NMC|$  with reasonable fitting accuracy (middle and right). (f) Correspondingly, we set different base line  $\rho$  to identify the case where  $\hat{I}(X'_t; Y_t) - \hat{I}(X_t, Y_t) \geq \rho$ . The probability  $\mathcal{P}[\cdot \geq \rho]$  and  $|NMC|$  feature a linear regression relation with reasonable fitting accuracy, relatively independent of the base line settings (left and middle). Finally, we illustrate a representative instance of the probability distributions  $\mathcal{P}[\cdot \leq \nu]$  and  $\mathcal{P}[\cdot \geq \rho]$  where  $(\nu, \rho) = (-0.0545, 0)$  (right).

current system (can be referred to as  $X'$  and  $\overline{X'}$ ). Fig. 9a illustrates the variation trends of the estimated mutual information in these random sequences, suggesting that the system in the previous iteration (includes fewer voxels) might encode more stimulus information ( $Y$ ) than the current system (includes more voxels). The corresponding differences between the encoded information quantities in previous and current systems are shown in Fig. 9b. In Fig. 9c, one can see that  $\hat{I}(X'_t; \overline{X'_t}; Y_t) > 0$  always holds across all iterations, revealing the existence of information redundancy. Here we need to emphasize that the iteration is implemented with randomization and therefore frequently adds disharmonious voxels into the system to create redundancy. This result does not deny the possibility of information synergy.

To relate the above findings with the internal orders of the human brain, we quantify the mean Pearson correlation coefficient between the voxels that previously exist ( $X'$ ) and the newly added voxels ( $\bar{X}'$ ). The correlation coefficients between voxels in functional states, reflecting the relations between neural clusters during neural information processing, usually vary across different cognitive processes and individuals. Moreover, the neural correlation usually concentrates on specific sub-intervals rather than pervades the whole interval of  $[-1, 1]$ . Therefore, we further normalize the mean Pearson correlation to ensure the generality of our analysis (here, we normalize the coefficients based on their absolute values and maintain their signs). In Fig. 9d, we visualize the relations between the estimated information quantities and the normalized mean correlation coefficients (NMC). When NMC approaches 0, it can be seen that the cases where  $\hat{\mathcal{I}}(X'_{(t)}; Y_{(t)}) \leq \hat{\mathcal{I}}(X_{(t)}, Y_{(t)})$  frequently occur and the 3-order mutual information drops sharply. This phenomenon inspires us to investigate the variation trend of  $\hat{\mathcal{I}}(X'_{(t)}; Y_{(t)}) - \hat{\mathcal{I}}(X_{(t)}, Y_{(t)})$  as a function of the absolute value of NMC. We implement binning on  $|\text{NMC}|$  following the Freedman-Diaconis approach [29], which maintains relative robustness on the non-smooth data. Then, we exclude the bins where no more than 5 sample exist to obtain the filtered data (covers 99.48% of the raw data and excludes 0.52% of the raw data as outliers). Our results in Fig. 9e suggest that the random variable  $\hat{\mathcal{I}}(X'_{(t)}; Y_{(t)}) - \hat{\mathcal{I}}(X_{(t)}, Y_{(t)}) \leq \nu$  ( $\nu \in [-0.1, 0]$ ) follows specific power law distributions  $\mathcal{P}(\cdot \leq \nu) \propto |\text{NMC}|^{-\lambda}$  (here  $\lambda \in [0.1149, 0.9899]$ , it varies across different  $\nu$ ). Meanwhile, we discover that the random variable  $\hat{\mathcal{I}}(X'_{(t)}; Y_{(t)}) - \hat{\mathcal{I}}(X_{(t)}, Y_{(t)}) \geq \rho$  ( $\rho \in [0, 0.001]$ ) follows a probability distribution  $\mathcal{P}(\cdot \geq \rho) = \alpha|\text{NMC}| + \beta$  (here  $\alpha \in [0.2507, 0.3470]$  and  $\beta \in [0.3019, 0.3989]$ ) in Fig. 9f. The probability density function fitting of these results obtains reasonable fitting accuracy (please see Fig. 9e-9f). In Fig. 9f, we illustrate a representative example of the fitted probability distributions  $\mathcal{P}(\cdot \leq \nu)$  and  $\mathcal{P}(\cdot \geq \rho)$ . One can see that the probability  $\mathcal{P}(\cdot \leq \nu)$  ( $\nu < 0$ ) significantly decreases along with the increasing intra-system coupling strength (quantified by  $|\text{NMC}|$ ) while the probability  $\mathcal{P}(\cdot \geq \rho)$  ( $\rho \geq 0$ ) shows an opposite pattern. The finding is in consistency with our theory that a system with stronger intra-system coupling more possibly allow specific sub-systems to encode more information than itself.

To this point, we have demonstrated our theoretical results in the brain, a non-isolated system of neurons that has complex intra-system coupling and is coupled with external stimuli. Although we can not calculate (1) directly, the indirect analysis demonstrates that the encoded information in a sub-system of neural tissues is not necessarily bound by the irreversible work of the whole system. In the brain, the possibility for a sub-system to encode more stimulus information than the whole system is modified by the neural correlation (the intra-system coupling strength). Suppose we consider the observer perspective discussed in Subsec. 4.1, a similar conclusion can be drawn that a specific amount of encoded information may be hidden in sub-systems.

## 5 Discussion

The current research pursues to explore the fundamental role of the order inside a non-isolated system in shaping the information thermodynamics coupling between this system and the environment. We begin with a thermodynamic perspective to characterize an arbitrary non-isolated system  $X$  as an information thermodynamics encoder when this system is coupled with an external source  $Y$ . The information thermodynamics encoder encodes the information of external source utilizing thermodynamics, providing a unified angle to analyze the inter-system coupling between  $X$  and  $Y$  in term of information and thermodynamics. Rather than stand alone, our idea is rooted in extensive explorations of the physics nature of information (please see [69] for a systematic review). Furthermore, similarities and differences coexist between the proposed information thermodynamics encoder and information engines [24, 63, 62, 60, 17]. Although these two kinds of systems both bridge information and thermodynamics, information thermodynamics encoder mainly focuses on the thermodynamic costs of perceiving external information rather than the potential of information to function as specific thermodynamic fuel [11, 10, 91, 5, 41, 100, 65].

To formalize the encoding process with solid physics backgrounds, we try to find an appropriate way to derive the thermodynamics of encoding based on the nonequilibrium second law of thermodynamics. One can find a reverse derivation process from information quantities to the nonequilibrium second law of thermodynamics and fluctuation theorem [80, 79]. Although our proposed encoding process is suggested as a possible generalization of the measurement process [69, 42, 77, 33, 80, 36, 28], we need to emphasize that there exist essential consistency between these two concepts. Given the thermodynamic definition of encoding, a fundamental law is suggested that the mutual information  $\mathcal{I}(X, Y)$  (the encoded information of  $Y$  by system  $X$  entirety) is bounded by the irreversible work  $\mathcal{W}_{\text{irr}}$  from system  $(X, Y)$  based on (8). This upper bound is implied by the nonequilibrium second law of thermodynamics intrinsically [28].



Built on the above foundations, we turn to verify if this bound is followed by the encoded information of  $Y$  in an arbitrary sub-system of  $X$  as well. The necessity to study this problem lies in that it provides an opportunity to glance at the effects of internal orders inside  $X$  on the information thermodynamics link between  $X$  and  $Y$ . Despite the substantial progress in exploring the physics of information [69], the critical roles of these internal orders remain elusive. Exploring this problem may help understand information thermodynamics in complex systems, where the diversities of information thermodynamics characteristics might emerge on multiple scales due to heterogeneous elements and intricate internal correlations. As a starting point, the present study is limited at a qualitative level. Combining the 3-order multiple mutual information and the theory of information synergy and redundancy [64, 83], the proposed thermodynamics of encoding suggests an intrinsic difference between the non-isolated system with internal correlations and the non-isolated system of independent elements when they act as information thermodynamics encoders. Unlike those with independent elements, an encoder  $X$  with internal correlations allows the encoded information of  $Y$  in its sub-systems to exceed the information thermodynamics bound on the joint system  $(X, Y)$ . More specifically, the encoded information of  $Y$  in an arbitrary sub-system  $X'$  is not necessarily bound by the irreversible work  $\mathcal{W}_{irr}$  from system  $(X, Y)$  following (8). This difference may originate from the nature of order inside the system  $X$ . There is frequently information synergy in the non-isolated system of independent elements; in comparison, there can be either information synergy or information redundancy in the non-isolated system with intra-system coupling. These theoretical findings can be mathematically derived utilizing the Yeung's inequality [103]. Furthermore, we have computationally verified them in an Ising model with a random external field and a real data set of the human brain during the perception process. Our analysis demonstrates that the stronger internal correlation inside these systems can create a higher possibility for their sub-systems to encode more information than the global one.

The essential idea that guides our research is to analyze the thermodynamic coupling relation between a non-isolated system and an external source from the information-theoretical perspective. Similar ideas are pervaded in various physics disciplines and become the foundation of information thermodynamics thanks to Szilard's [95] and Landauer's [57, 56] inspiring works.

Recently, this idea has received increasing attention in neuroscience. Neuroscientists have identified the fundamental role of information thermodynamics in supporting brain functions [21, 92, 93, 89, 81, 66]. Historically, the physics mechanisms underlying brain functions used to be elusive [23]. The brain is a system with complex topology [61], geometry [105], and dynamics [18], which creating numerous obstacles to study how the brain processes information. The application of information thermodynamics in neuroscience may help overcome these obstacles, suggesting a potential direction to understand neural information processing at a physically fundamental level [21]. Based on previous studies, our research moves one more step further to explore the information thermodynamics nature of the brain. As a non-isolated system of neurons, the brain is demonstrated to act as an information thermodynamics encoder with intra-system coupling. By deriving the thermodynamics of encoding based on the nonequilibrium second law of thermodynamics, our theory helps reveal the physics foundation of how the brain encodes the information of an external source. This framework might contribute to understanding the intrinsic relations between cognition and the physical brain [21, 69]. Moreover, we combine information thermodynamics with the theory of information synergy and redundancy to implement a unified analysis. We demonstrate that the intra-system coupling within a system allows the encoded information in specific sub-systems to exceed the information thermodynamics bound on the whole system. In other words, more information of the external source can be encoded (or hidden) in certain sub-systems in comparison with the whole system. The emergence probability of this phenomenon is affected by the internal correlation strength inside the brain positively. We suggest that this finding may provide insights into the synergy and labor division among neurons and cortices during neural information processing. Classically, the decomposition of the information encoded by a system of elements is implemented utilizing information synergy and redundancy. The possible significance of synergy and redundancy in the neural system has been discussed for decades, discovering various effects of redundant or synergistic collective dynamics on neural information processing [83, 31, 13, 68]. Our research further connects these previous theoretical findings with the thermodynamics of encoding, revealing the thermodynamic foundation of information synergy and redundancy during the encoding process. This finding may offer a possible explanation for the stimulus-dependent superiority of specific neural or cortical tissues in encoding efficiency comparing with the whole brain. For instance, the superiority of the visual cortex during visual perception [34] not only means the relative redundancy of other sensory cortices but also corresponds to the larger encoded information quantity that is not restricted by the information thermodynamics bound of the whole brain. In sum, the theory depicted here features the potential to be further applied in neuroscience and related fields, which may help deepen our understanding of the physics foundation of neural information processing.

When we turn to a more general perspective about thermodynamics and information, one might find that limitations and possible insights coexist in our theory. Under the classical framework of information thermodynamics, our theory helps generalize the measurement process [69, 42, 77, 33, 80, 36, 28], one of the basic concepts that bridge information and thermodynamics, to the encoding process. By doing so, we can analyze the origin of mutual information

from the nonequilibrium second law of thermodynamics in a more general aspect. Built on this framework, a new perspective is suggested to study the information thermodynamics connection that originates from the coupling between a non-isolated system and its environment. This perspective helps reveal how the intra-system attributes of a non-isolated system shape the information thermodynamics coupling between this system and the environment. The information thermodynamics difference discovered between the non-isolated system with internal correlations and the non-isolated system of independent elements is independent of detailed system definitions and free to be applied to any real system. However, we need to emphasize that our current work remains at a qualitative level. Although the 3-order multiple mutual information enables us to analyze the internal orders of a non-isolated system in terms of information synergy and redundancy [64, 83], it fails to offer a more precise and quantitative characterization of these orders. This deficiency remains an open challenge for future studies. By overcoming this challenge, our suggested perspective may help explore more information thermodynamics characteristics determined by the nature of order inside the non-isolated system or the inter-system coupling relations. Last but not least, an idea that may interest both physicists and computer scientists relates to the characterized information thermodynamics encoder in our research, which functions as a kind of generalized computer or memory system [69]. Given the arbitrariness of non-isolated system selection for defining the encoding process, it is possible for any system with multiple distinguishable states to function as a generalized computer (see similar ideas in thermodynamic computing [43, 69, 90]).

To conclude, the information thermodynamics link that originates from system coupling is explored in our research. We pursue a thermodynamics theory of encoding and propose a novel idea to analyze an arbitrary non-isolated system as an information thermodynamics encoder. This perspective indicates the critical role of the nature of order inside the non-isolated system in shaping information thermodynamics coupling. The presented theoretical findings might lay a cornerstone for the future exploration of various related topics.

## 6 Acknowledgements

The authors are grateful to Prof. Oren Raz for his inspiring suggestion on the research and valuable comments on the manuscript. Y.T appreciates the support of the YutChun Program. Moreover, the authors appreciate Mr. Ben Hou and Mr. Tianyi Wang for their proofreading on mathematics.

This research is supported by the Artificial and General Intelligence Research Program of Guo Qiang Research Institute at Tsinghua University (2020GQG1017) as well as the Tsinghua University Initiative Scientific Research Program.

## A The relative entropy and the mutual information

Here we show that for each  $\tau$

$$\begin{aligned} & \mathcal{D}[\mathcal{P}(\mathbf{X}_{(\tau)}, \mathbf{Y}_{(\tau)}) \parallel \hat{\mathcal{P}}(\mathbf{X}_{(\tau)}) \hat{\mathcal{P}}(\mathbf{Y}_{(\tau)})] \\ &= \sum_{\Omega(\mathbf{X}) \times \Omega(\mathbf{Y})} \mathcal{P}(\mathbf{X}_{(\tau)}, \mathbf{Y}_{(\tau)}) \log \left( \frac{\mathcal{P}(\mathbf{X}_{(\tau)}, \mathbf{Y}_{(\tau)})}{\mathcal{P}(\mathbf{X}_{(\tau)}) \mathcal{P}(\mathbf{Y}_{(\tau)})} \right) + \\ & \quad \sum_{\Omega(\mathbf{X}) \times \Omega(\mathbf{Y})} \mathcal{P}(\mathbf{X}_{(\tau)}, \mathbf{Y}_{(\tau)}) \log \left( \frac{\mathcal{P}(\mathbf{X}_{(\tau)}) \mathcal{P}(\mathbf{Y}_{(\tau)})}{\hat{\mathcal{P}}(\mathbf{X}_{(\tau)}) \hat{\mathcal{P}}(\mathbf{Y}_{(\tau)})} \right) \end{aligned} \quad (23)$$

$$= \mathcal{I}(\mathbf{X}_{(\tau)}; \mathbf{Y}_{(\tau)}) + \sum_{\Omega(\mathbf{X}) \times \Omega(\mathbf{Y})} \mathcal{P}(\mathbf{X}_{(\tau)}, \mathbf{Y}_{(\tau)}) \log \left( \frac{\mathcal{P}(\mathbf{X}_{(\tau)}) \mathcal{P}(\mathbf{Y}_{(\tau)})}{\hat{\mathcal{P}}(\mathbf{X}_{(\tau)}) \hat{\mathcal{P}}(\mathbf{Y}_{(\tau)})} \right), \quad (24)$$

Then, one can reformulate

$$\begin{aligned} & \sum_{\Omega(\mathbf{X}) \times \Omega(\mathbf{Y})} \mathcal{P}(\mathbf{X}_{(\tau)}, \mathbf{Y}_{(\tau)}) \log \left( \frac{\mathcal{P}(\mathbf{X}_{(\tau)}) \mathcal{P}(\mathbf{Y}_{(\tau)})}{\hat{\mathcal{P}}(\mathbf{X}_{(\tau)}) \hat{\mathcal{P}}(\mathbf{Y}_{(\tau)})} \right) \\ &= - \sum_{\Omega(\mathbf{X}) \times \Omega(\mathbf{Y})} \mathcal{P}(\mathbf{X}_{(\tau)}, \mathbf{Y}_{(\tau)}) \log \left( \frac{\hat{\mathcal{P}}(\mathbf{X}_{(\tau)}) \hat{\mathcal{P}}(\mathbf{Y}_{(\tau)})}{\mathcal{P}(\mathbf{X}_{(\tau)}) \mathcal{P}(\mathbf{Y}_{(\tau)})} \right), \end{aligned} \quad (25)$$

and prove the following derivations

$$\begin{aligned}
& - \sum_{\Omega(X) \times \Omega(Y)} \mathcal{P}(X_{(\tau)}, Y_{(\tau)}) \log \left( \frac{\hat{\mathcal{P}}(X_{(\tau)}) \hat{\mathcal{P}}(Y_{(\tau)})}{\mathcal{P}(X_{(\tau)}) \mathcal{P}(Y_{(\tau)})} \right) \\
& = - \sum_{\Omega(X) \times \Omega(Y)} \mathcal{P}(X_{(\tau)}, Y_{(\tau)}) \log \left( \frac{\hat{\mathcal{P}}(X_{(\tau)})}{\mathcal{P}(X_{(\tau)}, Y_{(\tau)})} \right) - \sum_{\Omega(X) \times \Omega(Y)} \mathcal{P}(X_{(\tau)}, Y_{(\tau)}) \log \hat{\mathcal{P}}(Y_{(\tau)}) - \\
& \quad \sum_{\Omega(X) \times \Omega(Y)} \mathcal{P}(X_{(\tau)}, Y_{(\tau)}) \log \left( \frac{\mathcal{P}(X_{(\tau)}, Y_{(\tau)})}{\mathcal{P}(X_{(\tau)}) \mathcal{P}(Y_{(\tau)})} \right). \tag{26}
\end{aligned}$$

Combining (25) and (26), we can apply the Jensen inequality [22] to obtain

$$\begin{aligned}
& \sum_{\Omega(X) \times \Omega(Y)} \mathcal{P}(X_{(\tau)}, Y_{(\tau)}) \log \left( \frac{\mathcal{P}(X_{(\tau)}) \mathcal{P}(Y_{(\tau)})}{\hat{\mathcal{P}}(X_{(\tau)}) \hat{\mathcal{P}}(Y_{(\tau)})} \right) \\
& \geq - \log \sum_{\Omega(X) \times \Omega(Y)} \hat{\mathcal{P}}(X_{(\tau)}) - \sum_{\Omega(X) \times \Omega(Y)} \mathcal{P}(X_{(\tau)}, Y_{(\tau)}) \log \hat{\mathcal{P}}(Y_{(\tau)}) - \\
& \quad \sum_{\Omega(X) \times \Omega(Y)} \mathcal{P}(X_{(\tau)}, Y_{(\tau)}) \log \left( \frac{\mathcal{P}(X_{(\tau)}, Y_{(\tau)})}{\mathcal{P}(X_{(\tau)}) \mathcal{P}(Y_{(\tau)})} \right). \tag{27}
\end{aligned}$$

One can immediately find that the first term in (27) satisfies  $-\log \sum_{\Omega(X) \times \Omega(Y)} \hat{\mathcal{P}}(X_{(\tau)}) = 0$ . Therefore, we have

$$\begin{aligned}
& \sum_{\Omega(X) \times \Omega(Y)} \mathcal{P}(X_{(\tau)}, Y_{(\tau)}) \log \left( \frac{\mathcal{P}(X_{(\tau)}) \mathcal{P}(Y_{(\tau)})}{\hat{\mathcal{P}}(X_{(\tau)}) \hat{\mathcal{P}}(Y_{(\tau)})} \right) \\
& \geq - \sum_{\Omega(X) \times \Omega(Y)} \mathcal{P}(X_{(\tau)}, Y_{(\tau)}) \log \hat{\mathcal{P}}(Y_{(\tau)}) - \sum_{\Omega(X) \times \Omega(Y)} \mathcal{P}(X_{(\tau)}, Y_{(\tau)}) \log \left( \frac{\mathcal{P}(X_{(\tau)}, Y_{(\tau)})}{\mathcal{P}(X_{(\tau)}) \mathcal{P}(Y_{(\tau)})} \right) \tag{28}
\end{aligned}$$

$$\geq \zeta - \mathcal{I}(X_{(\tau)}; Y_{(\tau)}), \tag{29}$$

where we mark

$$\zeta = - \sum_{\Omega(X) \times \Omega(Y)} \mathcal{P}(X_{(\tau)}, Y_{(\tau)}) \log \hat{\mathcal{P}}(Y_{(\tau)}). \tag{30}$$

To analyze the value range of  $\zeta$ , we define a measure  $\mu[\Omega(X)] = m[\Omega(X)] m^{-1}[\Omega(X) \times \Omega(Y)]$ , where  $m(\cdot)$  denotes the area (the area quantities of discrete sets can be obtained by element counting). Based on this measure, one can see that  $\sum_{\Omega(X) \times \Omega(Y)} \hat{\mathcal{P}}(Y_{(\tau)}) = \mu[\Omega(X)]$ . Then, we can reformulate (30) as

$$\zeta = - \sum_{\Omega(X) \times \Omega(Y)} \mathcal{P}(X_{(\tau)}, Y_{(\tau)}) \log \left\{ \frac{\hat{\mathcal{P}}(Y_{(\tau)})}{\mu[\Omega(X)]} \mu[\Omega(X)] \right\} \tag{31}$$

$$= - \sum_{\Omega(X) \times \Omega(Y)} \mathcal{P}(X_{(\tau)}, Y_{(\tau)}) \log \frac{\hat{\mathcal{P}}(Y_{(\tau)})}{\mu[\Omega(X)]} - \sum_{\Omega(X) \times \Omega(Y)} \mathcal{P}(X_{(\tau)}, Y_{(\tau)}) \log \mu[\Omega(X)]. \tag{32}$$

It is trivial to know that  $\sum_{\Omega(X) \times \Omega(Y)} \frac{\hat{\mathcal{P}}(Y_{(\tau)})}{\mu[\Omega(X)]} = 1$ . Therefore, we can apply the Gibbs inequality [22] on the first term of (32) to prove that

$$\zeta \geq - \sum_{\Omega(X) \times \Omega(Y)} \mathcal{P}(X_{(\tau)}, Y_{(\tau)}) \log \mathcal{P}(X_{(\tau)}, Y_{(\tau)}) - \sum_{\Omega(X) \times \Omega(Y)} \mathcal{P}(X_{(\tau)}, Y_{(\tau)}) \log \mu[\Omega(X)] \tag{33}$$

$$\geq \mathcal{S}(X_{(\tau)}, Y_{(\tau)}) - \sum_{\Omega(X) \times \Omega(Y)} \mathcal{P}(X_{(\tau)}, Y_{(\tau)}) \log \mu[\Omega(X)]. \tag{34}$$

As for the second term of (34), its non-negativity is trivial because  $\mu[\Omega(X)] \leq 1$  (the equality is satisfied if and only if  $m[\Omega(X)] = m[\Omega(X) \times \Omega(Y)]$ ). Therefore, we can know

$$\zeta \geq \mathcal{S}(X_{(\tau)}, Y_{(\tau)}). \tag{35}$$

Based on (29) and (35), one can derive

$$\begin{aligned} & \sum_{\Omega(X) \times \Omega(Y)} \mathcal{P}(X_{(\tau)}, Y_{(\tau)}) \log \left( \frac{\mathcal{P}(X_{(\tau)}) \mathcal{P}(Y_{(\tau)})}{\hat{\mathcal{P}}(X_{(\tau)}) \hat{\mathcal{P}}(Y_{(\tau)})} \right) \\ & \geq \mathcal{S}(X_{(\tau)}, Y_{(\tau)}) - \mathcal{I}(X_{(\tau)}; Y_{(\tau)}) \quad (36) \\ & \geq 0. \quad (37) \end{aligned}$$

To conclude, we can ensure that the term in (24) is non-negative. This result further implies that  $\mathcal{D}[\mathcal{P}(X_{(t)}, Y_{(t)}) \parallel \hat{\mathcal{P}}(X_{(t)}) \hat{\mathcal{P}}(Y_{(t)})] \geq \mathcal{I}(X_{(t)}; Y_{(t)})$  (please see (8) in our main text).

## B The UCE Condition and the mutual information

Assume that the thermodynamics of encoding defined in (7-8) is independent of the **UCE Condition**, then  $\mathcal{D}[\mathcal{P}(X_{(\tau)}, Y_{(\tau)}) \parallel \hat{\mathcal{P}}(X_{(\tau)}, Y_{(\tau)})]$  is not necessarily equivalent to  $\mathcal{D}[\mathcal{P}(X_{(\tau)}, Y_{(\tau)}) \parallel \hat{\mathcal{P}}(X_{(\tau)}) \hat{\mathcal{P}}(Y_{(\tau)})]$ . The derivations of  $\mathcal{D}[\mathcal{P}(X_{(t)}, Y_{(t)}) \parallel \hat{\mathcal{P}}(X_{(t)}) \hat{\mathcal{P}}(Y_{(t)})] \geq \mathcal{I}(X_{(t)}; Y_{(t)})$  do not necessarily hold under this condition, remaining for further verification.

Following the idea in appendix B, we know that for each  $\tau$

$$\begin{aligned} & \mathcal{D}[\mathcal{P}(X_{(\tau)}, Y_{(\tau)}) \parallel \hat{\mathcal{P}}(X_{(\tau)}, Y_{(\tau)})] \\ & = \mathcal{I}(X_{(\tau)}; Y_{(\tau)}) - \sum_{\Omega(X) \times \Omega(Y)} \mathcal{P}(X_{(\tau)}, Y_{(\tau)}) \log \left( \frac{\hat{\mathcal{P}}(X_{(\tau)}, Y_{(\tau)})}{\mathcal{P}(X_{(\tau)}) \mathcal{P}(Y_{(\tau)})} \right), \quad (38) \end{aligned}$$

where the second term of (38) is equivalent to

$$\begin{aligned} & - \sum_{\Omega(X) \times \Omega(Y)} \mathcal{P}(X_{(\tau)}, Y_{(\tau)}) \log \left( \frac{\hat{\mathcal{P}}(X_{(\tau)}, Y_{(\tau)})}{\mathcal{P}(X_{(\tau)}) \mathcal{P}(Y_{(\tau)})} \right) \\ & = - \sum_{\Omega(X) \times \Omega(Y)} \mathcal{P}(X_{(\tau)}, Y_{(\tau)}) \log \left( \frac{\hat{\mathcal{P}}(X_{(\tau)}, Y_{(\tau)})}{\mathcal{P}(X_{(\tau)}, Y_{(\tau)})} \right) - \\ & \quad \sum_{\Omega(X) \times \Omega(Y)} \mathcal{P}(X_{(\tau)}, Y_{(\tau)}) \log \left( \frac{\mathcal{P}(X_{(\tau)}, Y_{(\tau)})}{\mathcal{P}(X_{(\tau)}) \mathcal{P}(Y_{(\tau)})} \right) \quad (39) \end{aligned}$$

$$= - \sum_{\Omega(X) \times \Omega(Y)} \mathcal{P}(X_{(\tau)}, Y_{(\tau)}) \log \left( \frac{\hat{\mathcal{P}}(X_{(\tau)}, Y_{(\tau)})}{\mathcal{P}(X_{(\tau)}, Y_{(\tau)})} \right) - \mathcal{I}(X_{(\tau)}; Y_{(\tau)}) \quad (40)$$

$$\geq - \log \sum_{\Omega(X) \times \Omega(Y)} \hat{\mathcal{P}}(X_{(\tau)}, Y_{(\tau)}) - \mathcal{I}(X_{(\tau)}; Y_{(\tau)}) \quad (41)$$

$$\geq 0 - \mathcal{I}(X_{(\tau)}; Y_{(\tau)}). \quad (42)$$

Here (41) can be derived utilizing the Jensen inequality [22]. Combining (38) and (42), what we can prove is

$$\mathcal{D}[\mathcal{P}(X_{(\tau)}, Y_{(\tau)}) \parallel \hat{\mathcal{P}}(X_{(\tau)}, Y_{(\tau)})] \geq 0 \quad (43)$$

rather than the inequality (8) in our theory. The inequality (43) is trivial since the non-negativity of the relative entropy has already been known. To derive the stronger conclusion in (8), the **UCE Condition** should be included in our theory.

## C Simulation setting for Ising model

The following are the parameters used in our simulation for system  $(\sigma, h)$ . Specifically, the coupling strength  $J$ , temperature  $T$ , and transition rate  $\theta^{-1}$  conditions are set as following. Please note that the parameter  $\theta$  measures the interval between two times of transitions of the external field. Its inverse  $\theta^{-1}$  quantifies the transition speed.

In our experiment, we simulate the Ising model under each combination  $(J, T, \theta^{-1})$  to obtain the results. Moreover, we can treat  $J$ ,  $T$ , and  $\theta^{-1}$  as three parameter directions, respectively. By averaging experiment results in arbitrary two parameter directions, we can further analyze these results as the functions of the remaining parameter direction.

Table 1: Parameter settings in simulation

Parameter	Value	Meaning
$t$	10000	Duration length
$ \sigma $	8	Number of spins
$m$	5	Maximum external field
$K_J$	10	Number of coupling strength conditions
$K_T$	10	Number of temperature conditions
$K_\theta$	10	Number of transition rate conditions
$\xi$	$10 \times 2^{ \sigma }$	Times of repetition

Table 2: Condition settings in simulation

Parameter	Value
$J$	$J \in \{0.1, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5\}$
$T$	$T \in \{0.1, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5\}$
$\theta^{-1}$	$\theta^{-1} \in \{1, \frac{1}{5}, \frac{1}{10}, \frac{1}{20}, \frac{1}{50}, \frac{1}{100}, \frac{1}{200}, \frac{1}{500}, \frac{1}{1000}, \frac{1}{2000}\}$

## D fMRI data acquisition and pre-processing

The fMRI data set used in the present study is acquired from [38], which has been used in several neuroscience studies [37, 35, 67]. Neural responses, reflected by the fMRI BOLD signals, are measured from 6 subjects (5 females and 1 male) with gradient echo echoplanar imaging on a GE 3T scanner (General Electric, Milwaukee, WI). Repetition time (TR) is 2500 ms, echo time (TE) is 30 ms, the flip angle is 90, and the field of view (FOV) is 240 mm. In each time of repetition, 40 slices of 3.5-mm-thick sagittal images are obtained. T1-weighted spoiled gradient recall (SPGR) signals (124 slices of 1.2-mm-thick sagittal images) are obtained as the high-resolution information of detailed anatomy (please note that our analysis is implemented on the BOLD signals rather than the T1 signals). Stimuli are gray-scale images of 8 types of visual objects (faces, houses, cats, bottles, scissors, shoes, chairs, and the nonsense patterns that are generated as the phase-scrambled images of intact objects). More details about the stimulus sequences in visual perception experiments can be seen in [38]. In our analysis, these stimuli are represented by the indices of their types.

Pre-processing of the BOLD signals is performed based on SPM12 (a MATLAB toolbox designed for brain imaging data analysis, one can see <https://www.fil.ion.ucl.ac.uk/spm/> for details), including slice-timing correction and motion correction. These corrections are standard workflows in BOLD signal pre-processing and have no significant influence on the observed phenomena in Sec. 4.2. As for the T1 signals, we have not implemented further pre-processing on them since the anatomy information is not necessary for our study.

In general, our analysis in Sec. 4.2 does not rely on the data set and pre-processing techniques critically. One can implement the same analysis on any neural data set involving the perception of multiple stimuli.

## References

- [1] Moshe Abeles. *Corticonics: Neural circuits of the cerebral cortex*. Cambridge University Press, 1991.
- [2] David Abreu and Udo Seifert. Extracting work from a single heat bath through feedback. *EPL (Europhysics Letters)*, 94(1):10001, 2011.
- [3] Muktish Acharyya. Nonequilibrium phase transition in the kinetic ising model: Dynamical symmetry breaking by randomly varying magnetic field. *Physical Review E*, 58(1):174, 1998.
- [4] David Andrieux and Pierre Gaspard. Nonequilibrium generation of information in copolymerization processes. *Proceedings of the National Academy of Sciences*, 105(28):9516–9521, 2008.
- [5] A Cardoso Barato and Udo Seifert. An autonomous and reversible maxwell’s demon. *EPL (Europhysics Letters)*, 101(6):60001, 2013.
- [6] Andre C Barato, David Hartich, and Udo Seifert. Efficiency of cellular information processing. *New Journal of Physics*, 16(10):103024, 2014.
- [7] Antoine Bérut, Artak Arakelyan, Artyom Petrosyan, Sergio Ciliberto, Raoul Dillenschneider, and Eric Lutz. Experimental verification of landauer’s principle linking information and thermodynamics. *Nature*, 483(7388):187–189, 2012.

- [8] Antoine Bérut, Artyom Petrosyan, and Sergio Ciliberto. Detailed Jarzynski equality applied to a logically irreversible procedure. *EPL (Europhysics Letters)*, 103(6):60002, 2013.
- [9] Gyan Bhanot. The metropolis algorithm. *Reports on Progress in Physics*, 51(3):429, 1988.
- [10] Alexander B Boyd and James P Crutchfield. Maxwell demon dynamics: Deterministic chaos, the Szilard map, and the intelligence of thermodynamic systems. *Physical review letters*, 116(19):190601, 2016.
- [11] Alexander B Boyd, Dibyendu Mandal, and James P Crutchfield. Correlation-powered information engines and the thermodynamics of self-correction. *Physical Review E*, 95(1):012152, 2017.
- [12] Kay Brandner and Udo Seifert. Periodic thermodynamics of open quantum systems. *Physical Review E*, 93(6):062134, 2016.
- [13] Naama Brenner, Steven P Strong, Roland Koberle, William Bialek, and Rob R de Ruyter van Steveninck. Synergy in a neural code. *Neural computation*, 12(7):1531–1552, 2000.
- [14] Herbert B Callen. Thermodynamics and an introduction to thermostatistics, 1998.
- [15] Michele Campisi, Peter Talkner, and Peter Hänggi. Thermodynamics and fluctuation theorems for a strongly coupled open quantum system: an exactly solvable case. *Journal of Physics A: Mathematical and Theoretical*, 42(39):392002, 2009.
- [16] Yuansheng Cao, Zongping Gong, and HT Quan. Thermodynamics of information processing based on enzyme kinetics: An exactly solvable model of an information pump. *Physical Review E*, 91(6):062117, 2015.
- [17] Adrian Chapman and Akimasa Miyake. How an autonomous quantum Maxwell demon can harness correlated information. *Physical Review E*, 92(6):062125, 2015.
- [18] Dante R Chialvo. Emergent complex neural dynamics. *Nature physics*, 6(10):744–750, 2010.
- [19] Mark M Churchland, M Yu Byron, John P Cunningham, Leo P Sugrue, Marlene R Cohen, Greg S Corrado, William T Newsome, Andrew M Clark, Paymon Hosseini, Benjamin B Scott, et al. Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nature neuroscience*, 13(3):369–378, 2010.
- [20] Barry A Cipra. An introduction to the Ising model. *The American Mathematical Monthly*, 94(10):937–959, 1987.
- [21] Guillem Collell and Jordi Fauquet. Brain activity and cognition: a connection from thermodynamics and information theory. *Frontiers in psychology*, 6:818, 2015.
- [22] Thomas M Cover. *Elements of information theory*. John Wiley & Sons, 1999.
- [23] Peter Dayan and Laurence F Abbott. *Theoretical neuroscience: computational and mathematical modeling of neural systems*. Computational Neuroscience Series, 2001.
- [24] Giovanni Diana, G Baris Bağcı, and Massimiliano Esposito. Finite-time erasing of information stored in fermionic bits. *Physical Review E*, 87(1):012111, 2013.
- [25] Matthew J Donald. Free energy and the relative entropy. *Journal of statistical physics*, 49(1):81–87, 1987.
- [26] Kurt Dopfer and Jason Potts. *The general theory of economic evolution*. Routledge, 2007.
- [27] Massimiliano Esposito, Katja Lindenberg, and Christian Van den Broeck. Entropy production as correlation between system and reservoir. *New Journal of Physics*, 12(1):013013, 2010.
- [28] Massimiliano Esposito and Christian Van den Broeck. Second law and Landauer principle far from equilibrium. *EPL (Europhysics Letters)*, 95(4):40004, 2011.
- [29] David Freedman and Persi Diaconis. On the histogram as a density estimator: L<sup>2</sup> theory. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 57(4):453–476, 1981.
- [30] Weihao Gao, Sreeram Kannan, Sewoong Oh, and Pramod Viswanath. Estimating mutual information for discrete-continuous mixtures. In *Advances in neural information processing systems*, pages 5986–5997, 2017.
- [31] Timothy J Gawne and Barry J Richmond. How independent are the messages carried by adjacent inferior temporal cortical neurons? *Journal of Neuroscience*, 13(7):2758–2771, 1993.
- [32] Maxim F Gelin and Michael Thoss. Thermodynamics of a subensemble of a canonical ensemble. *Physical Review E*, 79(5):051121, 2009.
- [33] Léo Granger and Holger Kantz. Thermodynamic cost of measurements. *Physical Review E*, 84(6):061110, 2011.
- [34] Kalanit Grill-Spector and Rafael Malach. The human visual cortex. *Annu. Rev. Neurosci.*, 27:649–677, 2004.

- [35] Stephen José Hanson, Toshihiko Matsuka, and James V Haxby. Combinatorial codes in ventral temporal lobe for object recognition: Haxby (2001) revisited: is there a “face” area? *Neuroimage*, 23(1):156–166, 2004.
- [36] H-H Hasegawa, J Ishikawa, K Takara, and DJ Driebe. Generalization of the second law for a nonequilibrium initial state. *Physics Letters A*, 374(8):1001–1004, 2010.
- [37] James V Haxby, M Ida Gobbini, Maura L Furey, Alumit Ishai, Jennifer L Schouten, and Pietro Pietrini. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293(5539):2425–2430, 2001.
- [38] J.V. Haxby, M.I. Gobbini, M.L. Furey, A. Ishai, J.L. Schouten, and P. Pietrini. “visual object recognition”, 2016.
- [39] Donald T Haynie. *Biological thermodynamics*. Cambridge University Press, 2001.
- [40] David J Heeger and David Ress. What does fmri tell us about neuronal activity? *Nature Reviews Neuroscience*, 3(2):142–151, 2002.
- [41] Johannes Hoppenau and Andreas Engel. On the energetics of information exchange. *EPL (Europhysics Letters)*, 105(5):50002, 2014.
- [42] Jordan M Horowitz, Takahiro Sagawa, and Juan MR Parrondo. Imitating chemical motors with optimal information motors. *Physical review letters*, 111(1):010602, 2013.
- [43] Todd Hylton, Thomas M Conte, and Mark D Hill. A vision to compute like nature: thermodynamically. *Communications of the ACM*, 64(6):35–38, 2021.
- [44] Sosuke Ito. Stochastic thermodynamic interpretation of information geometry. *Physical review letters*, 121(3):030605, 2018.
- [45] Sosuke Ito and Takahiro Sagawa. Information thermodynamics on causal networks. *Physical review letters*, 111(18):180603, 2013.
- [46] C Jarzynski. Microscopic analysis of clausius–duhem processes. *Journal of statistical physics*, 96(1):415–427, 1999.
- [47] Chris Jarzynski. Nonequilibrium work theorem for a system strongly coupled to a thermal environment. *Journal of Statistical Mechanics: Theory and Experiment*, 2004(09):P09005, 2004.
- [48] Christopher Jarzynski. The thermodynamics of writing a random polymer. *Proceedings of the National Academy of Sciences*, 105(28):9451–9452, 2008.
- [49] Christopher Jarzynski. Stochastic and macroscopic thermodynamics of strongly coupled systems. *Physical Review X*, 7(1):011008, 2017.
- [50] Sven Erik Jorgensen and Yuri M Svirezhev. *Towards a thermodynamic theory for ecological systems*. Elsevier, 2004.
- [51] Yonggun Jun, Momčilo Gavrilov, and John Bechhoefer. High-precision test of landauer’s principle in a feedback trap. *Physical review letters*, 113(19):190601, 2014.
- [52] A Katchalsky and O Kedem. Thermodynamics of flow processes in biological systems. *Biophysical journal*, 2(2 Pt 2):53, 1962.
- [53] Ryoichi Kawai, Juan MR Parrondo, and Christian Van den Broeck. Dissipation: The phase-space perspective. *Physical review letters*, 98(8):080602, 2007.
- [54] Alexander Kraskov, Harald Stögbauer, and Peter Grassberger. Estimating mutual information. *Physical review E*, 69(6):066138, 2004.
- [55] Ellen Kuhl and Paul Steinmann. Mass–and volume–specific views on thermodynamics for open systems. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 459(2038):2547–2568, 2003.
- [56] Rolf Landauer. The physical nature of information. *Physics letters A*, 217(4-5):188–193, 1996.
- [57] Rolf Landauer et al. Information is physical. *Physics Today*, 44(5):23–29, 1991.
- [58] Joel L Lebowitz and Anders Martin-Löf. On the uniqueness of the equilibrium state for ising spin systems. *Communications in Mathematical Physics*, 25(4):276–282, 1972.
- [59] Harvey S Leff and Andrew F Rex. *Maxwell’s demon: entropy, information, computing*. Princeton University Press, 2014.
- [60] Zhiyue Lu, Dibyendu Mandal, and Christopher Jarzynski. Engineering maxwell’s demon. *Physics Today*, 67(8):60–61, 2014.

- [61] Christopher W Lynn and Danielle S Bassett. The physics of brain network structure, function and control. *Nature Reviews Physics*, 1(5):318–332, 2019.
- [62] Dibyendu Mandal and Christopher Jarzynski. Work and information processing in a solvable model of maxwell’s demon. *Proceedings of the National Academy of Sciences*, 109(29):11641–11645, 2012.
- [63] Dibyendu Mandal, HT Quan, and Christopher Jarzynski. Maxwell’s refrigerator: an exactly solvable model. *Physical review letters*, 111(3):030602, 2013.
- [64] William McGill. Multivariate information transmission. *Transactions of the IRE Professional Group on Information Theory*, 4(4):93–111, 1954.
- [65] Neri Merhav. Sequence complexity and work extraction. *Journal of Statistical Mechanics: Theory and Experiment*, 2015(6):P06037, 2015.
- [66] Leonard Mlodinow and Todd A Brun. Relation between the psychological and thermodynamic arrows of time. *Physical Review E*, 89(5):052102, 2014.
- [67] Alice J O’toole, Fang Jiang, Hervé Abdi, and James V Haxby. Partially distributed representations of objects and faces in ventral temporal cortex. *Journal of cognitive neuroscience*, 17(4):580–590, 2005.
- [68] G Palm, AMHJ Aertsen, and GL Gerstein. On the significance of correlations among neuronal spike trains. *Biological cybernetics*, 59(1):1–11, 1988.
- [69] Juan MR Parrondo, Jordan M Horowitz, and Takahiro Sagawa. Thermodynamics of information. *Nature physics*, 11(2):131–139, 2015.
- [70] Thomas G Philbin and Janet Anders. Thermal energies of classical and quantum damped oscillators coupled to reservoirs. *Journal of Physics A: Mathematical and Theoretical*, 49(21):215303, 2016.
- [71] Lorenzo Pucci, Massimiliano Esposito, and Luca Peliti. Entropy production in quantum brownian motion. *Journal of Statistical Mechanics: Theory and Experiment*, 2013(04):P04005, 2013.
- [72] Hong Qian. Relative entropy: Free energy associated with equilibrium fluctuations and nonequilibrium deviations. *Physical Review E*, 63(4):042103, 2001.
- [73] Hong Qian and Timothy C Reluga. Nonequilibrium thermodynamics and nonlinear kinetics in a cellular signaling switch. *Physical review letters*, 94(2):028101, 2005.
- [74] David Ruelle. On the use of “small external fields” in the problem of symmetry breakdown in statistical mechanics. *Annals of Physics*, 69(2):364–374, 1972.
- [75] Matthias Ruth. *Integrating economics, ecology and thermodynamics*, volume 3. Springer Science & Business Media, 2013.
- [76] Takahiro Sagawa. Thermodynamic and logical reversibilities revisited. *Journal of Statistical Mechanics: Theory and Experiment*, 2014(3):P03025, 2014.
- [77] Takahiro Sagawa and Masahito Ueda. Minimal energy cost for thermodynamic information processing: measurement and information erasure. *Physical review letters*, 102(25):250602, 2009.
- [78] Takahiro Sagawa and Masahito Ueda. Generalized jarzynski equality under nonequilibrium feedback control. *Physical review letters*, 104(9):090602, 2010.
- [79] Takahiro Sagawa and Masahito Ueda. Fluctuation theorem with information exchange: Role of correlations in stochastic thermodynamics. *Physical review letters*, 109(18):180602, 2012.
- [80] Takahiro Sagawa and Masahito Ueda. Role of mutual information in entropy production under information exchanges. *New Journal of Physics*, 15(12):125012, 2013.
- [81] Pablo Sartori, Léo Granger, Chiu Fan Lee, and Jordan M Horowitz. Thermodynamic costs of information processing in sensory adaptation. *PLoS Comput Biol*, 10(12):e1003974, 2014.
- [82] R Keith Sawyer and Robert Keith Sawyer Sawyer. *Social emergence: Societies as complex systems*. Cambridge University Press, 2005.
- [83] Elad Schneidman, Susanne Still, Michael J Berry, William Bialek, et al. Network information and connected correlations. *Physical review letters*, 91(23):238701, 2003.
- [84] Mark J Schnitzer and Markus Meister. Multineuronal firing patterns in the signal from eye to brain. *Neuron*, 37(3):499–511, 2003.
- [85] RB Schwarz and AG Khachaturyan. Thermodynamics of open two-phase systems with coherent interfaces. *Physical review letters*, 74(13):2523, 1995.



- [86] Udo Seifert. Stochastic thermodynamics, fluctuation theorems and molecular machines. *Reports on progress in physics*, 75(12):126001, 2012.
- [87] Udo Seifert. First and second law of thermodynamics at strong coupling. *Physical review letters*, 116(2):020601, 2016.
- [88] Ken Sekimoto. *Stochastic energetics*, volume 799. Springer, 2010.
- [89] Biswa Sengupta, Martin B Stemmler, and Karl J Friston. Information and efficiency in the nervous system—a synthesis. *PLoS Comput Biol*, 9(7):e1003157, 2013.
- [90] Susanne Still, David A Sivak, Anthony J Bell, and Gavin E Crooks. Thermodynamics of prediction. *Physical review letters*, 109(12):120604, 2012.
- [91] Philipp Strasberg, Gernot Schaller, Tobias Brandes, and Massimiliano Esposito. Thermodynamics of a physical model implementing a maxwell demon. *Physical review letters*, 110(4):040601, 2013.
- [92] Sterling Street. Neurobiology as information physics. *Frontiers in systems neuroscience*, 10:90, 2016.
- [93] Sterling Street. Upper limit on the thermodynamic information content of an action potential. *Frontiers in Computational Neuroscience*, 14:37, 2020.
- [94] Yuri M Svirezhev. Thermodynamics and ecology. *Ecological Modelling*, 132(1-2):11–22, 2000.
- [95] Leo Szilard. Über die entropieverminderung in einem thermodynamischen system bei eingriffen intelligenter wesen. *Zeitschrift für Physik*, 53(11-12):840–856, 1929.
- [96] Peter Talkner and Peter Hänggi. Open system trajectories specify fluctuating work but not heat. *Physical Review E*, 94(2):022143, 2016.
- [97] Peter Talkner and Peter Hänggi. Colloquium: Statistical mechanics and thermodynamics at strong coupling: Quantum and classical. *Reviews of Modern Physics*, 92(4):041002, 2020.
- [98] HU Kuo Ting. On the amount of information. *Theory of Probability & Its Applications*, 7(4):439–447, 1962.
- [99] Shoichi Toyabe, Takahiro Sagawa, Masahito Ueda, Eiro Muneyuki, and Masaki Sano. Experimental demonstration of information-to-energy conversion and validation of the generalized jarzynski equality. *Nature physics*, 6(12):988–992, 2010.
- [100] Jaegon Um, Haye Hinrichsen, Chulan Kwon, and Hyunggyu Park. Total cost of operating an information engine. *New Journal of Physics*, 17(8):085001, 2015.
- [101] Satosi Watanabe. Information theoretical analysis of multivariate correlation. *IBM Journal of research and development*, 4(1):66–82, 1960.
- [102] Stuart Yarrow, Edward Challis, and Peggy Seriès. Fisher and shannon information in finite neural populations. *Neural computation*, 24(7):1740–1780, 2012.
- [103] Raymond W Yeung. A new outlook on shannon’s information measures. *IEEE transactions on information theory*, 37(3):466–474, 1991.
- [104] Yirui Zhang and Andre C Barato. Critical behavior of entropy production and learning rate: Ising model with an oscillating field. *Journal of Statistical Mechanics: Theory and Experiment*, 2016(11):113207, 2016.
- [105] Muhua Zheng, Antoine Allard, Patric Hagmann, Yasser Alemán-Gómez, and M Ángeles Serrano. Geometric renormalization unravels self-similarity of the multiscale human connectome. *Proceedings of the National Academy of Sciences*, 117(33):20244–20253, 2020.