System Information Decomposition

Aobo Lyu*[1,2], Bing Yuan[2], Ou Deng[3], Mingzhe Yang[4,2], Andrew Clark[1], and Jiang Zhang*[4,2]

[1]Department of Electrical and Systems Engineering, Washington University in St. Louis, St. Louis, Missouri, United States of America, 63130
[2]Swarma Research, Beijing, China, 102308
[3]Graduate School of Human Sciences, Waseda University, Tokorozawa city, Saitama, Japan, 359-1192
[4]School of Systems Science, Beijing Normal University, Beijing, China, 100875

Abstract

To characterize complex higher-order interactions among variables in a system, we introduce a new framework for decomposing the information entropy of variables in a system, termed System Information Decomposition (SID). Diverging from Partial Information Decomposition (PID) correlation methods, which quantify the interaction between a single target variable and a collection of source variables, SID extends those approaches by equally examining the interactions among all system variables. Specifically, we establish the robustness of the SID framework by proving all the information atoms are symmetric, which detaches the unique, redundant, and synergistic information from the specific target variable, empowering them to describe the interactions between variables. Additionally, we analyze the relationship between SID and existing information measures and propose several properties that SID quantitative methods should follow. Furthermore, by employing an illustrative example, we demonstrate that SID uncovers higher-order interaction relationships among variables that cannot be captured by current measures of probability and information and provide two approximate calculation methods verified by this case. This advance in higher-order measures enables SID to explain why Holism posits that some systems cannot be decomposed without losing characteristics under existing measures and offers a potential quantification framework for higher-order relationships across a broad spectrum of disciplines.

***K*eywords** Information decomposition · Information entropy · Complex systems · Multivariate system · System decomposition

1 Introduction

Systems Science is a multidisciplinary field investigating the relationships and interactions among internal variables within a system, with applications spanning neuroscience, biology, social sciences, engineering, and finance [1, 2]. Complex systems are defined by many interconnected variables that engage in intricate interactions, the understanding of which is critical for predicting emergent properties, devising novel treatments, and optimizing system performance.

In information theory, mutual information is a widely employed method for quantifying interactions between two variables by encapsulating shared information or reducing uncertainty facilitated by each variable [3]. However, mutual information is restricted to describing pairwise interactions, which often needs to be revised for analyzing complex systems that necessitate multivariate interaction assessments.

As a solution, Beer et al. introduced the Partial Information Decomposition (PID) method, which characterizes information interactions between a target variable and multiple source variables by decomposing the mutual information shared among them [4]. In the past ten years, PID and related theories, such as Information Flow Modes [5] and integrated information theory [6], have been applied in many fields, such as quantitative identification of Causal Emergence [7], dynamical process analysis [8] and information disclosure [9, 10]. However, PID-related techniques only decompose the partial information of a single target variable at a time. This leads to the fact that selecting or constructing a suitable and plausible target variable can be challenging or even unfeasible when addressing complex systems problems and also raises questions as to why certain variables are prioritized as targets over others. Moreover, this variable-specific perspective results in a unidirectional relationship between the specified target variable and source variable, which makes information atoms bound to a specific target variable and insufficient for a comprehensive description of the relationships among variables. This further limits our exploration of system functions and properties, as many originate from the relationship between system variables rather than specific variables or their asymmetric properties.

To overcome these limitations, we need a system analysis method based on a system perspective, analogous to the synchronization model [11] or the Ising model [12], rather than a variable perspective like PID. Furthermore, this method should capture the nature and characteristics of the system without specifying or introducing any special variable and consider all the interactive relationships among all variables in the system, including pairwise and higher-order relationships. Therefore, we propose System Information Decomposition (SID), an innovative method that treats all system variables equally and effectively captures their intricate interactions. This novel approach enhances our capacity to scrutinize and understand the complexities of multivariate systems.

Specifically, we expand the PID framework to a system perspective by proving the symmetry of information atoms; that is, the information atoms obtained by decomposing the variables' information entropy are independent of the choice of a target variable. Based on this, we put forward a general SID framework, wherein redundant, synergistic, and unique information atoms become a variable system's property reflecting complex (pairwise and higher-order) relationships among variables. Furthermore, we explore the connections between existing information entropy indicators and the information atom within the SID framework while outlining the necessary properties for information atom quantification and proposing several viable calculation approaches. Through a detailed case analysis, we provide an intuitive demonstration that SID can unveil higher-order relationships within the system that cannot be captured by existing probability or information measures and then propose two viable calculation approaches. Finally, we discuss SID's potential application scenarios and implications from the philosophical perspective of system decomposition and from areas such as higher-order networks and causal science.

Our contributions to Information and System Science are twofold. Firstly, with its calculation approaches, the SID framework broadens the application of information decomposition methods in complex systems by introducing a methodology to decompose all variables within a system. Secondly, this framework reveals previously unexplored higher-order relationship dimensions that cannot be represented by existing probability or information measures, providing a potential data-driven quantitative framework for Higher-order Networks related research.

The remainder of this paper is organized as follows. Section 2 reviews the development of information theory, PID, and related research. Section 3 extends the PID method to multivariate system scenarios, defines SID, shows the connections between existing information entropy indicators and the information atom, and details the information atom calculation properties. Section 4 intuitively presents the characteristics of the SID framework through case analysis and gives two calculation approaches. The significance and potential applications of SID are discussed in Section 5.

2 Information Decomposition

2.1 Information Theory Framework

Shannon's classical information theory has provided a robust foundation for understanding information entropy [3]. Mutual information and conditional entropy further decompose information and joint entropy according to the pairwise relationship between variables. This can be intuitively shown in Venn diagrams 1, a precise tool for depicting the information composition within systems. In this paper, we explore the potential of Venn diagrams to provide valuable insights into the complex decomposition of multivariate systems and extend the entropy decomposition methods of classical information theory.

2.2 Partial Information Decomposition Framework

In classical information theory, the joint mutual information may occasionally be larger or smaller than the sum of the mutual information between individual variables. Consequently, traditional redundant information calculations may yield negative values, contradicting our intuitive understanding. To address this phenomenon, Beer et al. proposed the Partial Information Decomposition (PID) framework [4].

The PID framework facilitates the decomposition of joint mutual information between multiple source variables and a target variable. Specifically, for a random target variable Y and a random source variables $X = X_1, X_2, \cdots, X_n$, the PID framework allows for the decomposition of the information that X provides about Y into information atoms.

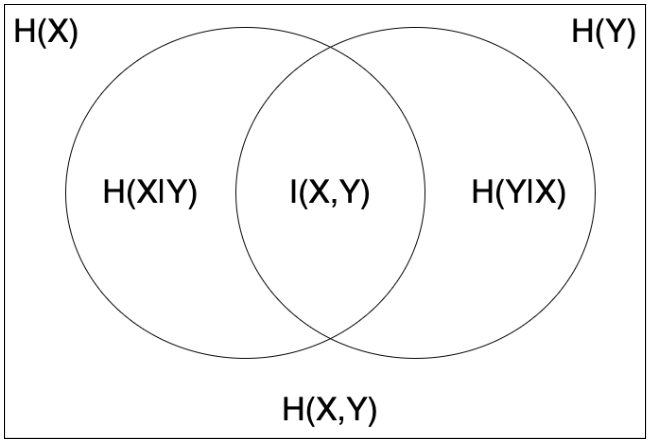


Figure 1: Information Theory Venn Diagram.

These atoms represent the partial information contributed by various subsets of X, individually or jointly, providing a more nuanced understanding of the relationships between the target and source variables.

Considering the simplest case of a system with three variables, one can employ a Venn diagram to elucidate their interactions [4]. The unique information $Un(Y : X_1)$ from X_1 signifies the information that X_1 provides to Y, which is not provided by X_2 and vice versa. In other words, unique information refers to the contribution made by a specific source variable to the target variable that is exclusive to that variable and not shared by other source variables. Redundant information $Red(Y : X_1, X_2)$ represents the common or overlapping information that X_1 and X_2 provide to Y. Synergistic information $Syn(Y : X_1, X_2)$ captures the combined contribution of X_1 and X_2 to Y, which cannot be obtained from either variable individually.

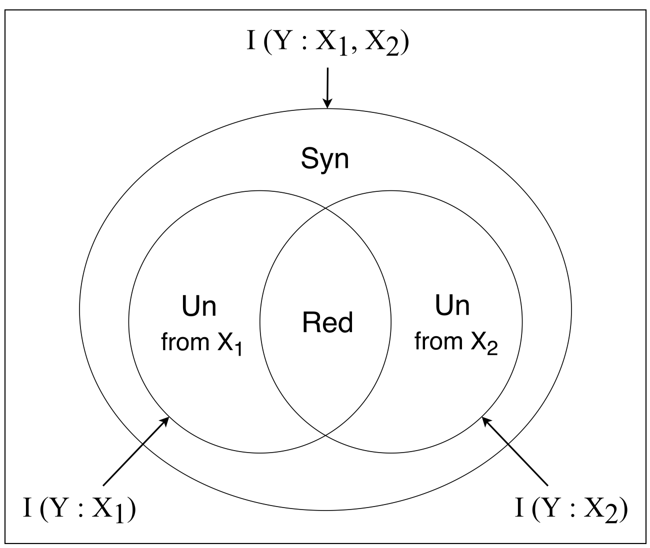


Figure 2: Venn Diagram of PID.

Definition 1 (Redundant Information)**.** *For an arbitrary variable system, we can select any variable as the target variable Y and the remaining variables as the source variables $X_1, \cdots, X_n$. The redundant information $Red(Y : X_1, \cdots, X_n)$ denotes the common or overlapping information provided by the source variables [4], which is contained in each source [13].*

Redundant information has the following properties [4]:

Axiom 1 (Symmetry of source variables)**.** *$Red(Y : X)$ is invariant to the permutation of X. For the source variables X_i and X_j from $\{X_1, \cdots, X_n\}$,$i, j \in \{1 \cdots n\}$, there is $Red(Y : X_i, \cdots X_j) = Red(Y : X_j, \cdots X_i)$.*

Axiom 2 (Self-redundancy)**.** *When there is only one source variable, the redundant information is equivalent to the mutual information between the target variable* Y *and the source variable* X_i*, i.e.,* $Red(Y : X_i) = I(Y : X_i)$*.*

Axiom 3 (Monotonicity)**.** *The redundancy should exhibit a monotonically decreasing behavior with the inclusion of additional inputs, i.e.* $Red(Y : X_1, \cdots, X_n) \leq Red(Y : X_1, \cdots, X_{n-1})$*, where* $n \in N$*.*

Despite numerous quantitative definitions for PID, a unified definition still needs to be discovered, primarily due to negative solutions. Such inconsistencies undermine the notion of information entropy as a non-negative measure of uncertainty. To circumvent reliance on a specific quantitative method, we employ classical mutual information and conditional entropy for calculating the sum of the information entropy of certain information atoms. Although this approach does not permit the precise calculation of individual information atoms [4, 14], it ensures that the framework remains independent of any specific PID calculation methods. Consequently, when a particular PID calculation method computes the value of one information atom, the information entropy of the remaining information atoms is determined.

Lemma 1 (Quantitative Computation)**.** *In a three-variable system with a target variable* Y *and source variables* X_i *and* X_j*, the following relationships hold:*

$Syn(Y : X_i, X_j) + Red(Y : X_i, X_j) + Un(Y : X_i) + Un(Y : X_j) = I(Y : X_i, X_j)$ *[4]*

$Un(Y : X_i) + Syn(Y : X_i, X_j) = I(Y : X_i|X_j) = H(Y|X_j) - H(Y|X_i, X_j)$ *[15]*

$Un(Y : X_i) + Red(Y : X_i, X_j) = I(X_i : Y)$ *[16]*

2.3 A set-theoretic understanding of PID

Although no perfect quantitative definition exists, several enlightening perspectives on PID have been proposed [4, 17, 18, 19, 20, 16]. In addition to those methods, a set theory may allow us to explore the properties of PID more deeply. Kolchinsky's work [13] offers an understanding based on set theory. Given that PID is inspired by an analogy between information theory and set theory, the redundant information can be understood as information sets that the sources provide to the target. More specifically, the definition of set intersection $\cap\{X_i\}$ in set theory means the most extensive set that is contained in all of the X_i. These set-theoretic definitions can be mapped into information-theoretic terms by treating "sets" as random variables, "set size" as entropy, and "set inclusion" as an ordering relation $\sqsubset$, which indicates when one random variable is more informative than another. One example of a partial order is Q $\sqsubset$ X if and only if $H(Q|X) = 0$.

Considering a set of sources variables $X_1, \cdots, X_n$ and a target Y, the PID aims to decompose $Red(Y : X_1, \cdots, X_n)$, the total same information provided by all sources about the target, into a set of non-negative terms. Therefore, redundant information can be viewed as the "intersection" of the information contributed by different sources, leading to the following definition:

Lemma 2 (Set Intersection of Information [13])**.** *For a variable system, the redundant information from the source variables* $X_1, \cdots, X_n$ *to the target variable* Y *is the information that all source variables can provide to the target variable, the largest mutual information between the target variable and information atom* Q *belonging to all source variables. That is*

$$Red(Y : X_1, \cdots, X_n) = I_\cap(X_1, \cdots, X_n \to Y) := \sup\{I(Q : Y) : Q \sqsubset X_i, \forall i \in \{1 \cdots n\}\}$$

3 System Information Decomposition

In this section, we develop a mathematical framework of SID. The objective of this framework is to decompose the information of all variables within a system based on their interrelationships. By addressing the limitation of PID, which focuses solely on a single target variable, we progress towards multi-variable information decomposition for systems.

3.1 Extension of PID in a System Scenario

The PID method only decomposes joint mutual information between multiple source variables and a specific target variable, as illustrated by the outermost circle of the Venn diagram in Figure 2. We redesign the Venn diagram to extend this method and encompass a system-wide perspective, as demonstrated in Figure 3. The system comprises two source variables,X_1 and X_2, and one target variable, Y, represented by the three intersecting circles.

The area size within the figure signifies the information entropy of the variables or information atoms, and the central area denotes the joint mutual information, encompassing redundant, unique from X_1, unique from X_2, and synergistic information. This arrangement aligns with the Venn diagram framework of PID.

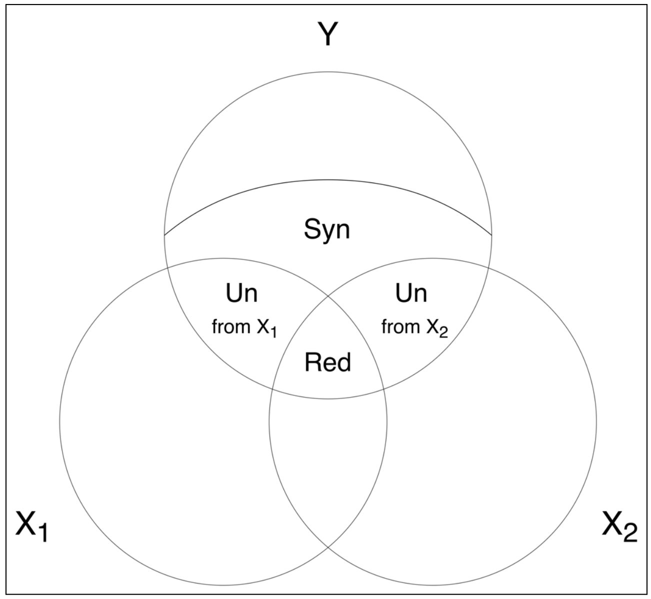


Figure 3: Venn diagram from different perspectives of PID.

To enhance the comprehensiveness of the framework, it is necessary to elucidate the unexplored section of the updated Venn diagram 3. In addition to the four sections of joint mutual information, the information entropy of the target variable Y contains an unaccounted-for area. According to Shannon's formula, this area corresponds to the joint conditional entropy of the source variables to the target variable $H(Y|X_1, X_2)$, which also characterizes the interrelationships between the target variable and the source variables. In the SID framework, numerous joint conditional entropy exists, including one that stands out: the joint conditional entropy originating from all variables except the target variable. To optimize the usefulness of the SID framework, we define this specific joint conditional entropy as the target variable's external information (Ext). The definition is grounded in the philosophical assumption that everything is interconnected. Since joint conditional entropy implies the uncertainty that the internal variables of the system cannot eliminate, the variables capable of providing this information must exist outside the system. External information can somewhat emphasize the relationship between the target variable and the entire system rather than just a simple relationship with other variables. Herefore, it is a kind of information atom within the SID framework.

Definition 2 (External Information)**.** *For a system containing variables Y and $\{X_1, \cdots, X_n\}$, the external information $Ext(Y)$ is defined as $Ext(Y) = H(Y|X_1, X_2, \cdots, X_n)$.*

Thus, we have decomposed the target variable's entropy into a finite number of non-repeated information atoms according to the relationship between it and the other variables in the system. Furthermore, we can apply this information decomposition method to each variable in the system to decompose the entire information entropy of the system, which results in a preliminary version of the SID. For the convenience of expression, we use Un_{i-j}, Syn_{ij-k}, and Red_{ij-k} to represent $Un(X_j, X_i)$, $Syn(X_k : X_i, X_j)$, and $Red(X_k : X_i, X_j)$ respectively. A Venn diagram for a three-variable system is shown in Figure 4:

3.2 Properties of information atoms

Although the preliminary version of SID can decompose all variables in a system, the decomposition of each variable is carried out separately, and the description of information atoms is directional (from source variables to the target variable). For instance, the unique information provided by X_1 to X_3 in Fig. 4 is not directly related to the unique information provided by X_3 to X_1. To make information atoms better reflect the relationship among variables, it is necessary to explore further the properties of information atoms within the SID framework. In this subsection, we will prove the symmetry property of information atoms by demonstrating that unique, redundant, and synergistic information atoms remain stable when different variables are considered target variables.

Theorem 1 (Symmetry of Redundant Information)**.** *Let $X_1, \cdots, X_n$ be the variables in a system. In SID, there is only one redundant information $Red(X_1, \cdots, X_n)$, which implies that the redundant information is equal irrespective of the chosen target variable. Formally, we write $Red(X_1, \cdots, X_n) = Red(X_i : X_1, \cdots, X_n \setminus X_i), \forall i \in \{1 \cdots n\}$.*

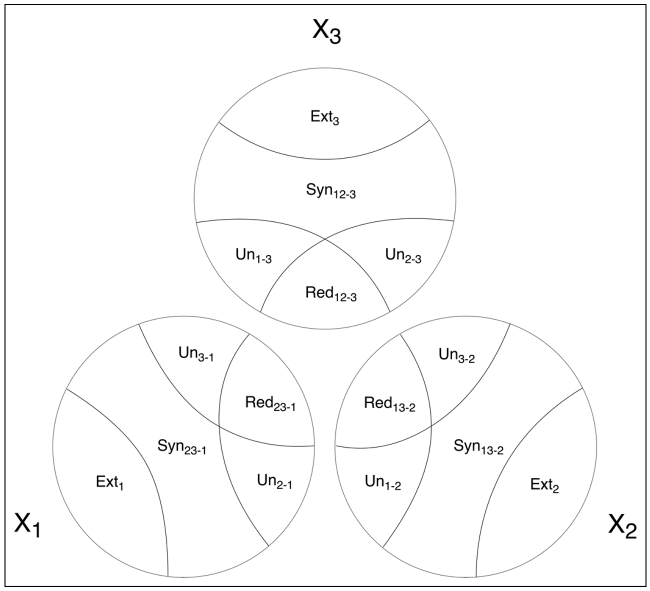


Figure 4: Venn diagram of SID's Preliminary version.

Proof. Suppose we have a variable system containing a target variable Y and source variables $X_1, \cdots, X_n$. For the convenience of expression, we use $\mathcal{X}$ to represent all the source variables $X_1, \cdots, X_n$. The proof is to show that $Red(Y : \mathcal{X}, Y) = Red(Y; \mathcal{X})$ and $Red(U : \mathcal{X}, Y) = Red(Y : \mathcal{X}, Y)$, where U is the union variable of Y and $\mathcal{X}$, such that $U = (\mathcal{X}, Y)$. Then, we can demonstrate that redundant information is equal regardless of which variable is chosen as the target variable.

Step One, to prove $Red(Y : \mathcal{X}, Y) = Red(Y; \mathcal{X})$: Firstly, we add a copy of the target variable Y' into the set of source variables $\{\mathcal{X}\}$ and obtain a new set of source variables $\{\mathcal{X}, Y'\}$. According to Axiom 3, we can deduce that $Red(Y : \mathcal{X}, Y') \leq Red(Y : \mathcal{X})$.

Secondly, by using the contradiction method, we assume $Red(Y : \mathcal{X}, Y') < Red(Y : \mathcal{X})$. Therefore, by Lemma 2, there exists an information atom Q_j such that $I(Q_j; Y) = Red(Y : \mathcal{X}) - Red(Y : \mathcal{X}, Y'), Q_j \sqsubset I_\cap(\mathcal{X} \to Y)$ and $(Q_j \not\sqsubset I_\cap(\mathcal{X}, Y' \to Y))$. Since $(I_\cap(\mathcal{X} \to Y) := \sup(I(Q; Y)$ such that $\forall i, Q \sqsubset X_i$, for all $i \in 1, \cdots, n, Q_j \sqsubset X_i$. And since $Q_j \not\sqsubset I_\cap(\mathcal{X}, Y' \to Y)$, $\exists Q_j \not\sqsubset Y'$. Therefore, $I(Q_j; Y') = 0$. Also, because $I(Q_j; Y) = Red(Y : \mathcal{X}) - Red(Y : \mathcal{X}, Y') > 0$, we get $Y \neq Y'$, which is contradiction, which means $Red(Y : \mathcal{X}, Y') \geq Red(Y : \mathcal{X})$.

Since we have $Red(Y : \mathcal{X}, Y') \leq Red(Y : \mathcal{X})$ and $Red(Y : \mathcal{X}, Y') \geq Red(Y : \mathcal{X})$, the result $Red(Y : \mathcal{X}, Y) = Red(Y : \mathcal{X}, Y') = Red(Y : \mathcal{X})$ is proved.

Step Two, to prove $Red(U : \mathcal{X}, Y) = Red(Y : \mathcal{X}, Y)$: Building upon the conclusion that $Red(Y : \mathcal{X}, Y) = Red(Y : \mathcal{X})$, we can replace the target variable with the union variable $U = (\mathcal{X}, Y)$, which combines the target variable Y and the source variables $\mathcal{X}$. (The entropy of the union variable U can be expressed as $H(U) = H(\mathcal{X}, Y)$.)

Firstly, let's employ the contradiction method by assuming that $Red(U : \mathcal{X}, Y) < Red(Y : \mathcal{X}, Y)$. Therefore, there exist a information atom Q_j, such that $I(Q_j; Y) = Red(Y : \mathcal{X}, Y) - Red(U : \mathcal{X}, Y), Q_j \sqsubset I_\cap(\mathcal{X}, Y \to Y)$ and $Q_j \not\sqsubset I_\cap(\mathcal{X}, Y \to U)$. This implies that there exists an information atom Q_k, satisfying $Q_k \sqsubset Y, Q_k \not\sqsubset U$ and $H(Q_k) > 0$, which creates a contradiction since it suggests $U \neq \cup(\mathcal{X}, Y)$. Consequently, we can conclude that $Red(U : \mathcal{X}, Y) \geq Red(Y : \mathcal{X}, Y)$.

Secondly, let's also use the contradiction method by assuming that $Red(U : \mathcal{X}, Y) > Red(Y : \mathcal{X}, Y)$. In this case, there exist P_j and P_k such that $I_\cap(P_j \to P_k) = I_\cap(\mathcal{X}, Y \to U) - I_\cap(\mathcal{X}, Y \to Y)$, with $P_k \sqsubset U$, $P_k \not\sqsubset Y$, and $P_j \sqsubset \cap(\mathcal{X}, Y)$, which implies that $P_j \sqsubset Y$. Considering Axiom 2, which states $I_\cap(Y \to Y) = Red(Y : Y) = I(Y : Y) = H(Y)$, we can deduce that for all $P_j \sqsubset Y$, if $I_\cap(P_j \to P_k) > 0$, then $P_k \sqsubset Y$, which leads to a contradiction. Therefore, we obtain $Red(U : \mathcal{X}, Y) \leq Red(Y : \mathcal{X}, Y)$.

Since we have both $Red(U : \mathcal{X}, Y) \geq Red(Y : \mathcal{X}, Y)$ and $Red(U : \mathcal{X}, Y) \leq Red(Y : \mathcal{X}, Y)$, $Red(U : \mathcal{X}, Y) = Red(Y : \mathcal{X}, Y)$ is proved.

In Summary: Since we have established that $Red(Y : \mathcal{X}, Y) = Red(Y : \mathcal{X})$, and $Red(U : \mathcal{X}, Y) = Red(Y : \mathcal{X}, Y)$, we can conclude that for all X_i in $\{\mathcal{X}\}$, $Red(X_i : Y, \{\mathcal{X}\} \setminus X_i) = Red(Y : \{\mathcal{X}\})$. Therefore, Theorem 1 is proved, and we can use $Red(X_1, \cdots, X_n)$ or $Red_{1\cdots n}$ denote the redundant information within the system $\{X_1, \cdots, X_n\}$.

□

Theorem 2 (Symmetry of Unique Information)**.** *Let $X_1, \cdots, X_n$ be the variables in a system. In SID, the unique information of any two variables relative to each other is equal, regardless of which is chosen as the target variable. Formally, we write $Un(X_i : X_j) = Un(X_j : X_i)$, $\forall i \neq j$ where $i, j \in \{1, \cdots, n\}$.*

Proof. According to Lemma 1, unique information is a part of the information provided by the source variable to the target variable, that is, mutual information minus redundant information. In a three-variable system $\{X_1, X_2, X_3\}$, we have $Un(X_i : X_j) + Red(X_i : X_j, X_k) = I(X_i; X_j)$, for all $i \neq j \in \{1, 2, 3\}$. Since $I(X_i : X_j) = I(X_j : X_i)$ according to the symmetry of Shannon's formula [3], and $Red(X_i : X_j, X_k) = Red(X_j : X_i, X_k) = Red(X_i, X_j, X_k)$ according to Theorem 1, we have $Un(X_i : X_j) = Un(X_j : X_i)$. Therefore, we can represent this information atom as $Un(X_i, X_j)$, or $Un_{i,j}$. □

Theorem 3 (Symmetry of Synergistic Information)**.** *Let $X_1, \cdots, X_n$ be the variables in a system. In SID, the synergistic information of any group of variables is equal, regardless of which is chosen as the target variable. Formally, we write $Syn(X_1, \cdots, X_n) = Syn(X_i : \{X_1, \cdots, X_n\} \setminus X_i), \forall i \in \{1 \cdots n\}$.*

Proof. According to Lemma 1, Theorem 1, Theorem 2, and the chain rule of Shannon formula, for a three-variable system with X_i, X_j, X_k:

$$\begin{aligned}
Syn(X_k : X_i, X_j) &= H(X_k|X_j) - H(X_k|X_i, X_j) - Un(X_i, X_k) \\
&= (H(X_j, X_k) - H(X_j)) - (H(X_i, X_j, X_k) - H(X_i, X_j)) - Un(X_i, X_k) \\
&= H(X_j, X_k) + H(X_i, X_j) - H(X_j) - H(X_i, X_j, X_k) - Un(X_i, X_k) \\
&= (H(X_i, X_j) - H(X_j)) - (H(X_i, X_j, X_k) - H(X_j, X_k)) - Un(X_i, X_k) \\
&= H(X_i|X_j) - H(X_i|X_j, X_k) - Un(X_i, X_k) \\
&= Syn(X_i : X_j, X_k)
\end{aligned}$$

Therefore, we proved Theorem 3 and we can write synergistic information in the form of $Syn(X_1, \cdots, X_n)$ or $Syn_{1\cdots n}$. □

Based on the Theorem 1 2 3 (the symmetry of information atoms), the SID framework can be merged into the formal version in Figure 5. In the formal version of SID, the concept of a target variable is canceled, and all variables are equally decomposed according to their relationship with other variables. Specifically, redundant information and unique information are merged. Redundant information (atoms) in any group of variables and unique information (atoms) between any two variables appear only once in the Venn diagram. In contrast, synergistic information (atoms) appear in each participating variable with the same value, and each variable contains one external information (atom). So far, we can give the formal definition of SID: SID is a system decomposition framework based on information entropy that can divide the whole information entropy of the system into non-overlapping information atoms according to the higher-order relationship between variables. In this framework, redundant information represents the common or overlapping information of all the variables; unique information represents information that is only owned by two variables but not by the third variable; and synergistic information represents the information that can be known from a variable only when the other two variables are observed simultaneously. The entropy of these information atoms reflects the properties of the set that contain different subsets of variables ($\{\{X_i, \cdots\}, \cdots, \{X_j, \cdots\}\}, i, j \in \{1 \cdots n\}$). The Venn diagrams and examples in this paper present only the simple case of a set with three subsets of variables, and each containing only one variable ($\{\{X_1\}, \{X_2\}, \{X_3\}\}$).

3.3 SID and Information Measure

In addition to Lemma 1 and Definition 2 for the relationship between SID and mutual information, conditional entropy, and joint conditional entropy, there are still some important information measures that deserve our attention.

Lemma 3 (Joint Entropy Decomposition)**.** *For any subsystem with 3 variables, $H(X_1, X_2, X_3) = Ext(X_1) + Ext(X_2) + Ext(X_3) + Un(X_1, X_2) + Un(X_1, X_3) + Un(X_2, X_3) + 2 * Syn(X_1, X_2, X_3) + Red(X_1, X_2, X_3)$.*

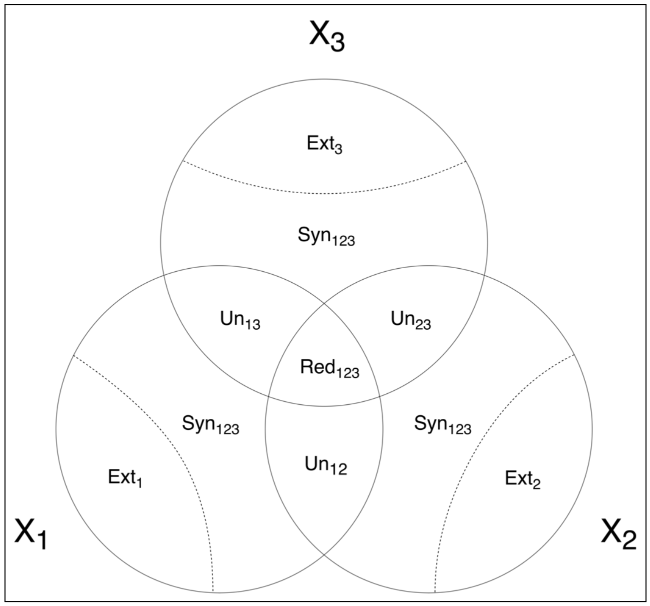


Figure 5: Venn diagram of SID's Formal Version.

Based on lemma 3, which can be easily proved by Lemma 1, we can have a deeper understanding of information atoms, that is, any information atom can be understood as some information stored by m variables, and at least n variables need to be known to obtain the information ($m > n, m, n \in \mathbb{Z}$). Specifically, the external information of the system is owned by the variable independently, so $m = 1$ and $n = 1$; redundant information is owned by all variables, so $m = number\ of\ variables$ and $n = 1$; unique information is owned by two variables, Therefore $m = 2$ and $n = 1$; synergistic information is shared by all variables, so $m = number\ of\ variables$ and $n = number\ of\ variables - 1$. Therefore, the joint entropy decomposition is the sum of each information atom multiplied by its $m - n$ quantity. This perspective will deepen our understanding of the essence of information atoms and facilitate our exploration of the joint entropy decomposition of systems with more than three variables.

Lemma 4 (Intersection Information Decomposition)**.** *For any system with 3 variables, its Intersection Information* $CoI(X_1, X_2, X_3) = Red(X_1, X_2, X_3) - Syn(X_1, X_2, X_3)$.

According to the calculation of $CoI(X, Y, Z) = H(X_1, X_2, X_3) + H(X_1) + H(X_2) + H(X_3) - H(X_1, X_2) - H(X_1, X_3) - H(X_2, X_3)$, Col is symmetry and unique for a system, which also verifies the symmetry of information atoms (Syn and Red) to some extent.

3.4 Calculation of SID

Although we have proposed the framework of SID and proved the symmetry of information atoms, the problem of exact computation still needs to be fully resolved. Therefore, in this paper, we alternatively propose the properties that the calculation method of the SID framework should satisfy and accept any method that can meet these properties.

Property 1 (Shannon's formula)**.** *The sum of certain information atoms should equal the mutual information and conditional information. It is Lemma 1 for a three-variable system.*

The information atoms can be regarded as a finer-grained division of Shannon's information entropy calculation, so calculation methods such as information entropy, mutual information, and conditional entropy can accurately calculate the sum of some information atoms, which means that the SID's calculation should conform to the Shannon formula. It is worth noting that when the specific PID calculation method calculates the value of one information atom, the rest of the information atoms will also get the results according to Lemma 1. This means that the calculation method of SID only needs to focus on one information atom in the system.

Property 2 (Computational Symmetry)**.** *The results of SID calculation should satisfy Theorems 1, 2, and 3.*

For the same system, the order of variables in the calculation method will not affect the results. This ensures that the SID framework provides a consistent decomposition of information entropy, regardless of the order of variables. Specifically, for redundant information and synergistic information, changing the order of any variable in the calculation method will not change the result; for unique information, exchanging the positions of the two focused variables or changing the order of the remaining variables will not change the result.

Property 3 (Non-negativity of information atoms)**.** *After applying SID, the value of any information atom is greater than or equal to zero. This non-negativity property holds because information measures and the degree of uncertainty are always non-negative as per the principles of information theory.*

Although the computational problem of information atoms has yet to be solved entirely, just like finding the Lyapunov function, for a specific case, we can often use specific methods, analysis, and some intuition to get the result. For example, a direct and rigorous method is to use properties 1 and 3.

Proposition 1 (Direct Method)**.** *Suppose certain mutual information or conditional entropy is zero. In that case, we can directly conclude that: (1) the redundant information and the corresponding unique information are zero if some mutual information is zero, or (2) the synergistic information and the corresponding unique information are zero if some conditional entropy is zero. Then, we can obtain the values of the remaining information atoms.*

For a more general scenario, in the next section, we will give a calculation formula that can be applied to most situations and a neural network method that can provide approximate values.

4 Measuring Higher-order Relationship

In this section, through a series of case analyses, we elucidate the unique properties of the SID framework and its capacity to uncover higher-order relationships that surpass the capabilities of current information and probability measures. We propose two novel methods for calculating information atoms and validate their accuracy and applicability by examining the cases.

4.1 Case Analysis

Without loss of generality, we can construct a case that includes both macro and micro perspectives, which can not only analyze the properties of SID at the macro level but also obtain ground truth through known micro properties. First, we construct six uniformly distributed Boolean variables a, b, c, d, e, f, ensuring that these variables are independent. We then create new variables by performing XOR operations on the existing variables: let $g = c \oplus e$, $h = d \oplus f$, $i = c \oplus f$, and $j = d \oplus e$.

Next, we form new macro variables by combining these micro variables: let $X_1 = abcd$, $X_2 = abef$, $X_3 = cdef$, $X_4 = aceh$, $X_5 = abgh$, $X_6 = abij$. The combination method involves simple splicing; e.g., when $a = 1$, $b = 0$, $c = 1$, $d = 1$, $X1$ is equal to 1011. Appendix A provides a concrete example that matches this design. As the micro-level variables are independent of each other, this combination ensures that the properties of the macro variables are a linear superposition of the properties of the micro variables.

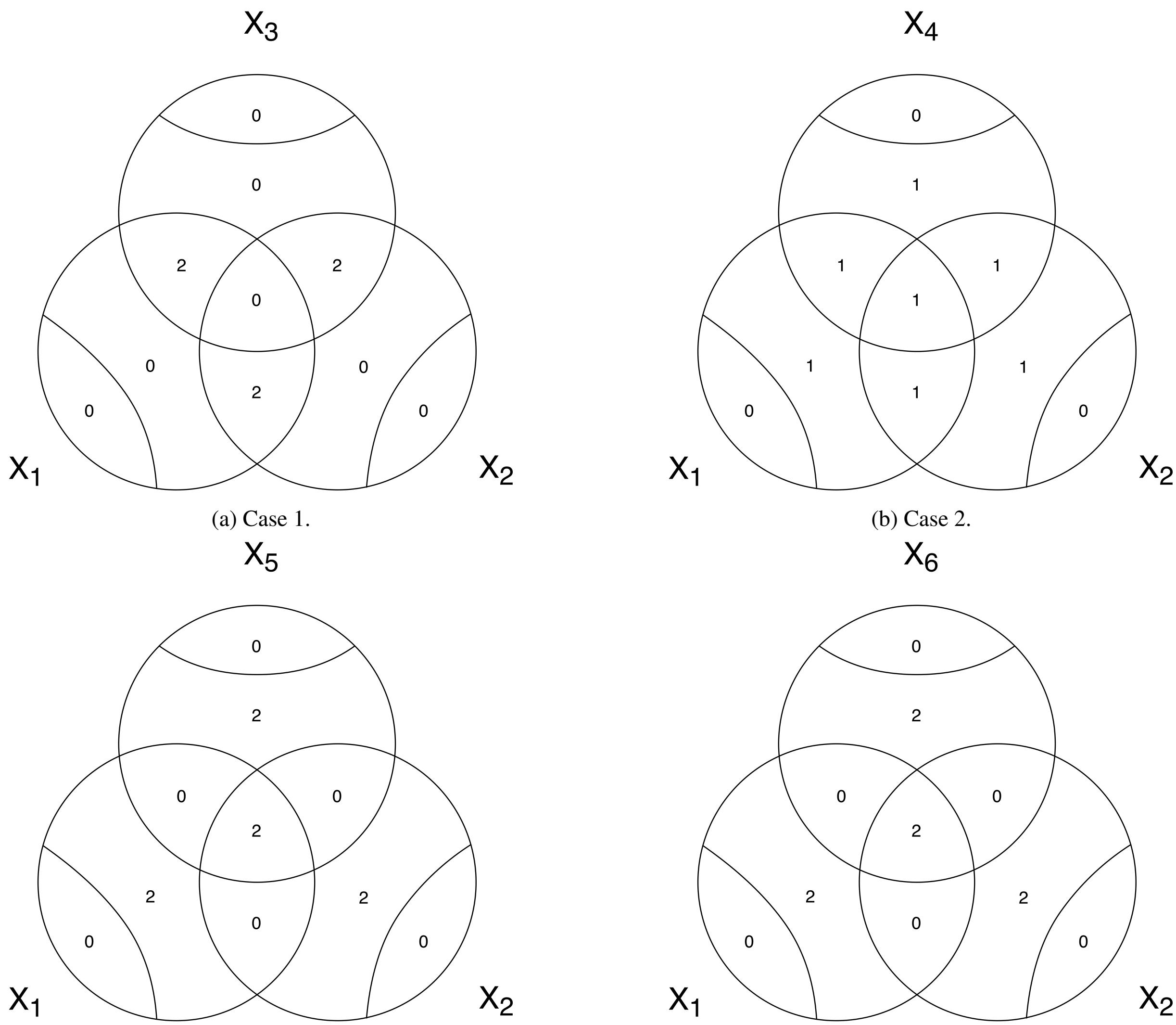
We then fix X_1 and X_2 as constants and form different three-variable systems (Cases 1-4) by adding X_3, X_4, X_5, and X_6, respectively. We analyze the differences between these three-variable systems.

It is worth noting that these four cases yield identical results under existing probability theory and information theory measures. The system has 64 equally probable outcomes, each variable has 16 equally probable outcomes, the total information amount in the system is 6, the pairwise mutual information between variables is 2, and the conditional entropy is 2. Existing systems analysis methods cannot identify the differences observed in these four examples.

However, the four systems exhibit three distinct internal characteristics under the SID framework. Since these examples comprise mutually independent micro variables, we can intuitively map the micro variables to the information atoms in each case. In Case 1, the micro variables a, b provide 2-bit unique information between X_1 and X_2 (c, d correspond to X_1 and X_3, e, f correspond to X_2 and X_3). In Case 2, micro variable a provides 1-bit redundant information, while b, c, and e provide 1-bit unique information between X_1 and X_2, X_1 and X_4, X_2 and X_4 respectively. The XOR relationship between $d - f - h$ provides 1-bit synergistic information between variables. In Cases 3 and 4, micro variables a and b provide 2-bit redundant information, and XOR relationships of $c - e - g$, $d - f - h$, and $c - f - i$, $d - e - j$ provide 2-bit synergistic information for the two cases, respectively. Figure 6 displays the SID Venn diagrams for Cases 1–4.

4.2 A Calculation Formula

Although we can calculate some cases through the Direct Method 1 or from the perspective of case construction like previous case analysis 4.1, to make the SID framework applicable in a broader range of scenarios, we need to find a



(a) Case 1. (b) Case 2.

(c) Case 3. (d) Case 4.

Figure 6: SID Venn Diagrams for Cases 1-4.

general solution for information atoms. After analyzing many known-result cases and combining some intuitions, we reveal the correspondence between the values of information atoms and certain structures on the data, which we call Synergistic Block and Unique Block. Based on this correspondence, we propose an identification method for unique information and synergistic information and further construct a formula for calculating synergistic information that is applicable in most cases.

Definition 3 (Synergistic Block and Unique Block)**.** *For a full probability table containing the values of all variables, if we fix a certain value of a variable (let $X_1 = x_1$), we can get the possible values ($j$and k) of the remaining variables under this condition ($j \in \{X_2|X_1 = x_1\}$, $k \in \{X_3|X_1 = x_1\}$). Then, mark all these values (j and k) of the remaining variables (X_2, X_3) while the fixed variables take other values ($X_2 = j|X_1 \neq x_1$, $X_3 = k|X_1 \neq x_1$). For all values of remaining variables where both occur simultaneously, such that $X_2 = j$ and $X_3 = k$ when $X_1 \neq x_1$, we call it Synergistic Block. For all values of remaining variables where only one occurs, we call it Unique Block, such that $X_2 = j$ and $X_3 \neq k$ when $X_1 \neq x_1$ for X_2, or $X_2 \neq j$ and $X_3 = k$ when $X_1 \neq x_1$ for X_3.*

Take Table A.1 as an example. We fixed the value of $X_1 = 0000$ and marked the values of all variables in this scenario in yellow. Then, we mark the values where X_2 to X_6 still take the same value when $X_1 \neq 0000$ as pink.Taking X_1, X_2 and X_4 as examples, we marked the synergistic blocks in **bold**, and marked the unique blocks of X2 and X3 in *italics*. Besides, although not as obvious as the previous two, redundant information also has corresponding redundant blocks.

Proposition 2 (Information Atom Identification)**.** *The synergistic information is greater than zero if and only if the synergistic block exists. For a three-variable system* $\{X_1, X_2, X_3\}$*,* $Syn(X_1, X_2, X_3) > 0$ *iff* $P(X_2 = j,\ X_3 = k,\ X_1 \neq x_1,\ j \in \{X_2|X_1 = x_1\},\ k \in \{X_3|X_1 = x_1\}) > 0$*. The unique information between two variables is greater than zero if and only if fix any of them. The remaining variable has a unique block for a three-variable system. That is* $Un(X_1, X_2) > 0$ *iff* $P(X_2 \neq j,\ X_3 = k,\ X_1 \neq x_1,\ j \in \{X_2|X_1 = x_1\},\ k \in \{X_3|X_1 = x_1\}) > 0$*.*

Based on the above Proposition, we construct a calculation formula to calculate synergistic information. The formula satisfies properties 1, 2, and 3 in most cases. The specific calculation method for synergistic information for a three-variable system involving X_1, X_2, and X_3 is as follows:

$$\begin{aligned}
&Syn(X_1, X_2, X_3) = (\sum P(x_1, x_2, x_3)* \\
&\log(\frac{P(X_2 = x_2, X_3 = k, k \in \{X_3|X_1 = x_1\})}{P(X_2 = x_2|X_1 = x_1)} * \frac{P(X_3 = x_3, X_2 = j, j \in \{X_2|X_1 = x_1\})}{P(X_3 = x_3|X_1 = x_1)} * \\
&\frac{P(X_1 = x_1)}{P(X_2 = j, X_3 = k, j \in \{X_2|X_1 = x_1\}, k \in \{X_3|X_1 = x_1\})})) - H(X_1|X_2, X_3)
\end{aligned} \tag{1}$$

In the previous case 4.1, since the data is relatively uniform, fixing any value of X_1 will have the same result, so we can quickly calculate the synergistic information of the four cases by fixing $X_1 = 0000$. In these cases, the $\log$ part of the formula can be intuitively understood as $\log$ (yellow + synergistic block / yellow), which is $\log(4/4) = 0$ in case 1; $\log(8/4) = 1$ in case 2; $\log(16/4) = 2$ in cases 3 and 4. Unique information can also be calculated by a similar method like $\log$(yellow + unique block / yellow).

4.3 An Approximate Method by Neural Information Squeezer

Another possible method is to use a generalized form of neural information squeezer (NIS, a machine learning framework using invertible neural networks proposed in Ref [21]) to calculate the redundancy of the system numerically and then to derive other information atoms.

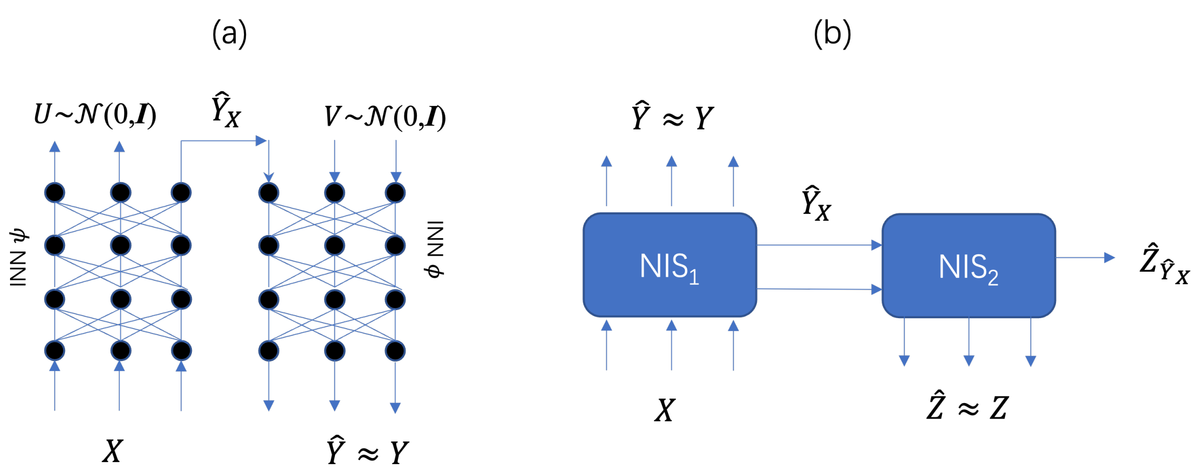


Figure 7: A generalized form of the Neural Information Squeezer network (NIS, see [21]) to calculate mutual information(a) and redundancy(b) of a trivariate system (X, Y, Z). In (a), two invertible neural networks (ψ, ϕ) can play the roles of encoder and decoder, respectively. The whole network accepts the input X to predict Y, and the intermediate variable $\hat{Y}_X$, which is the minimum low-dimensional representation of X, can be used to calculate the mutual information between X and Y. In (b), two NIS networks are stacked together. The first is just the network in (a), and the intermediate variable $\hat{Y}_X$ is fed into the second NIS network to predict Z. Then the intermediate variable, $\hat{Z}_{\hat{Y}_X}$ which is the minimum low-dimensional representation of $\hat{Y}_X$, can be used to calculate the redundancy of the system $\{X, Y, Z\}$.

As shown in Figure 7(a), the NIS framework has two parts: an encoder and a decoder. The encoder can accept any real vector variable with dimension p. It contains two operators: a bijector ψ modeled by an invertible neural network (see details in [21]) with dimension p and a projector χ which can drop out the last $p-q$ dimensions from the variable $\psi_p(X)$ to form variable U. The remaining part ($\hat{Y}_X$) can be regarded as a low-dimensional representation of X, which will be used to construct the target Y via another invertible neural network ϕ by mapping $[V, \hat{Y}_X]$ into $\hat{Y}$, where $V \sim \mathcal{N}(0, I)$ is a $p' - q$ dimensional random noise with Gaussian distribution, where p' is the dimension of Y. Then, we need to train the whole framework to conform that (1) $\hat{Y}$ approximates the target variable Y, and (2) U follows a $p-q$ dimensional standard normal distribution. It can be proven that the following proposition holds:

Proposition 3. *For any random variables X with p dimension and Y with p' dimension, and suppose p and p' are very large, then we can use the framework of Figure 7(a) to predict Y by squeezing the information channel of $\hat{Y}_X$ as the minimum dimension q^* but satisfying $\hat{Y} \approx Y$ and $U \sim \mathcal{N}(0, I)$. Further, if we suppose $H(X) > H(X|Y) > 0$, then:*

$$H(\hat{Y}_X) \approx I(X; Y), \tag{2}$$

and

$$H(U) \approx H(X|Y). \tag{3}$$

We will provide the proof in the appendix. We require that the dimensions of X, Y are significant because the maximal q for accurate predictions may not be an integer if p, p' are small. Therefore, we can enlarge the dimensions by duplicating the vectors.

We can use the NIS network twice to calculate the redundancy for a system with three variables: X, Y, Z, as shown in Figure 7(b). The first NIS network is to use the intermediate variable $\hat{Y}_X$, the dense low-dimensional representation of X with the minimum dimension q, to construct Y. Then, the second NIS network is to use $\hat{Z}_{\hat{Y}_X}$, the minimal dimensional dense low-dimensional representation of $\hat{Y}_X$ to construct Z. After these two steps, the Shannon entropy of the intermediate variable of NIS_2: $\hat{Z}_{\hat{Y}_X}$ can approach the redundancy. Thus, the redundancy of the system can be calculated approximately in the following way:

$$Red(X, Y, Z) \approx H(\hat{Z}_{\hat{Y}_X}). \tag{4}$$

To verify that $Red(X, Y, Z)$ calculated in this way can be regarded as the redundancy of the system, we need to prove that Equation 4 satisfies the property of symmetry for all the permutations of X, Y, Z, i.e., the following proposition:

Proposition 4. *For a system with three random variables X, Y, Z, without losing generality, we suppose that the conditional information satisfies $H(X) > H(X|Y) > 0$, $H(X) > H(X|Z) > 0$, and $H(Y) > H(Y|X) > 0$, then the redundancy calculated in Equation 4 is symmetric:*

$$Red(X, Y, Z) \approx Red(X, Z, Y). \tag{5}$$

It is noticed that $Red(X, Z, Y) \approx H(\hat{Y}_{\hat{Z}_X})$ is different from $Red(X, Y, Z)$ in the way that the order of the predictions from X is Z and then Y.

The proof of Theorem 4 is also provided in the appendix. By calculating redundancy, we can easily calculate unique and synergistic information atoms. Furthermore, we can extend the method to systems with more variables by stacking more NIS networks similarly as shown in Figure 7 (b).

However, there are two disadvantages to this method. One is that the calculation needs to be more accurate and requires many training epochs. Second, the number of dimensions of all variables must be large enough to discard the independent information among the variables by dropping out the dimensions. Further studies are needed.

To verify that the NIS framework can calculate redundant information, we conducted numerical experiments using Case3 as an example, as Figure8 shows, where the mutual information between each pair of variables and the redundant information is 2 bits.

In this experiment, variable X_1 is used as the input of NIS1 in the framework, with X_2 predicted as the target Y, and the intermediate variable $\hat{Y}_X$ is fed into NIS2 to predict X_3. This experiment expanded inputs and targets to 64 dimensions by directly replicating the original variables and letting the two intermediate variables in the NIS maintain consistent dimensions, denoted by q. The minimum dimension of $\hat{Y}_X$ and $\hat{Z}_{\hat{Y}_X}$ are selected by monitoring the changes in the loss curves.

From the above results, it can be seen that when q, the dimension of the intermediate variable, is relatively large, the entropy of the intermediate variable is entirely accurate for mutual information or redundant information. As the q drops below a threshold, the loss signally increases, indicating that the intermediate variable cannot capture all the mutual information and the redundant information.

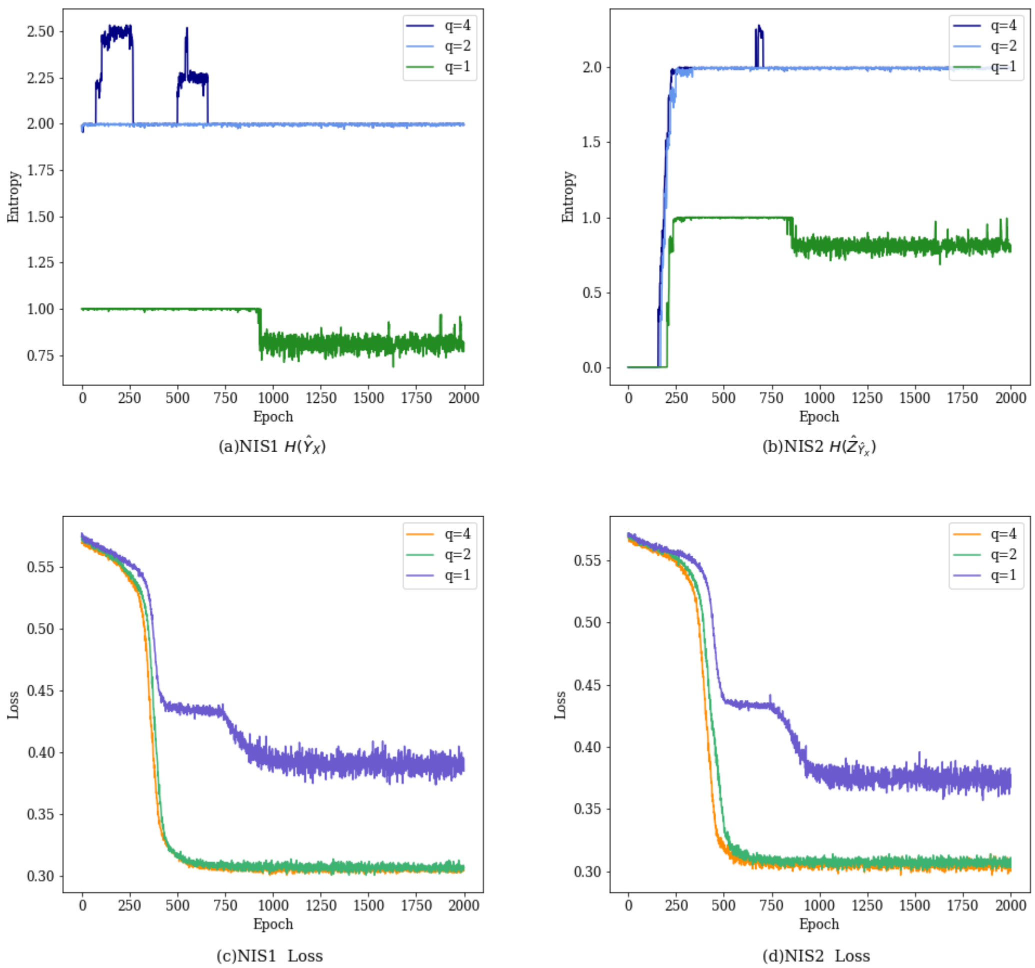


Figure 8: (a) The changes of $H(\hat{Y}_X)$ in NIS1 under $q = 4, 2, 1$ respectively; (b) The changes of $H(\hat{Z}_{\hat{Y}_X})$ in NIS2 under $q = 4, 2, 1$ respectively; (c) The changes of training loss in NIS1 under $q = 4, 2, 1$ respectively; (d) The changes of training loss in NIS2 under $q = 4, 2, 1$ respectively. The same experiments were conducted for the other three cases, and the redundant information could be accurately calculated under the NIS framework.

5 Discussion

The holism-versus-reductionism debate persists in modern literature [22]. Those with a reductionist view believe that any system can be divided into many subsystems. We can fully understand the entire system by studying the properties of the subsystems and their connections, which is also the research philosophy followed by most disciplines [23]. But holism holds that the system should be treated as a whole because the splitting of the system will inevitably lose the understanding of some of its properties [24]. This contradiction seems irreconcilable when we don't discuss how to split the system in detail.

However, the SID offers a perspective that can explain this conflict by accounting for higher-order relationships in the system that are not captured by previous measures. To better divide the different measures, we divide information entropy into first-order measures, which reflect a certain attribute of a single variable. On the other hand, mutual information and conditional entropy can be divided into second-order measures, which capture some aspects of pairwise relationships between variables [25]. Although among the second-order measurement, information theory's cross-entropy can measure the information shared among multiple variables. It still captures linear superpositions of second-order relationships, which provides limited insight into multivariate interactions. But under the SID framework, redundant, synergistic, and unique information can be regarded as three- or higher-order measures, revealing a new dimension of multivariate relationships entirely distinct from the first and second orders and facilitating a more profound comprehension of higher-order system relationships. In the case analysis, the internal structure of Case 1 aligns well with the results of the second-order measures and can be considered a reducible, decomposable system. Cases 2, 3, and 4, however, have internal structures that cannot be captured by second-order measures and are thus regarded by holism as systems that cannot be decomposed and understood individually. To some extent, SID and the case analysis offer an explanation that bridges the gap between holism and reductionism; some of the system properties that holism insists cannot be understood separately might be explained by higher-order measures or decomposition methods.

The understanding of SID and higher-order measures not only provides a philosophical perspective but also can be applied in many specific fields. A foreseeable application across many domains is that SID deepens our understanding of data, measures, and information. In the case analysis, the data in the table contains information about the construction of the four systems. Still, probability or existing information measures cannot capture this information. That means the incompleteness of measures may limit our ability to analyze existing data, even if we have obtained ideal data. Therefore, including higher-order information measures in the analysis of complex systems may offer valuable insights, especially in cases where traditional information theory measures fail to discern differences among systems. A direction worth exploring is extending SID to the field of higher-order network quantitative analysis [26]. It may potentially impact the analysis and understanding of real-world complex systems across various domains [27], since the framework offers a richer understanding of the relationships and interactions between system variables. For example, in studying neural networks and brain connectivity [28], the SID framework can provide further insights into the information flow between multiple neurons or brain regions; in ecological [29], financial or social systems, the quantitative characterization of high-order relationships among multiple variables can assist in the development of more accurate models and forecasts, as well as the design of effective control methods; in Machine Learning, the SID framework might be utilized in the development and analysis of machine learning models and algorithms. In summary, SID, as progress in the underlying measurement, may play a role in many application scenarios, which is also the focus of our next stage of work.

Another field where SID may impact is Causal Science since it studies the intrinsic relationships between multiple variables, just like the SID framework. Conditional independence plays an important role in causal discovery and causal inference in multivariate systems [30]. In the quantitative calculation of SID, conditional independence also plays a similar role in eliminating the uncertainty of higher-order relations. Refer to the calculation method first. Therefore, studying the properties of conditional independence within the framework of SID may provide a bridge between causal science and SID. The benefits of this association are mutual: from the perspective of Pearl Causal Hierarchy theory [31], SID is a research technique that utilizes observational data, which is at the lowest rung of the causal ladder. Investigating whether lifting the approach to higher rungs of the causal ladder can yield more profound insights into the system is an area worth exploring, for instance, by incorporating causal graphs (DAGs) into SID methods, etc.

In addition to the above-mentioned promising progress and expectations, several limitations are still worthy of attention. The first limitation is the need for a fully compatible quantitative method for the proposed framework, which restricts the practical application of SID in addressing real-world problems. As we continue to develop and refine the SID framework, it is a priority to develop robust computational methods for calculating SID components and consider how higher-order information measures can be integrated into existing analytical approaches. Furthermore, the existing proofs of framework properties and computational methods have only been established for three-variable systems. Although extending current work to more-than-three-variable systems is not a formidable challenge, it contains many aspects of work, such as how to intuitively present the decomposition results of multivariate systems on a two-dimensional plane, how to optimize the calculation algorithm to avoid the exponential calculation cost as the number of variables increases;

etc., which will be considered in the next stage of research. For those mentioned above and any possible problems, we cordially invite other scholars interested in this field to collaborate on addressing the existing challenges of SID and contribute to the model's refinement.

6 Conclusion

In this study, we introduced the System Information Decomposition (SID) framework, which offers novel insights for decomposing complex systems and analyzing higher-order relationships while addressing the limitations of existing information decomposition methods. By proving the symmetries of information atoms and connecting them to higher-order relationships, we show that the SID framework can provide insights and advance beyond existing measures in understanding the internal interactions and dynamics of complex systems. Furthermore, we explored the far-reaching implications of SID's unveiling of higher-order measures on the philosophical aspects of systems research, higher-order networks, and causal science. Even though current research on SID still faces challenges in quantitative calculations and multivariate analysis, continued collaboration and exploration by the scientific community will help overcome these obstacles. In conclusion, the SID framework signifies a promising new direction for investigating complex systems and information decomposition. We anticipate that the SID analysis framework will serve as a valuable tool across an expanding array of fields in the future.

Acknowledgments

We sincerely thank all the non-authors who played a crucial role in its successful completion. Our heartfelt appreciation goes to the Swarma Club, an open academic community for Complex Systems, where the Causal Emergence reading club provided the foundation for the ideas presented in this paper. We are also very grateful to Professor Duguid at UC Berkeley, whose course steadied an author's orientation towards understanding systems from an information perspective, serving as the genesis of this paper.

References

[1] Brian Castellani and Frederic William Hafferty. *Sociology and complexity science: a new field of inquiry*. Springer Science & Business Media, 2009.

[2] Cristoforo Sergio Bertuglia and Franco Vaio. *Nonlinearity, chaos, and complexity: the dynamics of natural and social systems*. Oxford University Press on Demand, 2005.

[3] Claude Elwood Shannon. A mathematical theory of communication. *ACM SIGMOBILE mobile computing and communications review*, 5(1):3–55, 2001.

[4] Paul L Williams and Randall D Beer. Nonnegative decomposition of multivariate information. *arXiv preprint arXiv:1004.2515*, 2010.

[5] Ryan G James, Blanca Daniella Mansante Ayala, Bahti Zakirov, and James P Crutchfield. Modes of information flow. *arXiv preprint arXiv:1808.06723*, 2018.

[6] Pedro AM Mediano, Fernando Rosas, Robin L Carhart-Harris, Anil K Seth, and Adam B Barrett. Beyond integrated information: A taxonomy of information dynamics phenomena. *arXiv preprint arXiv:1909.02297*, 2019.

[7] Fernando E Rosas, Pedro AM Mediano, Henrik J Jensen, Anil K Seth, Adam B Barrett, Robin L Carhart-Harris, and Daniel Bor. Reconciling emergences: An information-theoretic approach to identify causal emergence in multivariate data. *PLoS computational biology*, 16(12):e1008289, 2020.

[8] Ryan G James, Nix Barnett, and James P Crutchfield. Information flows? a critique of transfer entropies. *Physical review letters*, 116(23):238701, 2016.

[9] Fernando E Rosas, Pedro AM Mediano, Borzoo Rassouli, and Adam B Barrett. An operational information decomposition via synergistic disclosure. *Journal of Physics A: Mathematical and Theoretical*, 53(48):485001, 2020.

[10] Borzoo Rassouli, Fernando E Rosas, and Deniz Gündüz. Data disclosure under perfect sample privacy. *IEEE Transactions on Information Forensics and Security*, 15:2012–2025, 2019.

[11] Juan A Acebrón, Luis L Bonilla, Conrad J Pérez Vicente, Félix Ritort, and Renato Spigler. The kuramoto model: A simple paradigm for synchronization phenomena. *Reviews of modern physics*, 77(1):137, 2005.

[12] Barry A Cipra. An introduction to the ising model. *The American Mathematical Monthly*, 94(10):937–959, 1987.

[13] Artemy Kolchinsky. A novel approach to the partial information decomposition. *Entropy*, 24(3):403, 2022.

[14] Paul L Williams. *Information dynamics: Its theory and application to embodied cognitive systems*. PhD thesis, Indiana University, 2011.

[15] Xueyan Niu and Christopher J Quinn. A measure of synergy, redundancy, and unique information using information geometry. In *2019 IEEE International Symposium on Information Theory (ISIT)*, pages 3127–3131. IEEE, 2019.

[16] Nils Bertschinger, Johannes Rauh, Eckehard Olbrich, Jürgen Jost, and Nihat Ay. Quantifying unique information. *Entropy*, 16(4):2161–2183, 2014.

[17] Virgil Griffith, Edwin KP Chong, Ryan G James, Christopher J Ellison, and James P Crutchfield. Intersection information based on common randomness. *Entropy*, 16(4):1985–2000, 2014.

[18] Robin AA Ince. Measuring multivariate redundant information with pointwise common change in surprisal. *Entropy*, 19(7):318, 2017.

[19] Nils Bertschinger, Johannes Rauh, Eckehard Olbrich, and Jürgen Jost. Shared information—new insights and problems in decomposing information in complex systems. In *Proceedings of the European conference on complex systems 2012*, pages 251–269. Springer, 2013.

[20] Malte Harder, Christoph Salge, and Daniel Polani. Bivariate measure of redundant information. *Physical Review E*, 87(1):012130, 2013.

[21] Jiang Zhang and Kaiwei Liu. Neural information squeezer for causal emergence. *Entropy*, 25(1):26, 2023.

[22] Melanie Mitchell. *Complexity: A guided tour*. Oxford university press, 2009.

[23] Richard Gallagher and Tim Appenzeller. Beyond reductionism, 1999.

[24] Jan Christiaan Smuts. *Holism and evolution*. Macmillan, 1926.

[25] Thomas M Cover. *Elements of information theory*. John Wiley & Sons, 1999.

[26] Ginestra Bianconi. *Higher-order networks*. Cambridge University Press, 2021.

[27] Federico Battiston, Enrico Amico, Alain Barrat, Ginestra Bianconi, Guilherme Ferraz de Arruda, Benedetta Franceschiello, Iacopo Iacopini, Sonia Kéfi, Vito Latora, Yamir Moreno, et al. The physics of higher-order interactions in complex systems. *Nature Physics*, 17(10):1093–1098, 2021.

[28] Ed Bullmore and Olaf Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature reviews neuroscience*, 10(3):186–198, 2009.

[29] Simon A Levin. Self-organization and the emergence of complexity in ecological systems. *Bioscience*, 55(12):1075–1079, 2005.

[30] Judea Pearl. *Causality*. Cambridge university press, 2009.

[31] Judea Pearl and Dana Mackenzie. *The book of why: the new science of cause and effect*. Basic books, 2018.

A Appendix

A.1 Case Table

X_1				X_2				X_3				X_4				X_5				X_6			
a	b	c	d	a	b	e	f	c	d	e	f	a	c	e	h	a	b	g	h	a	b	i	j
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	1	0
0	0	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	0	1
0	0	0	0	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
0	0	0	1	**0**	**0**	**0**	**0**	0	1	0	0	**0**	**0**	**0**	**1**	0	0	0	1	0	0	0	1
0	0	0	1	**0**	**0**	**0**	**1**	0	1	0	1	**0**	**0**	**0**	**0**	0	0	0	0	0	0	1	1
0	0	0	1	**0**	**0**	**1**	**0**	0	1	1	0	**0**	**0**	**1**	**1**	0	0	1	1	0	0	0	0
0	0	0	1	**0**	**0**	**1**	**1**	0	1	1	1	**0**	**0**	**1**	**0**	0	0	1	0	0	0	1	0
0	0	1	0	*0*	*0*	*0*	*0*	1	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0
0	0	1	0	*0*	*0*	*0*	*1*	1	0	0	1	0	1	0	1	0	0	1	1	0	0	0	0
0	0	1	0	*0*	*0*	*1*	*0*	1	0	1	0	0	1	1	0	0	0	0	0	0	0	1	1
0	0	1	0	*0*	*0*	*1*	*1*	1	0	1	1	0	1	1	1	0	0	0	1	0	0	0	1
0	0	1	1	*0*	*0*	*0*	*0*	1	1	0	0	0	1	0	1	0	0	1	1	0	0	1	1
0	0	1	1	*0*	*0*	*0*	*1*	1	1	0	1	0	1	0	0	0	0	1	0	0	0	0	1
0	0	1	1	*0*	*0*	*1*	*0*	1	1	1	0	0	1	1	1	0	0	0	1	0	0	1	0
0	0	1	1	*0*	*0*	*1*	*1*	1	1	1	1	0	1	1	0	0	0	0	0	0	0	0	0
0	1	0	0	0	1	0	0	0	0	0	0	*0*	*0*	*0*	*0*	0	1	0	0	0	1	0	0
0	1	0	0	0	1	0	1	0	0	0	1	*0*	*0*	*0*	*1*	0	1	0	1	0	1	1	0
0	1	0	0	0	1	1	0	0	0	1	0	*0*	*0*	*1*	*0*	0	1	1	0	0	1	0	1
0	1	0	0	0	1	1	1	0	0	1	1	*0*	*0*	*1*	*1*	0	1	1	1	0	1	1	1
0	1	0	1	0	1	0	0	0	1	0	0	*0*	*0*	*0*	*1*	0	1	0	1	0	1	0	1
0	1	0	1	0	1	0	1	0	1	0	1	*0*	*0*	*0*	*0*	0	1	0	0	0	1	1	1
0	1	0	1	0	1	1	0	0	1	1	0	*0*	*0*	*1*	*1*	0	1	1	1	0	1	0	0
0	1	0	1	0	1	1	1	0	1	1	1	*0*	*0*	*1*	*0*	0	1	1	0	0	1	1	0
0	1	1	0	0	1	0	0	1	0	0	0	0	1	0	0	0	1	1	0	0	1	1	0
0	1	1	0	0	1	0	1	1	0	0	1	0	1	0	1	0	1	1	1	0	1	0	0
0	1	1	0	0	1	1	0	1	0	1	0	0	1	1	0	0	1	0	0	0	1	1	1
0	1	1	0	0	1	1	1	1	0	1	1	0	1	1	1	0	1	0	1	0	1	0	1
0	1	1	1	0	1	0	0	1	1	0	0	0	1	0	1	0	1	1	1	0	1	1	1
0	1	1	1	0	1	0	1	1	1	0	1	0	1	0	0	0	1	1	0	0	1	0	1
0	1	1	1	0	1	1	0	1	1	1	0	0	1	1	1	0	1	0	1	0	1	1	0
0	1	1	1	0	1	1	1	1	1	1	1	0	1	1	0	0	1	0	0	0	1	0	0
1	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0
1	0	0	0	1	0	0	1	0	0	0	1	1	0	0	1	1	0	0	1	1	0	1	0
1	0	0	0	1	0	1	0	0	0	1	0	1	0	1	0	1	0	1	0	1	0	0	1
1	0	0	0	1	0	1	1	0	0	1	1	1	0	1	1	1	0	1	1	1	0	1	1
1	0	0	1	1	0	0	0	0	1	0	0	1	0	0	1	1	0	0	1	1	0	0	1
1	0	0	1	1	0	0	1	0	1	0	1	1	0	0	0	1	0	0	0	1	0	1	1
1	0	0	1	1	0	1	0	0	1	1	0	1	0	1	1	1	0	1	1	1	0	0	0
1	0	0	1	1	0	1	1	0	1	1	1	1	0	1	0	1	0	1	0	1	0	1	0
1	0	1	0	1	0	0	0	1	0	0	0	1	1	0	0	1	0	1	0	1	0	1	0
1	0	1	0	1	0	0	1	1	0	0	1	1	1	0	1	1	0	1	1	1	0	0	0
1	0	1	0	1	0	1	0	1	0	1	0	1	1	1	0	1	0	0	0	1	0	1	1
1	0	1	0	1	0	1	1	1	0	1	1	1	1	1	1	1	0	0	1	1	0	0	1
1	0	1	1	1	0	0	0	1	1	0	0	1	1	0	1	1	0	1	1	1	0	1	1
1	0	1	1	1	0	0	1	1	1	0	1	1	1	0	0	1	0	1	0	1	0	0	1
1	0	1	1	1	0	1	0	1	1	1	0	1	1	1	1	1	0	0	1	1	0	1	0
1	0	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	0	0	0	1	0	0	0
1	1	0	0	1	1	0	0	0	0	0	0	1	0	0	0	1	1	0	0	1	1	0	0
1	1	0	0	1	1	0	1	0	0	0	1	1	0	0	1	1	1	0	1	1	1	1	0
1	1	0	0	1	1	1	0	0	0	1	0	1	0	1	0	1	1	1	0	1	1	0	1
1	1	0	0	1	1	1	1	0	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1
1	1	0	1	1	1	0	0	0	1	0	0	1	0	0	1	1	1	0	1	1	1	0	1
1	1	0	1	1	1	0	1	0	1	0	1	1	0	0	0	1	1	0	0	1	1	1	1
1	1	0	1	1	1	1	0	0	1	1	0	1	0	1	1	1	1	1	1	1	1	0	0
1	1	0	1	1	1	1	1	0	1	1	1	1	0	1	0	1	1	1	0	1	1	1	0
1	1	1	0	1	1	0	0	1	0	0	0	1	1	0	0	1	1	1	0	1	1	1	0
1	1	1	0	1	1	0	1	1	0	0	1	1	1	0	1	1	1	1	1	1	1	0	0
1	1	1	0	1	1	1	0	1	0	1	0	1	1	1	0	1	1	0	0	1	1	1	1
1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0	1	1	1	0	1
1	1	1	1	1	1	0	0	1	1	0	0	1	1	0	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	0	1	1	1	0	1	1	1	0	0	1	1	1	0	1	1	0	1
1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	1	1	1	0	1	1	1	1	0
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	0	1	1	0	0

A.2 Proof of Propositions for Neural Information Squeezer Network

Here we provide mathematical proves for the two propositions of the neural network framework to calculate mutual information and redundancy.

First, we rephrase Proposition 1, then give the proof here.

Proposition 1: For any random variables X and Y, we can use the framework of Figure 7(a) to predict Y by squeezing the information channel of $\hat{Y}_X$ as the minimum dimension but satisfying $\hat{Y} \approx Y$ and $U \sim \mathcal{N}(0, I)$. And we suppose the conditional entropy $H(X|Y) > 0$ holds, then:

$$H(\hat{Y}_X) \approx I(X;Y) \tag{6}$$

Proof. The whole structure of the alternative NIS network (Figure 7(a)) can be regarded as a similar structure as in Ref [21], but the dynamics learner is absent. However, we can understand the dynamic is a fixed identical mapping. In this way, all the conclusions proved in [21] can be applied here. Thus, we have:

$$I(X;Y) \approx I(\hat{Y}_X; \hat{Y}_X) = H(\hat{Y}_X) \tag{7}$$

if all the neural networks are well-trained. The first equation holds because of Theorem 2 (information bottle-neck) and Theorem 3(mutual information of the model will be closed to the data for a well-trained framework) in [21], the second holds when q is minimized such that the information channel of $\hat{Y}_X$ is squeezed as possible as we can and because of the property of mutual information.

Further, because U is an independent Gaussian noise, therefore:

$$H(U) = H(\psi(X)) - H(\hat{Y_X}) \approx H(X) - I(X;Y) = H(X|Y) \tag{8}$$

The approximated equation holds because ψ is a bijector that can keep the entropy unchanged, and Equation 7 holds. Therefore, we can prove Proposition 1. □

To calculate the redundancy for a system with three variables, we can further feed the variable of $\hat{Y}_X$ into another NIS network to predict Z and narrow down the information channel of the intermediate variable $\hat{Z}_{\hat{Y}_X}$ to get the minimum dimension $q^{*'}$ for $\hat{Z}_{\hat{Y}_X}$, then its Shannon entropy can approach the redundancy. The redundancy satisfies the property of permutation symmetry for all the variables. We can prove the following proposition:

Proposition 2: For a system with three random variables X, Y, Z, suppose the conditional information $H(X|Y) > 0, H(X|Z) > 0$, then the redundancy calculated by Equation 4 is symmetric, which means:

$$Red(X, Y, Z) \approx Red(X, Z, Y) \tag{9}$$

Proof. If we accept the definition of Equation 4, then:

$$Red(X, Y, Z) \approx H(\hat{Z}_{\hat{Y}_X}) = H(\hat{Y}_X) - H(U_{\hat{Y}_X}) = H(X) - H(X|Y) - H(\hat{Y}_X|Z), \tag{10}$$

where $U_{\hat{Y}_X}$ is the discarded Gaussian noise to predict $\hat{Y}_Z$.

In another way, we can use X to predict Z. The intermediate variable $\hat{Z}_X$ can be used to predict Y, and the intermediate variable $\hat{Y}_{\hat{Z}_X}$ can be used to approximate the redundancy which is denoted as $Red(X, Z, Y)$. Therefore,

$$Red(X, Z, Y) \approx H(X) - H(X|Z) - H(\hat{Z}_X|Y). \tag{11}$$

Because the discarded noise variable $U_{\hat{Y}_X}$ in the process of predicting Y by X is independent on all the variables, therefore:

$$H(U_{\hat{Y}_X}) = H(U_{\hat{Y}_X}|Z) = H(U_{\hat{Y}_X}|Y, Z) = H(X|Y, Z), \tag{12}$$

Similarly, the discarded noise variable $U_{\hat{Z}_{\hat{Y}_X}}$ in the process of predicting Z by $\hat{Y}_X$ is also independent on all the other variables, and $\psi(X)$ is the combination of $U_{\hat{Y}_X}$ and $\hat{Y}_X$, thus:

$$H(X|Y,Z) = H(U_{\hat{Y}_X}|Z) = H(X|Z) - H(\hat{Y}_X|Z). \tag{13}$$

In the same way, we can obtain:

$$H(X|Z,Y) = H(U_{\hat{Z}_X}|Y) = H(X|Y) - H(\hat{Z}_X|Y). \tag{14}$$

Because $H(X|Y,Z) = H(X|Z) - H(\hat{Y}_X|Z) = H(X|Y,Z) = H(X|Y) - H(\hat{Z}_X|Y)$, therefore:

$$H(X|Z) + H(\hat{Z}_X|Y) = H(X|Y) + H(\hat{Y}_X|Z) \tag{15}$$

and the Equation 10 and 11 lead to:

$$Red(X,Y,Z) = Red(X,Z,Y). \tag{16}$$

This equation is general for all the permutations of X, Y, and Z. Thus, the redundancy defined in the neural network NIS satisfies permutation symmetry.

□