Non-Negative Decomposition of Multivariate Information: From Minimum to Blackwell Specific Information

Tobias Mages *, Elli Anastasiadi, Christian Rohner

Posted Date: 6 March 2024

doi: 10.20944/preprints202403.0285.v1

Keywords: partial information decomposition; redundancy; synergy; information flow analysis; f-information; rényi-information

 doi:10.20944/preprints202403.0285.v1

Article

Non-Negative Decomposition of Multivariate Information: From Minimum to Blackwell Specific Information

Tobias Mages *, Elli Anastasiadi and Christian Rohner

Department of Information Technology, Uppsala University, 752 36 Uppsala, Sweden; Firstname.Surname@it.uu.se
* Correspondence: tobias.mages@it.uu.se

Abstract: Partial Information Decompositions (PIDs) aim to categorize how a set of source variables provide information about a target variable redundantly, uniquely, or synergetically. The original proposal for such an analysis used a lattice-based approach and gained significant attention. However, finding a suitable underlying decomposition measure is still an open research question, even at an arbitrary number of discrete random variables. This work proposes a solution to this case with a non-negative PID that satisfies an inclusion-exclusion relation for any f-information measure. The decomposition is constructed from a pointwise perspective of the target variable to take advantage of the equivalence between the Blackwell and zonogon order in this setting. We prove that the decomposition satisfies the axioms of the original decomposition framework and guarantees non-negative partial information results. We highlight that our decomposition behaves differently depending on the used information measure, which can be utilized for different applications. We additionally show how our proposal can be used to obtain a non-negative decomposition of Rényi-information at a transformed inclusion-exclusion relation, and for tracing partial information flows through Markov chains.

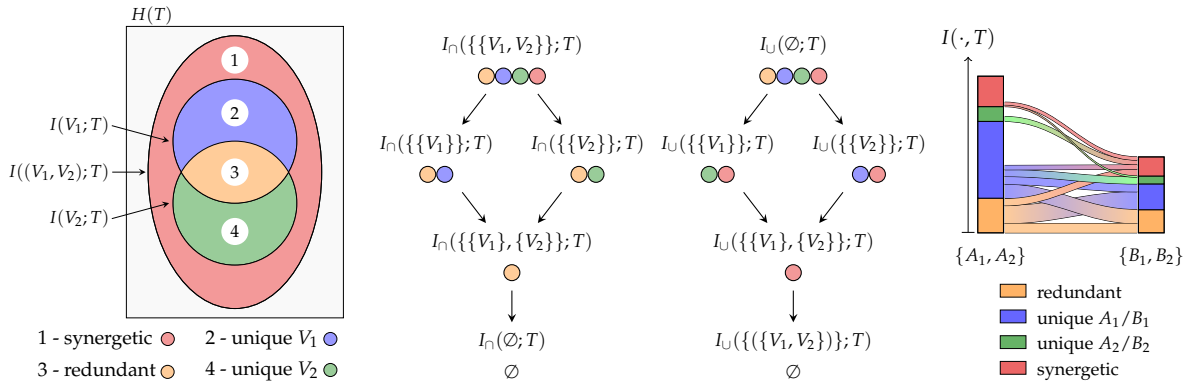
Keywords: partial information decomposition; redundancy; synergy; information flow analysis; f-information, rényi-information

1. Introduction

From computer science to neuroscience, we can find the following problem: We would like to know information about a random variable T, called the target, that we cannot observe directly. However, we can obtain information about the target indirectly from another set of variables $\mathbf{V} = \{V_1, ..., V_n\}$. We can use information measures to quantify how much information any set of variables provides about the target. When doing so, we can identify the concept of *redundancy*: For example, if we have two identical variables $V_1 = V_2$, then we can use one variable to predict the other and thus anything that this other variable can predict. Similarly, we can identify the concept of *synergy*: For example, if we have two independent variables and a target that corresponds to their XOR operation $T = (V_1 \text{ XOR } V_2)$, then both variables provide no advantage on their own for predicting the state of T, yet their combination fully determines it. Williams and Beer [1] suggested that it is possible to characterize information as visualized by the Venn diagram for two variables $\mathbf{V} = \{V_1, V_2\}$ in Figure 1a. This decomposition attributes the total information about the target to being redundant, synergetic, or unique to a particular variable. As indicated in Figure 1a by $I(\cdot, T)$, we can quantify three of the areas using information measures. However, this is insufficient to determine the four partial areas that represent the individual contributions. This causes the necessity to extend an information measure to either quantify the amount of redundancy or synergy between a set of variables.

Williams and Beer [1] first proposed a framework for Partial Information Decompositions (PIDs) and found favor by the community [2]. However, the proposed measure of redundancy was criticized for not distinguishing "the *same* information and the *same amount* of information" [3–6]. The proposal of Williams and Beer [1] focused specifically on mutual information. This work additionally studies the decomposition of any f-information or Rényi-information at discrete random variables. They have significance, among others, in parameter estimations, high-dimensional statistics, hypothesis testing, channel coding, data compression and privacy analyses [7,8].

 doi:10.20944/preprints202403.0285.v1



(a) Desired set-theoretic analogy **(b)** Redundancy Lattice $I_\cap$: using a *redundancy measure* **(c)** Synergy Lattice $I_\cup$: using a *loss measure* **(d)** Information flow visualization

Figure 1. Partial information decomposition representations at two variables $\mathbf{V} = \{V_1, V_2\}$. **(a)** Visualization of the desired intuition for multivariate information as Venn diagram. **(b)** Representation as redundancy lattice, where $I_\cap$ quantifies the information that is contained in all of its provided variables (inside their intersection). The ordering represents the expected subset relation of redundancy. **(c)** Representation as synergy lattice, where $I_\cup$ quantifies the information that is contained in neither of its provided variables (outside their union). **(d)** When having two partial information decompositions with respect to the same target variable, we can study how the partial information of one decomposition propagates into the next. We refer to this as information flow analysis of a Markov chain such as $T \to (A_1, A_2) \to (B_1, B_2)$.

1.1. Related work

Most of the literature focuses on the decomposition of mutual information. Here, many alternative measures have been proposed but cannot suitably replace the original measure of Williams and Beer [1]: They are either limited to two observable variables [5,9,10], provide negative partial contributions [11–13], or provide results in which the partial contributions do not sum to the total amount [14]. Griffith et al. [3] studied the decomposition of zero-error information and obtained negative partial contributions. Kolchinsky [14] proposed a decomposition framework that is applicable beyond Shannon information theory, however, where the partial contributions do not sum to the total amount.

In this work, we propose a decomposition measure for replacing the one presented by William and Beer's [1] while maintaining its desired properties. To achieve this, we combine several concepts from the literature: We use the Blackwell order, a preorder of information channels, for the decomposition and for deriving its operational interpretation, similar to Bertschinger et al. [9] and Kolchinsky [14]. We use its special case for binary input channels, the zonogon order studied by Bertschinger and Rauh [15], to achieve non-negativity at an arbitrary number of variables and provide it with a practical meaning. To utilize this special case for a general decomposition, we use the concept of a target pointwise decomposition as demonstrated by Williams and Beer [1] and related to Lizier et al. [16], Finn and Lizier [11], and Ince [12]. Finally, we apply the concepts from measuring on lattices, discussed by Knuth [17], to transform a non-negative decomposition with inclusion-exclusion relation from one information measure to another while maintaining the decomposition properties.

Remark 1. *We use the term 'target pointwise' or simply 'pointwise' within this work to refer to the analysis of each target state individually. This differs from [11,12,16], who use the latter term for the analysis of all joint sources-target realizations.*

 doi:10.20944/preprints202403.0285.v1

1.2. Contributions

In a recent work [18], we presented a decomposition of mutual information on the redundancy lattice (Figure 1b). This work aims to simplify, generalize and extend these ideas to make the following contributions to the area of Partial Information Decompositions:

- We propose a representation of distinct uncertainty and distinct information, which is used to demonstrate the unexpected behavior of the measure by Williams and Beer [1] (Section 2.2 and 3).
- We propose a decomposition for any f-information on both the redundancy lattice (Figure 1b) and synergy lattice (Figure 1c) that satisfies an inclusion-exclusion relation and provides a meaningful operational interpretation (Section 3.2).
- We prove that the proposed decomposition satisfies the original axioms of Williams and Beer [1] and guarantees non-negative partial information (Theorem 3).
- We propose to transform the non-negative decomposition of one information measure into another. This transformation maintains the non-negativity and its inclusion-exclusion relation under a re-definition of information addition (Section 3.3).
- We demonstrate the transformation of an f-information decomposition into a decomposition for Rényi- and Bhattacharyya-information (Section 3.3).
- We demonstrate that the proposed decomposition obtains different properties from different information measures and analyze the behavior of total variation in more detail (Section 4).
- We demonstrate the analysis of partial information flows through Markov chains (Figure 1d) for each information measure on both the redundancy and synergy lattice (Section 4.2).

2. Background

This section aims to provide the required background information and introduce the used notation. Section 2.1 discusses the Blackwell order and its special case at binary targets, the zonogon order, which will be used for operational interpretations and the representation of f-information for its decomposition. Section 2.2 discusses the PID framework of Williams and Beer [1] and the relation between a decomposition based on the redundancy lattice and one based on the synergy lattice. We also demonstrate the unintuitive behavior of the original decomposition measure which will be resolved by our proposal in Section 3. Section 2.3 provides the considered definitions of f-information, Rényi-information, and Bhattacharyya information for the later demonstration of transforming decomposition results between measures.

Notation 1 (Random variables and their distribution)**.** *We use the notation T (upper case) to represent a random variable, ranging over the event space $\mathcal{T}$ (calligraphic) containing events $t \in \mathcal{T}$ (lower case), and use the notation P_T (P with subscript) to indicate its probability distribution. The same convention applies to other variables, such as a random variable S with events $s \in \mathcal{S}$ and distribution P_S. We indicate the outer product of two probability distributions as $P_S \otimes P_T$, which assigns the product of their marginals $P_S(s) \cdot P_T(t)$ to each event (s, t) of the Cartesian product $\mathcal{S} \times \mathcal{T}$. Unless stated otherwise, we use the notation T, S and V to represent random variables throughout this work.*

2.1. Blackwell and Zonogon Order

Definition 1 (Channel)**.** *A channel $\mu = T \to S$ from $\mathcal{T}$ to $\mathcal{S}$ represents a garbling of the input variable T that results in variable S. Within this work, we represent an information channel μ as (row) stochastic matrix, where each element is non-negative, and all rows sum to one.*

For the context of this work, we consider a variable S to be the observation of the output from an information channel $T \to S$ from the target variable T, such that the corresponding channel can be

 doi:10.20944/preprints202403.0285.v1

obtained from their conditional probability distribution, as shown in Equation 1 where $\mathcal{T} = \{t_1, ..., t_n\}$ and $\mathcal{S} = \{s_1, ..., s_m\}$.

$$\mu = (T \to S) = P_{(S|T)} = \begin{bmatrix} p(s_1 \mid t_1) & \dots & p(s_m \mid t_1) \\ \vdots & \ddots & \vdots \\ p(s_1 \mid t_n) & \dots & p(s_m \mid t_n) \end{bmatrix} \tag{1}$$

Notation 2 (Binary input channels). *Throughout this work, we reserve the symbol κ for binary input channels, meaning κ signals a stochastic matrix of dimension $2 \times m$. We use the notation $\vec{v} \in \kappa$ to indicate a column of this matrix.*

Definition 2 (More informative [15,19]). *An information channel $\mu_1 = T \to S_1$ is more informative than another channel $\mu_2 = T \to S_2$ if - for any decision problem involving a set of actions $a \in \Omega$ and a reward function $u : (\Omega, \mathcal{T}) \to \mathbb{R}$ that depends on the chosen action and state of the variable T - an agent with access to S_1 can always achieve an expected reward at least as high as another agent with access to S_2.*

Definition 3 (Blackwell order [15,19]). *The Blackwell order is a preorder of channels. A channel μ_1 is Blackwell superior to channel μ_2, if we can pass its output through a second channel λ to obtain an equivalent channel to μ_2, as shown in Equation 2.*

$$\mu_2 \sqsubseteq \mu_1 \iff \mu_2 = \mu_1 \cdot \lambda \qquad \text{for some stochastic matrix } \lambda \tag{2}$$

Blackwell [19] showed that a channel is more informative if and only if it is Blackwell superior. Bertschinger and Rauh [15] showed that the Blackwell order does not form a lattice for channels $\mu = T \to S$ if $|\mathcal{T}| > 2$ since the ordering does not provide unique meet and join elements. However, binary target variables $|\mathcal{T}| = 2$ are a special case where the Blackwell order is equivalent to the zonogon order (discussed next) and does form a lattice [15].

Definition 4 (Zonogon [15]). *The zonogon $Z(\kappa)$ of a binary input channel $\kappa = T \to S$ is defined using the Minkowski sum from the collection of vector segments as shown in Equation 3. The zonogon $Z(\kappa)$ can similarly be defined as image of the unit cube $[0,1]^{|\mathcal{S}|}$ under the linear map of κ.*

$$Z(\kappa) := \left\{ \sum_i x_i \vec{v}_i \ : \ 0 \le x_i \le 1, \vec{v}_i \in \kappa \right\} = \left\{ \kappa a \ : \ a \in [0,1]^{|\mathcal{S}|} \right\} \tag{3}$$

The zonogon $Z(\kappa)$ is a centrally symmetric convex polygon, and the set of vectors $\vec{v}_i \in \kappa$ span its perimeter. Figure 2 shows the example of a binary input channel and its corresponding zonogon.

Definition 5 (Zonogon sum). *The addition of two zonogons corresponds to their Minkowski sum as shown in Equation 4.*

$$Z(\kappa_1) + Z(\kappa_2) := \{a + b \ : \ a \in Z(\kappa_1), b \in Z(\kappa_2)\} = Z\left(\begin{bmatrix} \kappa_1 & \kappa_2 \end{bmatrix}\right) \tag{4}$$

Definition 6 (Zonogon order [15]). *A zonogon $Z(\kappa_1)$ is zonogon superior to another $Z(\kappa_2)$ if and only if $Z(\kappa_2) \subseteq Z(\kappa_1)$.*

Bertschinger and Rauh [15] showed that for binary input channels, the zonogon order is equivalent to the Blackwell order and forms a lattice (Equation 5). In the remaining work, we will only discuss

 doi:10.20944/preprints202403.0285.v1

binary input channels, such that the orderings of Definition 2, 3 and 6 are equivalent and can be thought of as zonogons with subset relation.

$$\kappa_1 \sqsubseteq \kappa_2 \iff Z(\kappa_1) \subseteq Z(\kappa_2) \tag{5}$$

For obtaining an interpretation of what a channel zonogon $Z(\kappa)$ represents, we can consider a binary decision problem by aiming to predict the state $t \in \mathcal{T}$ of a *binary* target variable T using the output of channel $\kappa = T \to S$. Any decision strategy $\lambda \in [0,1]^{|\mathcal{S}|\times 2}$ for obtaining a binary prediction $\hat{T}$ can be fully characterized by its resulting pair of True-Positive Rate (TPR) and False-Positive Rate (FPR), as shown in Equation 6

$$\kappa \cdot \lambda = (T \to S \to \hat{T}) = P_{(\hat{T}|T)} = \begin{bmatrix} p(\hat{T}=t \mid T=t) & p(\hat{T}\neq t \mid T=t) \\ p(\hat{T}=t \mid T\neq t) & p(\hat{T}\neq t \mid T\neq t) \end{bmatrix} = \begin{bmatrix} \text{TPR} & 1-\text{TPR} \\ \text{FPR} & 1-\text{FPR} \end{bmatrix} \tag{6}$$

Therefore, a channel zonogon $Z(\kappa)$ provides the set of all achievable (TPR,FPR)-pairs for a given channel κ [18,20]. This can also be seen from Equation 3, where the unit cube $a \in [0,1]^{|\mathcal{S}|}$ represents all possible first columns of the decision strategy λ. The first column of λ fully determines the second since each row has to sum to one. As a result, κa provides the (TPR,FPR)-pair for the decision strategy $\lambda = \begin{bmatrix} a & (1-a) \end{bmatrix}$ and the definition of Equation 3 all achievable (TPR,FPR)-pairs for predicting the state of a binary target variable. Since this will be helpful for operational interpretations, we label the axis of zonogon plots accordingly, as shown in Figure 2, and refer to regions within this plot as reachable decision regions:

Definition 7 (Reachable decision region)**.** *A reachable decision region for a binary decision problem is a set of achievable (TPR,FPR) performance pairs and can be visualized in a TPR/FPR-plot such as Figure 2.*

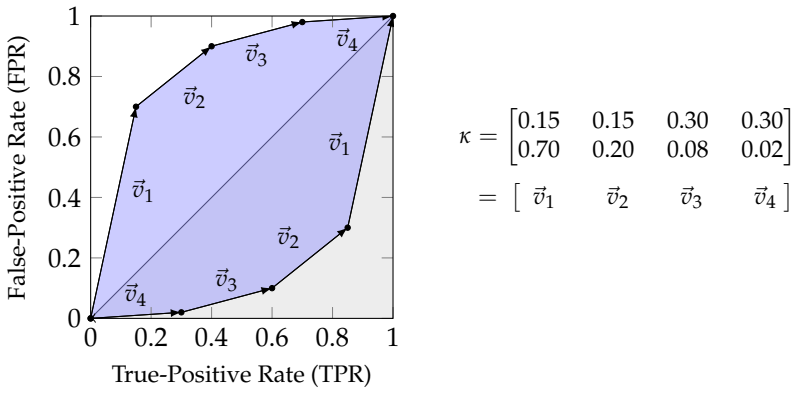


Figure 2. An example zonogon (blue) for a binary input channel κ. The zonogon perimeter corresponds to the vectors $\vec{v}_i \in \kappa$ sorted by increasing/decreasing slope for the lower/upper half. The zonogon also corresponds to the set of (TPR,FPR)-pairs that are achievable when predicting the binary target variable.

Notation 3 (Channel lattice)**.** *We use the notation $\kappa_1 \sqcap \kappa_2$ for the meet element of binary input channels under the Blackwell order and $\kappa_1 \sqcup \kappa_2$ for their join element. We use the notation $\top_{BW} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ for the top element of binary input channels under the Blackwell order and $\bot_{BW} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ for the bottom element.*

For binary input channels, the meet element of the Blackwell order corresponds to the zonogon intersection $Z(\kappa_1 \sqcap \kappa_2) = Z(\kappa_1) \cap Z(\kappa_2)$ and the join element of the Blackwell order corresponds to

 doi:10.20944/preprints202403.0285.v1

the convex hull of their union $Z(\kappa_1 \sqcup \kappa_2) = \text{Conv}(Z(\kappa_1) \cup Z(\kappa_2))$. Equation 7 describes this for an arbitrary number of channels.

$$Z\left(\bigsqcap_{\kappa \in \mathbf{A}} \kappa\right) = \bigcap_{\kappa \in \mathbf{A}} Z(\kappa) \qquad \text{and} \qquad Z\left(\bigsqcup_{\kappa \in \mathbf{A}} \kappa\right) = \text{Conv}\left(\bigcup_{\kappa \in \mathbf{A}} Z(\kappa)\right) \tag{7}$$

2.2. Partial Information Decomposition

The commonly used framework for PIDs was introduced by Williams and Beer [1]. A PID is computed with respect to a particular random variable that we would like to know information *about*, called the target, and tries to identify *from* which variables that we have access to, called visible variables, we obtain this information. Therefore, this section considers sets of variables that represent their joint distribution.

Notation 4. *Throughout this work, we use the notation T for the target variable and $\mathbf{V} = \{V_1, ..., V_n\}$ for the set of visible variables. We use the notation $\mathcal{P}(\mathbf{V})$ for the power set of $\mathbf{V}$, and $\mathcal{P}_1(\mathbf{V}) = \mathcal{P}(\mathbf{V}) \setminus \emptyset$ for its power set without the empty set.*

Definition 8 (Sources, Atoms [1])**.**

- *A source $\mathbf{S}_i \in \mathcal{P}_1(\mathbf{V})$ is a non-empty set of visible variables.*
- *An atoms $\alpha \in \mathcal{A}(\mathbf{V})$ is a set of sources constructed by Equation 8.*

$$\mathcal{A}(\mathbf{V}) = \{\alpha \in \mathcal{P}_1(\mathcal{P}_1(\mathbf{V})) \ : \ \forall \mathbf{S}_a, \mathbf{S}_b \in \alpha, \mathbf{S}_a \not\subset \mathbf{S}_b\}, \tag{8}$$

The used filter for obtaining the set of atoms (Equation 8) removes sets that would be equivalent to other elements. This is required for obtaining a lattice from the following two ordering relations:

Definition 9 (Redundancy-/Gain-lattice [1])**.** *The redundancy lattice $(\mathcal{A}(\mathbf{V}), \preccurlyeq)$ is obtained by applying the ordering relation of Equation 9 to all atoms $\alpha, \beta \in \mathcal{A}(\mathbf{V})$.*

$$\alpha \preccurlyeq \beta \iff \forall \mathbf{S}_b \in \beta, \exists \mathbf{S}_a \in \alpha, \mathbf{S}_a \subseteq \mathbf{S}_b \tag{9}$$

The redundancy lattice for three visible variables is visualized in Figure 3a. On this lattice, we can think of an atom as representing the information that can be obtained from all of its sources about the target T (their redundancy or informational intersection). For example, the atom $\alpha = \{\{V_1, V_2\}, \{V_1, V_3\}\}$ represents on the redundancy lattice the information that is contained in both (V_1, V_2) and (V_1, V_3) about T. Since both sources in α provide the information of V_1, their redundancy contains at least this information, and the atom $\beta = \{\{V_1\}\}$ is considered its predecessor. Therefore, the ordering indicates an informational subset relation for the redundancy of atoms, and the information that is represented by an atom increases as we move up. The up-set of an atom α on the redundancy lattice indicates the information that is lost when losing all of its sources. Considering the example from above, if we lose access to $\{V_1$ (or) $V_2\}$ and $\{V_1$ (or) $V_3\}$, then we lose access to all atoms in the up-set of $\alpha = \{\{V_1, V_2\}, \{V_1, V_3\}\}$.

Definition 10 (Synergy-/Loss-lattice [21])**.** *The synergy lattice $(\mathcal{A}(\mathbf{V}), \preceq)$ is obtained by applying the ordering relation of Equation 10 to all atoms $\alpha, \beta \in \mathcal{A}(\mathbf{V})$.*

$$\alpha \preceq \beta \iff \forall \mathbf{S}_b \in \beta, \exists \mathbf{S}_a \in \alpha, \mathbf{S}_b \subseteq \mathbf{S}_a \tag{10}$$

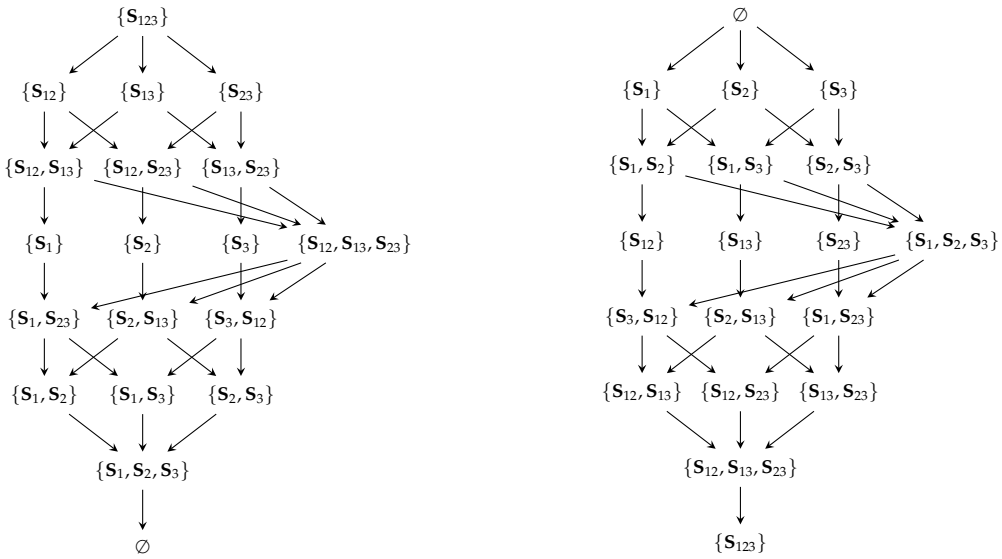
The synergy lattice for three visible variables is visualized in Figure 3b. On this lattice, we can think of an atom as representing the information that is contained in neither of its sources (information outside their union). For example, the atom $\alpha = \{\{V_1, V_2\}, \{V_1, V_3\}\}$ represents on the synergy

 doi:10.20944/preprints202403.0285.v1

lattice the information that is obtained from neither (V_1, V_2) nor (V_1, V_3) about T. The ordering again indicates their expected subset relation: the information that is obtained from neither $\{V_1$ (and) $V_2\}$ nor $\{V_1$ (and) $V_3\}$ is fully contained in the information that cannot be obtained from $\beta = \{\{V_1\}\}$ and thus α is a predecessor of β.

With an intuition for both ordering relations in mind, we can see how the filter in the construction of atoms (Equation 8) removes sets that would be equivalent to another atom: the set $\{\{V_1, V_2\}, \{V_1\}\}$ is removed from the power set of sources since it would be equivalent to the atom $\{\{V_1\}\}$ under the ordering of the redundancy lattice and equivalent to the atom $\{\{V_1, V_2\}\}$ under the ordering of the synergy lattice.

Notation 5 (Redundancy/Synergy lattices)**.** *We use the notation $(\mathcal{A}(V), \curlyvee, \curlywedge)$ for the join and meet operators on the redundancy lattice, and $(\mathcal{A}(V), \vee, \wedge)$ for the join and meet operators on the synergy lattice. We use the notation $\top_{RL} = \{\mathbf{V}\}$ for the top and $\bot_{RL} = \emptyset$ for the bottom atom on the redundancy lattice, and $\top_{SL} = \emptyset$ and $\bot_{SL} = \{\mathbf{V}\}$ for the top and bottom atom on the synergy lattice. For an atom α, we use the notation $\downarrow\alpha$ for its down-set, $\downarrow\!\!\!\!\!\!\cdot\,\alpha$ for its strict down-set, and α^- for its cover set. These definitions will only appear in the Möbius inverse of a function that is directly associated with either the synergy or redundancy lattice such that there is no ambiguity about which ordering relation has to be considered.*



(a) Redundancy-/Gain-Lattice $(\mathcal{A}(\{V_1, V_2, V_3\}), \preccurlyeq)$ (quantifies information present in all sources)

(b) Synergy-/Loss-Lattice $(\mathcal{A}(\{V_1, V_2, V_3\}), \preceq)$ (quantifies information present in neither source)

Figure 3. For the visualization, we abbreviated the notation by indicating the contained visible variable as index of the source, for example, $\mathbf{S}_{12} = \{V_1, V_2\}$ to represent their joint distribution: **(a)** A redundancy lattice based on the ordering $\preccurlyeq$ of Equation 9. **(b)** A synergy lattice based on the ordering $\preceq$ of Equation 10 for the partial information decomposition at $\mathbf{V} = \{V_1, V_2, V_3\}$. On the redundancy lattice, the redundancy of all sources within an atom increases while moving up. On the synergy lattice, the information that is obtained from neither source of an atom increases while moving up.

The redundant, unique, or synergetic information (partial contributions) can be calculated based on either lattice. They are obtained by quantifying each atom of the redundancy or synergy lattice with a cumulative measure that increases as we move up in the lattice. The partial contributions are then obtained in a second step from a Möbius inverse.

Definition 11 ([Cumulative] redundancy measure [1])**.** *A redundancy measure $I_\cap(\alpha; T)$ is a function that assigns a real value to each atom of the redundancy lattice. It is interpreted as a cumulative information measure that quantifies the redundancy between all sources $\mathbf{S} \in \alpha$ of an atom $\alpha \in \mathcal{A}(\mathbf{V})$ about the target T.*

 doi:10.20944/preprints202403.0285.v1

Definition 12 ([Cumulative] loss measure [21]). *A loss measure $I_\cup(\alpha; T)$ is a function that assigns a real value to each atom of the synergy lattice. It is interpreted as a cumulative measure that quantifies the information about T that is provided by neither of the sources $\mathbf{S} \in \alpha$ of an atom $\alpha \in \mathcal{A}(\mathbf{V})$.*

To ensure that a redundancy measure actually captures the desired concept of redundancy, Williams and Beer [1] defined three axioms that a measure $I_\cap$ should satisfy. For the synergy lattice, we consider the equivalent axioms discussed by Chicharro and Panzeri [21]:

Axiom 1 (Commutativity [1,21]). *Invariance in the order of sources (σ permuting the order of indices):*

$$I_\cap(\{\mathbf{S}_1, ..., \mathbf{S}_i\}; T) = I_\cap(\{\mathbf{S}_{\sigma(1)}, ..., \mathbf{S}_{\sigma(i)}\}; T)$$
$$I_\cup(\{\mathbf{S}_1, ..., \mathbf{S}_i\}; T) = I_\cup(\{\mathbf{S}_{\sigma(1)}, ..., \mathbf{S}_{\sigma(i)}\}; T)$$

Axiom 2 (Monotonicity [1,21]). *Additional sources can only decrease redundant information. Additional sources can only decrease the information that is in neither source.*

$$I_\cap(\{\mathbf{S}_1, ..., \mathbf{S}_{i-1}\}; T) \geq I_\cap(\{\mathbf{S}_1, ..., \mathbf{S}_i\}; T)$$
$$I_\cup(\{\mathbf{S}_1, ..., \mathbf{S}_{i-1}\}; T) \geq I_\cup(\{\mathbf{S}_1, ..., \mathbf{S}_i\}; T)$$

Axiom 3 (Self-redundancy [1,21]). *For a single source, redundancy equals mutual information. For a single source, the information loss equals the difference between the total available mutual information and the mutual information of the considered source with the target.*

$$I_\cap(\{\mathbf{S}_i\}; T) = I(\mathbf{S}_i; T) \quad \textit{and} \quad I_\cup(\{\mathbf{S}_i\}; T) = I(\mathbf{V}; T) - I(\mathbf{S}_i; T)$$

The first axiom states that an atom's redundancy and information loss should not depend on the order of its sources. The second axiom states that adding sources to an atom can only decrease the redundancy of all sources (redundancy lattice) and decrease the information from neither source (synergy lattice). The third axiom binds the measures to be consistent with mutual information and ensures that the bottom element of both lattices is quantified to zero.

Once a lattice with corresponding cumulative measure ($I_\cap / I_\cup$) is defined, we can use the Möbius inverse to compute the partial contribution of each atom. This partial information can be visualized as partial area in a Venn diagram (see Figure 1a) and corresponds to the desired redundant, unique, and synergetic contributions. However, the same atom represents different partial contributions on each lattice: As visualized for the case of two visible variables in Figure 1, the unique information of variable V_1 is represented by $\alpha = \{\{V_1\}\}$ on the redundancy lattice and by $\beta = \{\{V_2\}\}$ on the synergy lattice.

Definition 13 (Partial information [1,21]). *Partial information $\Delta I_\cap(\alpha; T)$ and $\Delta I_\cup(\alpha; T)$ corresponds to the Möbius inverse of its corresponding cumulative measure on the respective lattice.*

$$\textit{Redundancy-Lattice:} \qquad \Delta I_\cap(\alpha; T) = I_\cap(\alpha; T) - \sum_{\beta \in \downarrow\alpha} \Delta I_\cap(\beta; T), \tag{11a}$$

$$\textit{Synergy-Lattice:} \qquad \Delta I_\cup(\alpha; T) = I_\cup(\alpha; T) - \sum_{\beta \in \downarrow\alpha} \Delta I_\cup(\beta; T). \tag{11b}$$

 doi:10.20944/preprints202403.0285.v1

Remark 2. *Using the Möbius inverse for defining partial information enforces an inclusion-exclusion relation in that all partial information contributions have to sum to the corresponding cumulative measure. Kolchinsky [14] argues that an inclusion-exclusion relation should not be expected to hold for PIDs and proposes an alternative decomposition framework. In this case, the sum of partial contributions (unique/redundant/synergetic information) is no longer expected to sum to the total amount $I(\mathbf{V}; T)$.*

Property 1 (Local positivity, non-negativity [1])**.** *A partial information decomposition satisfies non-negativity or local positivity if its partial information contributions are always non-negative, as shown in Equation 12.*

$$\forall \alpha \in \mathcal{A}(V). \qquad \Delta I_\cap(\alpha; T) \geq 0 \qquad or \qquad \Delta I_\cup(\alpha; T) \geq 0 \tag{12}$$

The non-negativity property is important if we assume an inclusion-exclusion relation since it states that the unique, redundant, or synergetic information cannot be negative. If an atom α provides a negative partial contribution in the framework of Williams and Beer [1], then this may indicate that we over-counted some information in its down-set.

Remark 3. *Several additional axioms and properties have been suggested since the original proposal of Williams and Beer [1], such as target monotonicity and target chain rule [4]. However, this work will only consider the axioms and properties of Williams and Beer [1]. To the best of our knowledge, no other measure since the original proposal (discussed below) has been able to satisfy these properties for an arbitrary number of visible variables while ensuring an inclusion-exclusion relation for their partial contributions.*

It is possible to convert between both representations due to a lattice duality:

Definition 14 (Lattice duality and dual decompositions [21])**.** *Let $C = (\mathcal{A}(\mathbf{V}), \preccurlyeq)$ be a redundancy lattice with associated measure $I_\cap$ and let $D = (\mathcal{A}(\mathbf{V}), \preceq)$ be a synergy lattice with measure $I_\cup$, then the two decompositions are said to be dual if and only if the down-set on one lattice corresponds to the up-set in the other as shown in Equation 13.*

$$\forall \alpha \in C,\ \exists \beta \in D : \Delta I_\cap(\alpha; T) = \Delta I_\cup(\beta; T) \tag{13a}$$
$$\forall \alpha \in D, \exists \beta \in C\ : \Delta I_\cup(\alpha; T) = \Delta I_\cap(\beta; T) \tag{13b}$$
$$\forall \alpha \in C,\ \exists \beta \in D : \quad I_\cap(\alpha; T) = \sum_{\gamma \in \downarrow \alpha} \Delta I_\cap(\gamma; T) = \sum_{\gamma \in \uparrow \beta} \Delta I_\cup(\gamma; T) \tag{13c}$$
$$\forall \alpha \in D, \exists \beta \in C\ : \quad I_\cup(\alpha; T) = \sum_{\gamma \in \downarrow \alpha} \Delta I_\cup(\gamma; T) = \sum_{\gamma \in \uparrow \beta} \Delta I_\cap(\gamma; T) \tag{13d}$$

Williams and Beer [1] proposed $I_\cap^{\min}$, as shown in Equation 14, to be used as measure of redundancy and demonstrated that it satisfies the three required axioms and local positivity. They define redundancy (Equation 14b) as the expected value of the minimum *specific information* (Equation 14a).

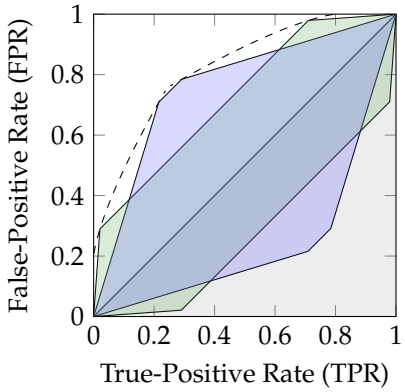
Remark 4. *Throughout this work, we use the term 'target pointwise information' or simply 'pointwise information' to refer to 'specific information'. This shall avoid confusion when naming their corresponding binary input channels in Section 3.*

$$I(\mathbf{S}_i; T = t) = \sum_{s \in \mathcal{S}_i} p(s \mid t) \left[\log\left(\frac{1}{p(t)}\right) - \log\left(\frac{1}{p(t \mid s)}\right) \right] \tag{14a}$$
$$I_\cap^{\min}(\mathbf{S}_1, ..., \mathbf{S}_k; T) = \sum_{t \in \mathcal{T}} p(t) \min_{i \in 1..k} I(\mathbf{S}_i; T = t). \tag{14b}$$

 doi:10.20944/preprints202403.0285.v1

To the best of our knowledge, this measure is the only existing non-negative decomposition that satisfies all three axioms listed above for an arbitrary number of visible variables while providing an inclusion-exclusion relation of partial information.

However, the measure $I_\cap^{\min}$ could be criticized for not providing a notion of distinct information due to its use of a pointwise minimum (for each $t \in \mathcal{T}$) over the sources. This leads to the question of distinguishing "the *same* information and the *same amount* of information" [3–6]. We can use the definition through a pointwise minimum (Equation 14) to construct examples of unexpected behavior: consider for example a uniform binary target variable T and two visible variables as output of the channels visualized in Figure 4. The channels are constructed to be equivalent for both target states and provide access to distinct decision regions while ensuring a constant pointwise information $\forall t \in \mathcal{T} : I(V_x, T = t) = 0.2$.



$$\kappa(x,y) = \begin{bmatrix} x & 1-x-y & y \\ y & 1-x-y & x \end{bmatrix}, \text{where } x \leq y.$$

$$I(V_1;T) = I_\cap^{\min}(\{V_1, V_2\};T) = I(V_2,T) = 0.2$$

Figure 4. Example of the unexpected behavior of $I_\cap^{\min}$: the dashed isoline indicates the pairs (x,y) for which channel $\kappa(x,y) = T \to V_i$ results in a pointwise information $\forall t \in \mathcal{T} : I(V_i, T = t) = 0.2$ for a uniform binary target variable. Even though observing the output of both indicated example channels (blue/green) provides significantly different abilities for predicting the target variable state, the measure $I_\cap^{\min}$ indicates full redundancy.

Even though our ability to predict the target variable significantly depends on which of the two indicated channel outputs we observe (blue or green in Figure 4), the measure $I_\cap^{\min}$ concludes full redundancy between them $I(V_1;T) = I_\cap^{\min}(\{V_1, V_2\};T) = I(V_2,T) = 0.2$. We think this behavior is undesired and, as discussed in the literature, caused by an underlying lack of distinguishing the *same* information. To resolve this issue, we will present a representation of f-information in Section 3.1, which allows the use of all (TPR,FPR)-pairs for each state of the target variable to represent a distinct notion of uncertainty.

2.3. Information Measures

This section discusses two generalizations of mutual information at discrete random variables based on f-divergences and Rényi divergences [22,23]. While mutual information has interpretational significance in channel coding and data compression, other f-divergences have their significance in parameter estimations, high-dimensional statistics, and hypothesis testing [7, p. 88], while Rényi-divergences can be found among others in privacy analysis [8]. Finally, we introduce Bhattacharyya information for demonstrating that it is possible to chain decomposition transformations in Section 3.3. All definitions in this section only consider the case of discrete random variables (which is what we need for the context of this work).

Definition 15 (f-divergence [22]). *Let $f : (0,\infty) \to \mathbb{R}$ be a function that satisfies the following three properties.*

- *f is convex,*

 doi:10.20944/preprints202403.0285.v1

- $f(1) = 0$,
- $f(z)$ *is finite for all* $z > 0$.

By convention we understand that $f(0) = \lim_{z\to 0^+} f(z)$ *and* $0f\left(\frac{0}{0}\right) = 0$. *For any such function* f *and two discrete probability distributions* P *and* Q *over the event space* $\mathcal{X}$, *the* f*-divergence for discrete random variables is defined as shown in Equation 15.*

$$D_f(P \parallel Q) \coloneqq \sum_{x\in\mathcal{X}} Q(x) f\left(\frac{P(x)}{Q(x)}\right) = \mathbb{E}_Q\left[f\left(\frac{P(X)}{Q(X)}\right)\right] \tag{15}$$

Notation 6. *Throughout this work, we reserve the name* f *for functions that satisfy the required properties for an* f*-divergence of Definition 15.*

An f-divergence quantifies a notion of dissimilarity between two probability distributions P and Q. Key properties of f-divergences are their non-negativity, their invariance under bijective transformations, and them satisfying a data-processing inequality [7, p. 89]. A list of commonly used f-divergences is shown in Table 1. Notably, the continuation for $a = 1$ of both the Hellinger- and α-divergence result in the KL-divergence [24].

Table 1. Commonly used functions for f-divergences.

D_{KL}	Kullback-Leiber (KL) divergence	$f(z) = z\log z$
D_{TV}	Total Variation (TV)	$f(z) = \frac{1}{2}\lvert z-1\rvert$
D_{χ^2}	χ^2-divergence	$f(z) = (z-1)^2$
D_{H^2}	Squared Hellinger distance	$f(z) = (1-\sqrt{z})^2$
D_{LC}	Le Cam distance	$f(z) = \frac{1-z}{2z+2}$
D_{JS}	Jensen-Shannon divergence	$f(z) = z\log\frac{2z}{z+1} + \log\frac{2}{z+1}$
D_{H_a}	Hellinger divergence with $a \in (0,1)\cup(1,\infty)$	$f(z) = \frac{z^a-1}{a-1}$
$D_{\alpha=a}$	α-divergence with $a \in (0,1)\cup(1,\infty)$	$f(z) = \frac{z^a-1-a(z-1)}{a(a-1)}$

The generator function of an f-divergence is not unique since $D_{f(z)} = D_{f(z)+c(z-1)}$ for a real constant $c \in \mathbb{R}$ [7, p. 90f.]. As a result, the considered α-divergence is a linear scaling of the Hellinger divergence ($D_{\mathrm{H}_a} = a \cdot D_{\alpha=a}$) as shown in Equation 16.

$$\frac{z^a-1}{a-1} + c(z-1) = a\cdot\frac{z^a-1-a(z-1)}{a(a-1)} \qquad \text{for } c = -\frac{a}{a-1} \tag{16}$$

Definition 16 (f-information [7])**.** *An* f*-information is defined based on an* f*-divergence from the joint distribution of two discrete random variables and the product of their marginals as shown in Equation 17.*

$$\begin{aligned} I_f(S;T) &\coloneqq D_f\left(P_{(S,T)} \parallel P_S \otimes P_T\right) \\ &= \sum_{(s,t)\in\mathcal{S}\times\mathcal{T}} P_S(s)\cdot P_T(t)\cdot f\left(\frac{P_{(S,T)}(s,t)}{P_S(s)\cdot P_T(t)}\right) \\ &= \sum_{t\in\mathcal{T}} P_T(t)\left[\sum_{s\in\mathcal{S}} P_S(s)\cdot f\left(\frac{P_{S|T}(s\mid t)}{P_S(s)}\right)\right] \end{aligned} \tag{17}$$

Definition 17 (f-entropy)**.** *A notion of* f*-entropy for a discrete random variable is obtained from the self-information of a variable* $H_f(T) \coloneqq I_f(T;T)$.

Notation 7. *Using the KL-divergence results in the definition of mutual information and Shannon entropy. Therefore, we use the notation* I_{KL} *for mutual information (KL-information) and* H_{KL} *(KL-entropy) for the Shannon entropy.*

 doi:10.20944/preprints202403.0285.v1

The remaining part of this section will define Rényi- and Bhattacharyya-information to highlight that they can be represented as an invertible transformation of Hellinger-information. This will be used in Section 3.3 to transform the decomposition of Hellinger-information to a decomposition of Rényi- and Bhattacharyya-information.

Remark 5. *We could similarly choose to represent Renyi divergence as a transformation of the α-divergence. A liner scaling of the considered f-divergence will however not affect our later results (see Section 3.3).*

Definition 18 (Rényi divergence [23])**.** *Let P and Q be two discrete probability distributions over the event space $\mathcal{X}$, then Rényi divergence R_a is defined as shown in Equation 18 for $a \in (0,1) \cup (1,\infty)$, and extended to $a \in \{0,1,\infty\}$ by continuation.*

$$\begin{aligned} R_a(P \parallel Q) &:= \frac{1}{a-1} \log\left(\mathbb{E}_Q\left[\left(\frac{P(X)}{Q(X)} \right)^a \right] \right) \\ &= \frac{1}{a-1} \log\left(1 + (a-1)\mathbb{E}_Q\left[\frac{\left(\frac{P(X)}{Q(X)} \right)^a - 1}{a-1} \right] \right) \\ &= \frac{1}{a-1} \log(1 + (a-1) D_{H_a}(P \parallel Q)) \end{aligned} \tag{18}$$

Notably, the continuation of Rényi divergence for $a = 1$ also equals the KL-divergence [7, p. 116]. Renyi divergence can be expressed as an invertible transformation of the Hellinger divergence (D_{H_a}, see Equation 18) [24].

Definition 19 (Rényi-information [7])**.** *Rényi-information is defined equivalent to f-information as shown in Equation 19 and corresponds to an invertible transformation of Hellinger-information (I_{H_a}).*

$$\begin{aligned} I_{R_a}(S;T) &:= R_a\left(P_{(S;T)} \parallel P_S \otimes P_T \right) \\ &= \frac{1}{a-1} \log(1 + (a-1) I_{H_a}(S;T)) \end{aligned} \tag{19}$$

Finally, we consider the Bhattacharyya distance (Definition 20), which is equivalent to a linear scaling from a special case of Rényi divergence (Equation 20) [24]. It is applied, among others, in signal processing [25] and coding theory [26]. The corresponding information measure (Equation 21) is like its distance the scaling of a special case of Rényi-information.

Definition 20 (Bhattacharyya distance [27])**.** *Let P and Q be two discrete probability distributions over the event space $\mathcal{X}$, then the Bhattacharyya distance is defined as shown in Equation 20.*

$$\begin{aligned} B(P \parallel Q) &:= -\log\left(\sum_{x \in \mathcal{X}} \sqrt{P(x)Q(x)} \right) \\ &= -\log\left(\sum_{x \in \mathcal{X}} Q(x) \sqrt{\frac{P(x)}{Q(x)}} \right) \\ &= -\log\left(1 - 0.5 \cdot \mathbb{E}_Q\left[\frac{\left(\frac{P(X)}{Q(X)} \right)^{0.5} - 1}{0.5 - 1} \right] \right) \\ &= -\log(1 - 0.5 \cdot D_{H_{0.5}}(P \parallel Q)) \\ &= 0.5 \cdot R_{0.5}(P \parallel Q) \end{aligned} \tag{20}$$

 doi:10.20944/preprints202403.0285.v1

Definition 21 (Bhattacharyya-information)**.** *Bhattacharyya-information is defined equivalent to f-information as shown in Equation 21.*

$$I_B(S;T) \coloneqq B\Big(P_{(S,T)} \parallel P_S \otimes P_T\Big) = 0.5 \cdot I_{R_{0.5}}(S;T) \tag{21}$$

3. Decomposition Methodology

To construct a partial information decomposition in the framework of Williams and Beer [1], we only have to define a cumulative redundancy measure ($I_\cap$) or cumulative loss measure ($I_\cup$). However, doing this requires a meaningful definition of when information is the *same*. Therefore, Section 3.1 presents an interpretation of f-information that enables a representation of distinct information. Specifically, we demonstrate that f-information quantifies the expected perimeter for the channel zonogons of each state $t \in \mathcal{T}$. This allows for the interpretation that each distinct (TPR,FPR)-pair for predicting a state of the target variable provides a distinct notion of uncertainty. Information corresponds accordingly to reachable decision regions, which are quantified by their perimeter. This interpretation of f-information is used in Section 3.2 to construct a partial information decomposition under the zonogon order for each state $t \in \mathcal{T}$ individually. These individual decompositions are then combined into the final result. Therefore, we decompose specific information based on the Blackwell order rather than using its minimum, like Williams and Beer [1]. We use the resulting decomposition of any f-information in Section 3.3 to transform a Hellinger-information decomposition into a Rényi-information decomposition while maintaining its non-negativity and an inclusion-exclusion relation. In Sections 3.2 and 3.3, we first demonstrate the decomposition on the synergy lattice and then its corresponding dual decomposition on the redundancy lattice. To achieve the desired axioms and properties, we combine different aspects of the existing literature:

- Like Bertschinger et al. [9] and Kolchinsky [14] we base the decomposition on the Blackwell order and use this to obtain the operational interpretation of the decomposition.
- Like Williams and Beer [1] and related to Lizier et al. [16], Finn and Lizier [11], and Ince [12], we perform a decomposition from a pointwise perspective but only for the target variable.
- In a similar manner to how Finn and Lizier [11] used probability mass exclusion to differentiate distinct information, we use the achievable decision regions for each state of a target variable to differentiate distinct information.
- We propose applying the concepts about lattice re-graduations discussed by Knuth [17] to PIDs to transform the decomposition of one information measure to another while maintaining its consistency.

We extend Axiom 3 of Williams and Beer [1] as shown below, to allow binding any information measure to the decomposition.

Axiom 3* (Self-redundancy)**.** *For a single source, redundancy $I_{\cap,*}$ and information loss $I_{\cup,*}$ correspond to information measure I_* as shown below:*

$$I_{\cap,*}(\{\mathbf{S}_i\};T) = I_*(\mathbf{S}_i;T) \quad \textit{and} \quad I_{\cup,*}(\{\mathbf{S}_i\};T) = I_*(\mathbf{V};T) - I_*(\mathbf{S}_i;T) \tag{22}$$

3.1. Representing f-Information

We begin with an interpretation of f-information, for which we define a pointwise variable $\pi(T,t)$ that represents one state of the target variable (Equation 23a) and construct its pointwise information channel (Definition 22). Then, we define a function r_f based on the generator function of an f-divergence. This function acts like a pseudo-distance for measuring half the length of the zonogon perimeters for each pointwise information channel (see Figure 2). These zonogon perimeter lengths are pointwise f-information.

 doi:10.20944/preprints202403.0285.v1

Definition 22 ([Target] pointwise binary input channel)**.** *We define a target pointwise binary input channel* $\kappa(\mathbf{S}, T, t)$ *from one state of the target variable* $t \in \mathcal{T}$ *to an information source* $\mathbf{S}$ *with event space* $\mathcal{S} = \{s_1, \ldots, s_m\}$ *as shown in Equation 23b.*

$$\pi(T, t) \coloneqq \begin{cases} 1 & \text{if } T = t \\ 0 & \text{otherwise} \end{cases} \tag{23a}$$

$$\kappa(\mathbf{S}, T, t) \coloneqq \pi(T, t) \to \mathbf{S} = \begin{bmatrix} p(S = s_1 \mid T = t) & \ldots & p(S = s_m \mid T = t) \\ p(S = s_1 \mid T \neq t) & \ldots & p(S = s_m \mid T \neq t) \end{bmatrix} \tag{23b}$$

Definition 23 ([Target] pointwise f-information)**.**

- *We define a function* r_f *as shown in Equation 24a to quantify a vector, where* $0 \leq p, x, y \leq 1$.
- *We define a target pointwise* f*-information function* i_f*, as shown in Equation 24b, to quantify half the zonogon perimeter for the corresponding pointwise channel* $Z(\kappa(\mathbf{S}, T, t))$.

$$r_f(p, \left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]) \coloneqq (px + (1-p)y) \cdot f\left(\frac{x}{px + (1-p)y}\right) \tag{24a}$$

$$i_f(p, \kappa) \coloneqq \sum_{\vec{v} \in \kappa} r_f(p, \vec{v}) \tag{24b}$$

Theorem 1 (Properties of r_f)**.** *For a constant* $0 \leq p \leq 1$*: (1) the function* $r_f(p, \vec{v})$ *is convex in* $\vec{v}$*, (2) scales linearly in* $\vec{v}$*, (3) satisfies a triangle inequality in* $\vec{v}$*, (4) quantifies any vector of slope one to zero, and (5) quantifies the zero vector to zero.*

Proof.

1. The convexity of $r_f(p, \vec{v})$ in $\vec{v}$ is shown separately in Lemma A1 of Appendix A.
2. That $r_f(p, \ell\vec{v}) = \ell r_f(p, \vec{v})$ scales linearly in $\vec{v}$ can directly be seen from Equation 24a.
3. The triangle inequality of $r_f(p, \vec{v})$ in $\vec{v}$ is shown separately in Corollary A1 of Appendix A.
4. A vector of slope one is quantified to zero $r_f(p, \left[\begin{smallmatrix} \ell \\ \ell \end{smallmatrix}\right]) = \ell \cdot f(1) = 0$, since $f(1) = 0$ is a requirement on the generator function of an f-divergence (Definition 15).
5. The zero vector is quantified to zero $r_f(p, \left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right]) = 0 \cdot f(\frac{0}{0}) = 0$ by the convention of generator functions for an f-divergence (Definition 15).

□

The properties of the function r_f are useful since we can interpret it as a pseudo-distance[1] when quantifying the perimeter length of a channel zonogon $Z(\kappa)$. The function r_f provides the following properties to the pointwise information measure i_f.

Theorem 2 (Properties of i_f)**.** *The pointwise information measure* i_f *(1) maintains the ordering relation of the Blackwell order for binary input channels and (2) is non-negative.*

Proof.

1. That the function r_f maintains the ordering relation of the Blackwell order on binary input channels is shown separately in Lemma A2 of Appendix A (Equation 25a).
2. The bottom element $\perp_{\text{BW}} = \left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right]$ consists of a single vector of slope one, which is quantified to zero by Theorem 1 (Equation 25b). The combination with Equation 25a ensures the non-negativity.

 doi:10.20944/preprints202403.0285.v1

$$\kappa_1 \sqsubseteq \kappa_2 \implies i_f(p, \kappa_1) \leq i_f(p, \kappa_2), \tag{25a}$$

$$i_f(p, \bot_{\mathrm{BW}}) = 0. \tag{25b}$$

□

An f-information corresponds to the expected value of the target pointwise f-information function defined above (Equation 26). As a result, we can interpret f-information as quantifying (half) the expected zonogon perimeter length for the pointwise channels $Z(\kappa(\mathbf{S}, T, t))$, where the function r_f acts as a pseudo-distance.

$$\begin{aligned} I_f(\mathbf{S}; T) &= \sum_{t \in \mathcal{T}} P_T(t) \cdot i_f(P_T(t), \kappa(\mathbf{S}, T, t)) \\ &= \sum_{t \in \mathcal{T}} P_T(t) \cdot \left[\sum_{\vec{v} \in \kappa(\mathbf{S}, T, t)} r_f(P_T(t), \vec{v}) \right] \\ &= \sum_{t \in \mathcal{T}} P_T(t) \cdot \left[\sum_{s \in \mathcal{S}} P_S(s) \cdot f\left(\frac{P_{S|T}(s \mid t)}{P_S(s)} \right) \right] \end{aligned} \tag{26}$$

3.2. Decomposing f-Information

With the representation of Section 3.1 in mind, we can define a non-negative partial information decomposition for a set of visible variables $\mathbf{V} = \{V_1, ..., V_n\}$ about a target variable T for any f-information. The decomposition is performed from a pointwise perspective, which means that we decompose the pointwise measure i_f on the synergy lattice $(\mathcal{A}(\mathbf{V}), \preceq)$ for each $t \in \mathcal{T}$. The pointwise synergy lattices are then combined using a weighted sum to obtain the decomposition of I_f.

We map each atom of the synergy lattice to the join of pointwise channels for its contained sources.

Definition 24 (From atoms to channels)**.** *We define the channel corresponding to an atom $\alpha \in \mathcal{A}(\mathbf{V})$ as shown in Eqation 27.*

$$\kappa_{\sqcup}(\alpha, T, t) \coloneqq \begin{cases} \bot_{BW} & \text{if } \alpha = \top_{SL} = \emptyset \\ \bigsqcup_{\mathbf{S} \in \alpha} \kappa(\mathbf{S}, T, t)) & \text{otherwise} \end{cases} \tag{27}$$

Lemma 1. *For any set of sources $\alpha, \beta \in \mathcal{P}(\mathcal{P}_1(\mathbf{V}))$ and target variable T with state $t \in \mathcal{T}$, the function $\kappa_{\sqcup}$ maintains the ordering of the synergy lattice under the Blackwell order as shown in Equation 28.*

$$\alpha \preceq \beta \implies \kappa_{\sqcup}(\beta, T, t) \sqsubseteq \kappa_{\sqcup}(\alpha, T, t) \tag{28}$$

Lemma 1 is shown separately in Section B.1 of Appendix B. The mapping from Definition 24 provides a lattice that can be quantified using pointwise f-information to construct a cumulative loss measure for its decomposition using the Möbius inverse.

Definition 25 ([Target] pointwise cumulative and partial loss measures)**.** *We define the target pointwise cumulative and partial loss functions as shown in Equation 29a and 29b.*

$$i_{\cup, f}(\alpha, T, t) \coloneqq i_f(P_T(t), \kappa(\mathbf{V}, T, t)) - i_f(P_T(t), \kappa_{\sqcup}(\alpha, T, t)) \tag{29a}$$

$$\Delta i_{\cup, f}(\alpha, T, t) \coloneqq i_{\cup, f}(\alpha, T, t) - \sum_{\beta \in \downarrow \alpha} \Delta i_{\cup, f}(\beta, T, t) \tag{29b}$$

 doi:10.20944/preprints202403.0285.v1

The combined cumulative and partial measures are the expected value of their corresponding pointwise measures. This corresponds to combining the pointwise decomposition lattices by a weighted sum.

Definition 26 (Combined cumulative and partial loss measures)**.** *The cumulative loss measure $I_{\cup,f}$ is defined by Equation 30 and the decomposition result $\Delta I_{\cup,f}$ by Equation 31.*

$$I_{\cup,f}(\alpha;T) \coloneqq \sum_{t\in\mathcal{T}} P_T(t)\cdot i_{\cup,f}(\alpha,T,t) \tag{30}$$

$$\begin{aligned}\Delta I_{\cup,f}(\alpha;T) &\coloneqq \sum_{t\in\mathcal{T}} P_T(t)\cdot \Delta i_{\cup,f}(\alpha,T,t)\\ &= I_{\cup,f}(\alpha;T) - \sum_{\beta\in{\downarrow}\alpha} \Delta I_{\cup,f}(\beta;T)\end{aligned} \tag{31}$$

Theorem 3. *The presented definitions for the pointwise and expected loss measures ($i_{\cup,f}$ and $I_{\cup,f}$) provide a non-negative PID on the synergy lattice with inclusion-exclusion relation that satisfies the Axioms 1, 2 and 3* for any f-information measure.*

Proof.

- **Axiom 1**: The measure $i_{\cup,f}$ (Equation 29a) is invariant to permuting the order of sources in α, since the join operator of the zonogon order ($\bigsqcup_{\mathbf{S}\in\alpha}$) is. Therefore, also $I_{\cup,f}$ satisfies Axiom 1.
- **Axiom 2**: The monotonicity of both $i_{\cup,f}$ and $I_{\cup,f}$ on the synergy lattice is shown separately as Corollary A2 in Appendix B.
- **Axiom 3***: For a single source, $i_{\cup,f}$ equals the pointwise information loss by definition (see Equation 22, 24b and 29a). Therefore, $I_{\cup,f}$ satisfies Axiom 3*.
- **Non-negativity**: The non-negativity of $\Delta i_{\cup,f}$ and $\Delta I_{\cup,f}$ is shown separately as Lemma A7 in Appendix B.

□

Consider the pointwise and combined redundancy measures ($i_{\cap,f}$ and $I_{\cap,f}$) as shown in Equation 32, which are defined through an additional inclusion-exclusion relation on the sources of an atom (pointwise using Equation 32a and 32b or combined using Equation 32c). The partial contributions ($\Delta i_{\cap,f}$ and $\Delta I_{\cap,f}$) are obtained from the Möbius inverse. This defines the corresponding dual decomposition on the redundancy lattice.

Definition 27 (Dual decomposition on the redundancy lattice)**.** *We define the pointwise and cumulative redundancy measure as shown in Equation 32.*

$$i_{\cap,f}(\alpha,T,t) \coloneqq \sum_{\beta\in\mathcal{P}_1(\alpha)} (-1)^{|\beta|-1} i_f(P_T(t),\kappa_{\sqcup}(\beta,T,t)) \tag{32a}$$

$$I_{\cap,f}(\alpha;T) \coloneqq \sum_{t\in\mathcal{T}} P_T(t)\cdot i_{\cap,f}(\alpha,T,t) \tag{32b}$$

$$= I_f(\mathbf{V};T) - \sum_{\beta\in\mathcal{P}_1(\alpha)} (-1)^{|\beta|-1} I_{\cup,f}(\beta;T) \tag{32c}$$

Corollary 1. *The dual decomposition as defined by Equation 32 provides a non-negative PID which satisfies an inclusion-exclusion relation and the axioms of Williams and Beer [1] on the redundancy lattice.*

Proof. The Axioms 1 and 3* are transformed from Theorem 3 by Equation 32c. The non-negativity is obtained from Theorem 3 since the partial contributions are identical between dual decompositions.

 doi:10.20944/preprints202403.0285.v1

The non-negativity ensures monotonicity (Axiom 2) since the cumulative measure $I_{\cap,f}$ is the sum of (non-negative) partial contributions in its down-set due to the Möbius inverse. □

Remark 6. *As discussed before [18], it is possible to further split the redundant component in the decomposition for extracting the pointwise meet under the Blackwell order (zonogon intersection). The second component of redundancy as defined above contains decision regions that are part of the convex hull but not the individual channel zonogons (discussed as shared information in [18]).*

From a pointwise perspective, there always exists a dependency between the sources for which the synergy of this state becomes zero. This dependence corresponds, by definition, to the join of their channels. This is helpful for the operational interpretation in the following paragraph since, individually, each pointwise synergy becomes fully volatile to the dependence between the sources.

Operational interpretation: The decomposition obtains the operational interpretation that if a variable provides pointwise unique information, then there exists a unique decision region for some $t \in \mathcal{T}$ that this variable provides access to. Moreover, if a set of variables provides synergetic information, then a decision region for some $t \in \mathcal{T}$ may become inaccessible if the dependence between the variables changes. Due to the equivalence of the zonogon and Blackwell order for binary input variables, these interpretations can also be transferred to a set of actions $a \in \Omega$ and a *pointwise* reward function $u(a, \pi(T, t))$, which only depends on one state of the target variable $\pi(T, t)$ (see Section 2.1): If a variable provides unique information, then it provides an advantage for some set of actions and pointwise reward function, while synergy indicates that the advantage for some pointwise reward function is based on the dependence between variables.

The implication of the interpretation does not hold in the other direction, which we will also highlight in the example of $I_{\cup,\mathrm{TV}}$ in Section 4.1. Finally, the definition of the Blackwell order through the chaining of channels (Equation 2) highlights its suitability for tracing the flows of information in Markov chains (see Section 4.2).

3.3. Decomposing Rényi-Information

Since Rényi-information is an invertible transformation of Hellinger-information and α-information, we argue that their decompositions should be consistent. We propose to view the decomposition of Rényi-information as a transformation from an f-information and demonstrate the approach by transferring the Hellinger-information decomposition to a Rényi-information decomposition. Then, we demonstrate that the result is invariant to a linear scaling of the considered f-information, such that the transformation from α-information provides identical results. The obtained Rényi-information decomposition is non-negative and satisfies the three axioms proposed by Williams and Beer [1] (see below). However, its inclusion-exclusion relation is based on a transformed addition operator. For transforming the decomposition, we consider Rényi-information to be a re-graduation of Hellinger-information, as shown in Equation 33.

$$v_a(z) \coloneqq \frac{1}{a-1}\log(1+(a-1)z) \tag{33a}$$

$$I_{\mathrm{R}_a}(\mathbf{S};T) = v_a(I_{\mathrm{H}_a}(\mathbf{S};T)) \tag{33b}$$

To maintain consistency when transforming the measure, we also have to transform its operators [17, p. 6 ff.]:

 doi:10.20944/preprints202403.0285.v1

Definition 28 (Addition of Rényi-information)**.** *We define the addition of Rényi-information* $\oplus_a$ *with its corresponding inverse function* $\ominus_a$ *by Equation 34.*

$$x \oplus_a y := v_a(v_a^{-1}(x) + v_a^{-1}(y)) = \frac{\log\left(e^{(a-1)x} + e^{(a-1)y} - 1\right)}{a-1} \tag{34a}$$

$$x \ominus_a y := v_a(v_a^{-1}(x) - v_a^{-1}(y)) = \frac{\log\left(e^{(a-1)x} - e^{(a-1)y} + 1\right)}{a-1} \tag{34b}$$

To transform a decomposition of the synergy lattice, we define the cumulative loss measures as shown in Equation 35 and use the transformed operators when computing the Möbius inverse (Equation 36a) to maintain consistency in the results (Equation 36b).

Definition 29. *The cumulative and partial Rényi-information loss measures are defined as transformations of the cumulative and partial Hellinger-information loss measures, as shown in Equation 35 and 36.*

$$I_{\cup,R_a}(\alpha;T) := v_a(I_{\cup,H_a}(\alpha;T)) \tag{35}$$

$$\Delta I_{\cup,R_a}(\alpha;T) := I_{\cup,R_a}(\alpha;T) \;\ominus_a \sum_{\beta\in\downarrow\alpha} \Delta I_{\cup,R_a}(\beta;T) \qquad \textit{where: } + := \oplus_a \tag{36a}$$

$$= v_a(\Delta I_{\cup,H_a}(\alpha;T)) \tag{36b}$$

Remark 7. *We show in Lemma A8 of Appendix C that re-scaling the original* f*-information does not affect the resulting decomposition or transformed operators. Therefore, transforming a Hellinger-information decomposition or a* α*-information decomposition to a Rényi-information decomposition provides identical results.*

The operational interpretation presented in Section 3.2 is similarly applicable to partial Rényi-information ($\Delta I_{\cup,\mathrm{R}_a}$, Equation 36b), since the function v_a satisfies $v_a(0) = 0$ and $x \leq 0 \implies 0 \leq v_a(x)$.

Theorem 4. *The presented definitions for the cumulative loss measure* $I_{\cup,R_a}$ *provide a non-negative PID on the synergy lattice with inclusion-exclusion relation under the transformed addition (Definition 28) that satisfies the Axioms 1, 2 and 3* for any Rényi-information measure.*

Proof.

- **Axiom 1**: $I_{\cup,\mathrm{R}_a}(\alpha;T)$ is invariant to permuting the order of sources, since $I_{\cup,\mathrm{H}_a}(\mathbf{S};T)$ satisfies Axiom 1 (see Section 3.2).
- **Axiom 2**: $I_{\cup,\mathrm{R}_a}(\alpha;T)$ satisfies monotonicity, since $I_{\cup,\mathrm{H}_a}(\mathbf{S};T)$ satisfies Axiom 2 (see Section 3.2) and the transformation function v_a is monotonically increasing for $a \in (0,1) \cup (1,\infty)$.
- **Axiom 3***: Since $I_{\cup,\mathrm{H}_a}$ satisfies Axiom 3* (see Section 3.2, Equation 33 and 35), $I_{\cup,\mathrm{R}_a}$ satisfies the self-redundancy axiom by definition, however, at a transformed operator: $I_{\cup,\mathrm{R}_a}(\{\mathbf{S}_i\};T) = I_{\mathrm{R}_a}(\{\mathbf{V}\};T) \ominus I_{\mathrm{R}_a}(\{\mathbf{S}_i\};T)$.
- **Non-negativity**: The decomposition of $I_{\cup,\mathrm{R}_a}$ is non negative, since $\Delta I_{\cup,\mathrm{H}_a}$ is non-negative (see Section 3.2), the Möbius inverse is computed with transformed operators (Equation 36b) and the function v_a satisfies $x \leq 0 \implies 0 \leq v_a(x)$.

□

 doi:10.20944/preprints202403.0285.v1

Remark 8. *To obtain an equivalent decomposition of Rényi-information on the redundancy lattice, we can correspondingly transform the dual decomposition from the redundancy lattice of Hellinger-Information as shown in Equation 37. The resulting decomposition will satisfy the non-negativity, axioms of Williams and Beer [1] and an inclusion-exclusion relation under the transformed operators (Definition 28) for the same reasons described above from Corollary 1.*

$$I_{\cap,R_a}(\alpha;T) \coloneqq v_a(I_{\cap,H_a}(\alpha;T)) \tag{37a}$$
$$\Delta I_{\cap,R_a}(\alpha;T) \coloneqq v_a(\Delta I_{\cap,H_a}(\alpha;T)) \tag{37b}$$

Remark 9. *The relation between the redundancy and synergy lattice can be used for the definition of a bivalution [17] in calculations as discussed in [18]. This is also possible for Rényi-information at transformed operators.*

When taking the limit of Rényi-information for $a \to 1$, we obtain mutual information (I_{KL}). Since mutual information is also an f-information, we expect its operators in the Möbius inverse to be addition. This is indeed the case (Equation 38), and the measures will be consistent.

$$\begin{aligned} \lim_{a\to 1} x \oplus_a y &= x + y \\ \lim_{a\to 1} x \ominus_a y &= x - y \end{aligned} \tag{38}$$

Finally, the decomposition of Bhattacharyya-information can be obtained by re-scaling the decomposition of Rényi-information at $a = 0.5$, which causes another transform of the addition operator for the inclusion-exclusion relation.

4. Evaluation

A comparison of the proposed decomposition with other methods of the literature can be found in [18] for mutual information. Therefore, this section first compares different f-information measures at typical decomposition examples and discusses the special case of total variation (TV)-information to explain its distinct behavior. Since we can see larger differences between measures in more complex scenarios, we compare the measures by analyzing the information flows in a Markov chain. We provide the used implementation for both dual decompositions of f-information and the examples used in this work at [28].

4.1. Partial Information Decomposition

4.1.1. Comparison of different f-information measures

We use the examples discussed by Finn and Lizier [11] to compare different f-information decompositions and add a generic example from [18]. All used probability distributions and their abbreviations can be found in Appendix D. We normalize the decomposition results to the f-Entropy of the target variable for the visualization in Figure 5.

 doi:10.20944/preprints202403.0285.v1

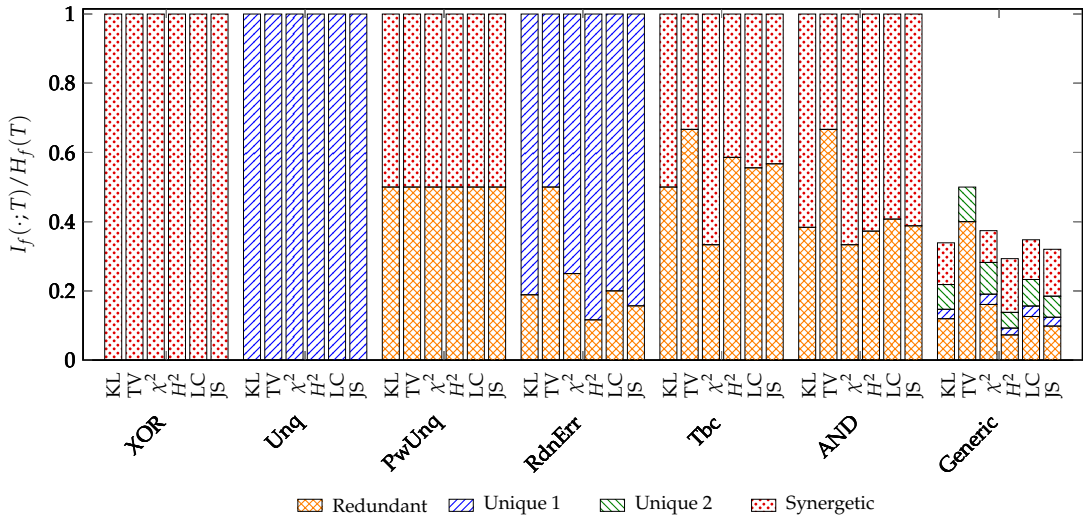


Figure 5. Comparison of different f-information measures normalized to the f-Entropy of the target variable. All distributions are shown in Appendix D and correspond to the examples of [11,18]. The example name abbreviations are listed below Table A1. The measures behave mostly similarly since the decompositions follow an identical structure. However, it can be seen that total variation attributes more information to being redundant than other measures and appears to behave differently in the generic example since it does not attribute any partial information to the first variable or their synergy.

Since all results are based on the same framework, they behave similarly at examples that analyze a specific aspect of the decomposition function (XOR, Unq, PwUnq, RdnErr, Tbc, AND). However, it can be observed that the decomposition of total variation (TV) appears to differ from others: (1) In all examples, total variation attributes more information to being redundant than other measures. (2) In the generic example, total variation is the only measure that does not attribute any information to being unique to variable one or synergetic. We discuss the case of total variation in Section 4.1.2 to explain its distinct behavior.

We visualize the zonogons for the generic example in Figure A2, which shall highlight that the implication of the operational interpretation does not hold in the other direction: the existence of partial information implies an advantage for the expected reward towards some state of the target variable, but an advantage for the expected reward towards some state of the target variable does not imply partial information in the example of total variation.

4.1.2. The special case of total variation

The behavior of total variation appears different compared to other f-information measures (Figure 5). This is due to total variation measuring the perimeter of a zonogon such that the result corresponds to a linear scaling of the maximal (Euclidean) height h^* that the zonogon reaches above the diagonal as visualized in Figure 6:

Lemma 2.

a) The pointwise total variation (i_{TV}) is a linear scaling of the maximal (Euclidean) height h^* that the corresponding zonogon reaches above the diagonal, as visualized in Figure 6 (Equation 39a).
b) For a non-empty set of pointwise channels $\mathbf{A}$, pointwise total variation i_{TV} quantifies the join element to the maximum of its individual channels (Equation 39b).
c) The loss measure $i_{\cup,TV}$ quantifies the meet for a set of sources on the synergy lattice to their minimum (Equation 39c).

 doi:10.20944/preprints202403.0285.v1

$$i_{TV}(p,\kappa) = \frac{1-p}{2}\sum_{v\in\kappa}|v_x - v_y| = (1-p)\frac{h^*}{\sqrt{2}} \tag{39a}$$

$$i_{TV}(p, \bigsqcup_{\kappa\in\mathbf{A}}\kappa) = \max_{\kappa\in\mathbf{A}} i_{TV}(p,\kappa) \tag{39b}$$

$$i_{\cup,TV}(\bigwedge_{\alpha\in\mathbf{A}}\alpha, T, t) = \min_{\alpha\in\mathbf{A}} i_{\cup,TV}(\alpha, T, t) \tag{39c}$$

Proof. The proof of the first two statements (Equation 39b and 39b) is provided separately in Appendix E, which imply the third (Equation 39c) by Definition 25. □

Quantifying the meet element on the synergy lattice to the minimum has the following consequences for total variation: (1) It attributes a minimum amount of synergy, and therefore more information to redundancy than other measures. (2) For each state of the target, at most one variable can provide unique information. In the case of $|\mathcal{T}| = 2$, the pointwise channels are symmetric (see Equation 6), such that the same variable provides the maximal zonogon height both times. This is the case in the generic example of Figure 5, and the reason why at most one variable can provide unique information in this setting. However, beyond binary targets ($|\mathcal{T}| > 2$), both variables may provide unique information at the same time since different sources can provide the maximal zonogon height for different target states (see later example in Figure 7).

Remark 10. *Using the pointwise minimum on the synergy lattice results in a similar structure to the proposed measure of Williams and Beer [1]. However, TV-information is based on a different pointwise measure i_{TV}, which displays the same behavior (Equation 39b), unlike pointwise KL-information.*

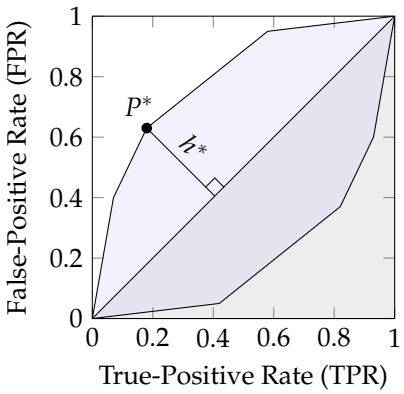


Figure 6. Visualization of the maximal (Euclidean) height h^* at point P^* that a zonogon (blue) reaches above the diagonal.

4.2. Information Flow Analysis

The differences between f-information measures in Section 4.1 appear more visible in complex scenarios. Therefore, this section compares different measures in the information flow analysis of a Markov chain.

Consider a Markov chain $\mathbf{M}_1 \to \mathbf{M}_2 \to \cdots \to \mathbf{M}_5$, where $\mathbf{M}_i = (X_i, Y_i)$ is the joint distribution of two variables. Assume that we are interested in state three and thus define $T = \mathbf{M}_3$ as the target variable. Using the approach described in Section 3, we can compute an information decomposition for each state $\mathbf{M}_i$ of the Markov chain with respect to the target. Now, we are additionally interested in how the partial information decomposition from stage $\mathbf{M}_i$ propagates into the next $\mathbf{M}_{i+1}$, as visualized in Figure 7.

 doi:10.20944/preprints202403.0285.v1

Definition 30 (Partial information flow)**.** *The partial information flow of an atom $\alpha \in \mathcal{A}(\mathbf{M}_i)$ into the atom $\beta \in \mathcal{A}(\mathbf{M}_{i+1})$ quantifies the redundancy between the partial contributions of their respective decomposition lattices.*

Notation 8. *We use the notation $I_{\circ,f}$ with $\circ \in \{\cup, \cap\}$ to refer to either the loss measure $I_{\cup,f}$ or redundancy measure $I_{\cap,f}$. The same applies to the functions $J_{\circ\to\circ,f}$ and $J_{\Delta\to\circ,f}$ of Equation 40.*

Let $\alpha \in \mathcal{A}(\mathbf{M}_i)$ and $\beta \in \mathcal{A}(\mathbf{M}_{i+1})$, then we compute information flows equivalently on the redundancy or synergy lattice as shown in Equation 40. When using a redundancy measure $\circ = \cap$, then the strict down-set of α refers to the strict down-set on its redundancy lattice $(\mathcal{A}(\mathbf{M}_i), \preccurlyeq)$ and when using a loss measure $\circ = \cup$, then the strict down-set refers to the strict down-set on its synergy lattice $(\mathcal{A}(\mathbf{M}_i), \preceq)$. We obtain the intersection of cumulative measures by quantifying their meet, which is on both lattice equivalent to their union of sources ($J_{\circ\to\circ,f}$, Equation 40a). To obtain how much of the partial contribution of α can be found in the cumulative measure of β ($J_{\Delta\to\circ,f}$), we remove the contributions of its down-set ($\dot{\downarrow}\alpha$ on lattice for $\mathcal{A}(\mathbf{M}_i)$, see Equation 40b). To finally obtain the flow from the partial contribution of α to the partial contribution of β ($J_{\Delta\to\Delta,f}$), we similarly remove the contributions of the down-set of β ($\dot{\downarrow}\beta$ on lattice for $\mathcal{A}(\mathbf{M}_{i+1})$, see Equation 40c). The approach can be extended for tracing information flows over multiple steps, however, we will only trace one step in this example.

$$J_{\circ\to\circ,f}(\alpha, \beta, T) \coloneqq I_{\circ,f}(\alpha \cup \beta; T) \tag{40a}$$

$$J_{\Delta\to\circ,f}(\alpha, \beta, T) \coloneqq J_{\circ\to\circ,f}(\alpha, \beta, T) - \sum_{\gamma \in \dot{\downarrow}\alpha} J_{\Delta\to\circ,f}(\gamma, \beta, T) \tag{40b}$$

$$J_{\Delta\to\Delta,f}(\alpha, \beta, T) \coloneqq J_{\Delta\to\circ,f}(\alpha, \beta, T) - \sum_{\gamma \in \dot{\downarrow}\beta} J_{\Delta\to\Delta,f}(\alpha, \gamma, T) \tag{40c}$$

Remark 11. *The resulting partial information flows are equivalent (dual) between the redundancy and loss measure except for the bottom element since their functionality differs: The flow from or to the bottom element on the redundancy lattice is always zero. In contrast, the flow from or to the bottom element on the synergy lattice quantifies the information gained or lost in the step.*

Remark 12. *The information flow analysis of Rényi- and Bhattacharyya-information can be obtained as a transformation of the information flow from Hellinger-information. Alternatively, the information flow can be computed directly using Equation 40 under the corresponding definition of addition and subtraction for the used information measure.*

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 March 2024 doi:10.20944/preprints202403.0285.v1

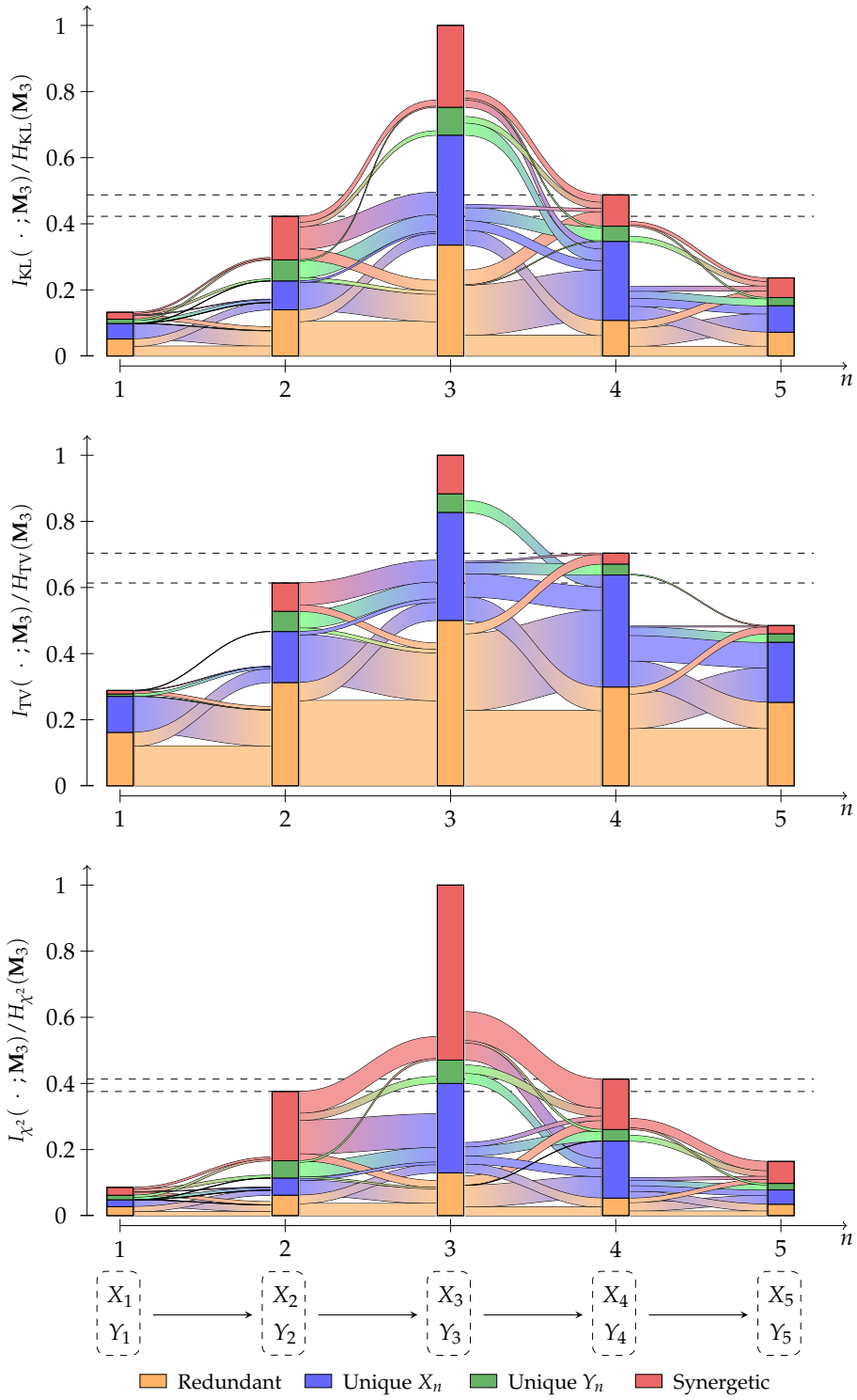


Figure 7. Analysis of the Markov chain information flow (Equation A27). Visualized results for the information measures: KL, TV, and χ^2. The remaining results (H^2-, LC-, and JS-information) can be found in Figure A3.

We randomly generate an initial distribution and each row of a transition matrix under the constraint that at least one value shall be above 0.8 to avoid an information decay that is too rapid through the chain. The specific parameters of the example are shown in Appendix F. The used event spaces are $\mathcal{X} = \{0,1,2\}$ and $\mathcal{Y} = \{0,1\}$ such that $|\mathcal{M}_i| = 6$. We construct a Markov chain of five steps with the target $T = \mathbf{M}_3$ and trace each partial information for one step using Equation 40. We

 doi:10.20944/preprints202403.0285.v1

visualized the results for KL-, TV-, and χ^2-information in Figure 7, and the results for H^2-, LC- and JS-information in Figure A3 of Appendix F.

All results display the expected behavior that the information that $\mathbf{M}_i$ provides about $\mathbf{M}_3$ increases for $1 \leq i \leq 3$ and decreases for $3 \leq i \leq 5$. The information flow results of KL-, H^2-, LC-, and JS-information are conceptually similar. Their main differences appear in the rate at which the information decays and, therefore, how much of the total information we can trace. In contrast, the results of TV- and χ^2-information display different behavior, as shown in Figure 7: TV-information indicates significantly more redundancy, and χ^2-information displays significantly more synergy than the other measures. Additionally, the decomposition of TV-information contains fever information flows. For example, it is the only analysis that does not show any information flow from $\mathbf{M}_2$ into the unique contribution of Y_3 or from $\mathbf{M}_2$ into the synergy of (X_3, Y_3). This demonstrates that the same decomposition method can obtain different behaviors from different f-divergences.

5. Discussion

We perform the decomposition on a pointwise lattice using the zonogon join since it is possible to represent f-information as quantification of the zonogon perimeter on pointwise channels. Correspondingly, if we identified an information measure that was based on quantifying the zonogons of the pointwise channels by their area, then we would need to decompose it on a pointwise lattice using the zonogon meet to achieve non-negativity.

In the literature, PIDs have been defined based on different ordering relations [14]. This diversity is desirable since each approach provides a different operational interpretation of redundancy and synergy. For this reason, we wonder if obtaining a non-negative decomposition with inclusion-exclusion relation for other ordering relations was possible when transferring them to a pointwise perspective or from mutual information to other information measures.

We think studying the relations between different information measures for the same decomposition method may provide further insights into their properties, as demonstrated by the example of total variation in Section 4.2. Similarly, allowing the re-definition of addition for different information measures, as demonstrated in the example of Rényi-information in Section 3.3, opens new possibilities for satisfying the inclusion-exclusion relation and providing consistent decomposition results between related information measures.

Finally, there is currently no universally accepted definition of conditional Rényi information. Assuming that $I_{\mathrm{R}_a}(T; \mathbf{S}_i \mid \mathbf{S}_j)$ should capture the information that $\mathbf{S}_i$ provides about T when already knowing the information from $\mathbf{S}_j$, then one could propose that this quantity should correspond to the according partial information contributions (unique/synergetic) and thus the definition of Equation 41.

$$I_{\mathrm{R}_a}(T; \mathbf{S}_i \mid \mathbf{S}_j) \coloneqq I_{\mathrm{R}_a}(T; \mathbf{S}_i, \mathbf{S}_j) \ominus I_{\mathrm{R}_a}(T; \mathbf{S}_j) \tag{41}$$

6. Conclusions

In this work, we demonstrated a non-negative PID in the framework of Williams and Beer [1] for any f-information with practical operational interpretation. We demonstrated that the decomposition of f-information can be used to obtain a non-negative decomposition of Rényi-information, for which we re-defined the addition to demonstrate that its results satisfy an inclusion-exclusion relation. Finally, we demonstrated how the proposed decomposition method can be used for tracing the flow of information through Markov chains and how the decomposition obtains different properties depending on the chosen information measure.

Author Contributions: Conceptualization, Writing - original draft: T.M.; Analysis of non-negativity: T.M. and E.A.; Writing - review & editing: E.A. and C.R.; Supervision: C.R.

Funding: This research was funded by Swedish Civil Contingencies Agency (MSB) through the project RIOT grant number MSB 2018-12526.

 doi:10.20944/preprints202403.0285.v1

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A. Quantifying zonogon perimeters

Lemma A1. *If the function f is convex, then the function $r_f(p, \vec{v})$ as defined in Equation 24a is convex in its second argument ($\vec{v}$) for a constant $p \in [0,1]$ and $\vec{v} \in [0,1]^2$.*

Proof. We use the following definitions for abbreviating the notation. Let $0 \leq t \leq 1$ and $\vec{v}_i = \left[\begin{smallmatrix} x_i \\ y_i \end{smallmatrix}\right]$:

$$\begin{aligned} a_1 &:= x_1 p + y_1(1-p) \\ a_2 &:= x_2 p + y_2(1-p) \\ b_1 &:= \frac{t a_1}{t a_1 + (1-t) a_2} \\ b_2 &:= \frac{(1-t) a_2}{t a_1 + (1-t) a_2} \end{aligned}$$

The case of $a_i = 0$ is handled by the convention that $0 \cdot f\left(\frac{0}{0}\right) = 0$. Therefore, we can assume that $a_i \neq 0$ and use $0 \leq b_1 \leq 1$ with $b_2 = 1 - b_1$ to apply the definition of convexity on the function f:

$$\begin{aligned} r_f\left(p, \left[\begin{smallmatrix} t x_1 + (1-t) x_2 \\ t y_1 + (1-t) y_2 \end{smallmatrix}\right]\right) &= (t a_1 + (1-t) a_2) \cdot f\left(\frac{t x_1 + (1-t) x_2}{t a_1 + (1-t) a_2}\right) \\ &= (t a_1 + (1-t) a_2) \cdot f\left(b_1 \frac{x_1}{a_1} + b_2 \frac{x_2}{a_2}\right) \\ &\leq (t a_1 + (1-t) a_2) \cdot \left(b_1 f\left(\frac{x_1}{a_1}\right) + b_2 f\left(\frac{x_2}{a_2}\right)\right) \quad \text{(by convexity of } f\text{)} \\ &= t a_1 \cdot f\left(\frac{x_1}{a_1}\right) + (1-t) a_2 \cdot f\left(\frac{x_2}{a_2}\right) \\ &= t \cdot r_f(p, \left[\begin{smallmatrix} x_1 \\ y_1 \end{smallmatrix}\right]) + (1-t) \cdot r_f(p, \left[\begin{smallmatrix} x_2 \\ y_2 \end{smallmatrix}\right]) \end{aligned}$$

□

Corollary A1. *For a constant $p \in [0,1]$ and $\vec{v}_1, \vec{v}_2, (\vec{v}_1 + \vec{v}_2) \in [0,1]^2$, the function $r_f(p, \vec{v})$ as defined in Equation 24a satisfies a triangle inequality on its second argument: $r_f(p, \vec{v}_1 + \vec{v}_2) \leq r_f(p, \vec{v}_1) + r_f(p, \vec{v}_2)$.*

Proof.

$$\begin{aligned} r_f(p, \ell\vec{v}_1 + (1-\ell)\vec{v}_2) &\leq \ell r_f(p, \vec{v}_1) + (1-\ell) r_f(p, \vec{v}_2) && \text{(be Lemma A1)} \\ r_f(p, 0.5(\vec{v}_1 + \vec{v}_2)) &\leq 0.5\Big(r_f(p, \vec{v}_1) + r_f(p, \vec{v}_2)\Big) && \text{(let } \ell = 0.5\text{)} \\ r_f(p, \vec{v}_1 + \vec{v}_2) &\leq r_f(p, \vec{v}_1) + r_f(p, \vec{v}_2) && \text{(by } r_f(p, \ell\vec{v}) = \ell r_f(p, \vec{v})\text{)} \end{aligned}$$

□

Lemma A2. *For a constant $p \in [0,1]$, the function i_f maintains the ordering relation from the Blackwell order on binary input channels: $\kappa_1 \sqsubseteq \kappa_2 \implies i_f(p, \kappa_1) \leq i_f(p, \kappa_2)$.*

Proof. Let κ_1 be represented by a $2 \times n$ matrix and κ_2 by a $2 \times m$ matrix. By the definition of the Blackwell order ($\kappa_1 \sqsubseteq \kappa_2$, Equation 2), there exists a stochastic matrix λ such that $\kappa_1 = \kappa_2 \cdot \lambda$. We use the notation $\kappa_2[:, i]$ to refer to the i^{th} column of matrix κ_2 and indicate the element at row $i \in \{1..m\}$

 doi:10.20944/preprints202403.0285.v1

and column $j \in \{1..n\}$ of λ by $\lambda[i,j]$. Since λ is a valid (row) stochastic matrix of dimension $m \times n$, its rows sum to one $\forall i \in \{1..m\}$. $\sum_{j=1}^{n} \lambda[i,j] = 1$.

$$\begin{aligned}
i_f(p,\kappa_1) &= \sum_{j=1}^{n} r_f(p,\kappa_1[:,j]) && \text{(by Equation 24b)}\\
&= \sum_{j=1}^{n} r_f(p,\sum_{i=1}^{m} \kappa_2[:,i]\lambda[i,j]) && \text{(by Equation 2)}\\
&\leq \sum_{j=1}^{n}\sum_{i=1}^{m} r_f(p,\kappa_2[:,i]\lambda[i,j]) && \text{(by Corollary A1)}\\
&= \sum_{j=1}^{n}\sum_{i=1}^{m} \lambda[i,j] r_f(p,\kappa_2[:,i]) && \text{(by } r_f(p,\ell\vec{v}) = \ell r_f(p,\vec{v})\text{)}\\
&= \sum_{i=1}^{m} r_f(p,\kappa_2[:,i]) && \text{(by } \sum_{j=1}^{n} \lambda[i,j] = 1\text{)}\\
&= i_f(p,\kappa_2) && \text{(by Equation 24b)}
\end{aligned}$$

□

Lemma A3. *Consider two non-empty sets of binary input channels with equal cardinality ($|\mathbf{A}| = |\mathbf{B}|$) and a constant $p \in [0,1]$. If the Minkowski sum for the zonogons of channels in $\mathbf{A}$ is a subset of the Minkowski sum for the zonogons of channels in $\mathbf{B}$, then the sum of pointwise information for the channels in $\mathbf{A}$ is less than the sum of pointwise information for the channels in $\mathbf{B}$ as shown in Equation A1.*

$$\sum_{\kappa\in\mathbf{A}} Z(\kappa) \subseteq \sum_{\kappa\in\mathbf{B}} Z(\kappa) \implies \sum_{\kappa\in\mathbf{A}} i_f(p,\kappa) \leq \sum_{\kappa\in\mathbf{B}} i_f(p,\kappa) \tag{A1}$$

Proof. Let $n = |\mathbf{A}| = |\mathbf{B}|$. We use the notation $\mathbf{A}[i]$ with $1 \leq i \leq n$ to indicate the channel κ_i within the set $\mathbf{A}$.

$$\begin{aligned}
\sum_{i=1}^{n} Z(\mathbf{A}[i]) &\subseteq \sum_{i=1}^{n} Z(\mathbf{B}[i])\\
Z\left(\begin{bmatrix}\mathbf{A}[1] & \dots & \mathbf{A}[n]\end{bmatrix}\right) &\subseteq Z\left(\begin{bmatrix}\mathbf{B}[1] & \dots & \mathbf{B}[n]\end{bmatrix}\right) && \text{(by Equation 4)}\\
Z\left(\frac{1}{n}\cdot\begin{bmatrix}\mathbf{A}[1] & \dots & \mathbf{A}[n]\end{bmatrix}\right) &\subseteq Z\left(\frac{1}{n}\cdot\begin{bmatrix}\mathbf{B}[1] & \dots & \mathbf{B}[n]\end{bmatrix}\right) && \text{(scale to sum } (1,1)\text{)}\\
i_f\left(p,\frac{1}{n}\cdot\begin{bmatrix}\mathbf{A}[1] & \dots & \mathbf{A}[n]\end{bmatrix}\right) &\leq i_f\left(p,\frac{1}{n}\cdot\begin{bmatrix}\mathbf{B}[1] & \dots & \mathbf{B}[n]\end{bmatrix}\right) && \text{(by Eq. 5, Lem. A2)}\\
\sum_{i=1}^{n} i_f\left(p,\frac{1}{n}\mathbf{A}[i]\right) &\leq \sum_{i=1}^{n} i_f\left(p,\frac{1}{n}\mathbf{B}[i]\right) && \text{(by Equation 24b)}\\
\frac{1}{n}\sum_{i=1}^{n} i_f(p,\mathbf{A}[i]) &\leq \frac{1}{n}\sum_{i=1}^{n} i_f(p,\mathbf{B}[i]) && \text{(by } r_f(p,\ell\vec{v}) = \ell r_f(p,\vec{v})\text{)}\\
\sum_{\kappa\in\mathbf{A}} i_f(p,\kappa) &\leq \sum_{\kappa\in\mathbf{B}} i_f(p,\kappa)
\end{aligned}$$

□

Appendix B. The non-negativity of partial f-information

The proof of non-negativity can be divided into three parts. First, we show that the loss measure maintains the ordering relation of the synergy lattice and how the quantification of a meet element

 doi:10.20944/preprints202403.0285.v1

$i_{\cup,f}(\alpha \wedge \beta, T, t)$ can be computed. Second, we demonstrate the construction of a bijective mapping between all subsets of even and odd cardinality that maintains a required subset relation for any selection function. Finally, we combine these two results to demonstrate that an inclusion-exclusion relation using the convex hull of zonogons is greater than their intersection and obtain the non-negativity of the decomposition by transitivity.

Appendix B.1. Properties of the loss measure on the synergy lattice

We require some of the following properties to hold for any set of sources $\alpha \in \mathcal{P}(\mathcal{P}_1(\mathbf{V}))$. Therefore, we define an equivalence relation from the ordering of the synergy lattice ($\cong$) as shown in Equation A2.

Notation A1 (Equivalence under synergy-ordering)**.** *We use the notation $\alpha \cong \beta$ for the equivalence of two sets of sources $\alpha, \beta \in \mathcal{P}(\mathcal{P}_1(\mathbf{V}))$ on the synergy lattice.*

$$(\alpha \cong \beta) \iff (\alpha \preceq \beta \text{ and } \beta \preceq \alpha) \tag{A2}$$

Lemma A4. *Any set of sources $\alpha \in \mathcal{P}(\mathcal{P}_1(\mathbf{V}))$ is equivalent ($\cong$) to some atom of the synergy lattice $\gamma \in \mathcal{A}(\mathbf{V})$.*

$$\forall \alpha \in \mathcal{P}(\mathcal{P}_1(\mathbf{V})).\ \exists \gamma \in \mathcal{A}(\mathbf{V}).\ \gamma \cong \alpha$$

The union for two sets of sources is equivalent to the meet of their corresponding atoms on the synergy lattice. Let $\alpha, \beta \in \mathcal{P}(\mathcal{P}_1(\mathbf{V}))$ and $\gamma, \delta \in \mathcal{A}(\mathbf{V})$:

$$\gamma \cong \alpha \text{ and } \delta \cong \beta \implies (\gamma \wedge \delta) \cong (\alpha \cup \beta)$$

Proof. The used filter in the definition of an atom ($\mathcal{A}(\mathbf{V}) \subseteq \mathcal{P}(\mathcal{P}_1(\mathbf{V}))$, Equation 8) only removes sets of cardinality $2 \leq |\alpha|$ and for any removed set of sources, we can construct an equivalent set which contains one less source by removing the subset $\mathbf{S}_a \subset \mathbf{S}_b$ as shown in Equation A3a. Therefore, all sets of sources $\alpha \in \mathcal{P}(\mathcal{P}_1(\mathbf{V}))$ are equivalent to some atom $\gamma \in \mathcal{A}(\mathbf{V})$ within the lattice (Equation A3b).

$$\mathbf{S}_a \subset \mathbf{S}_b \implies \alpha \cong (\alpha \setminus \mathbf{S}_a) \qquad \text{where: } \mathbf{S}_a, \mathbf{S}_b \in \alpha \tag{A3a}$$

$$\forall \alpha \in \mathcal{P}(\mathcal{P}_1(\mathbf{V})), \exists \gamma \in \mathcal{A}(\mathbf{V}).\ \alpha \cong \gamma \tag{A3b}$$

The union of two sets of sources $\alpha \in \mathcal{P}(\mathcal{P}_1(\mathbf{V}))$ is inferior to each individual set α and β:

$$(\alpha \cup \beta) \preceq \alpha \qquad \text{(by Equation 10)}$$

$$(\alpha \cup \beta) \preceq \beta \qquad \text{(by Equation 10)}$$

All sets of sources $\varepsilon \in \mathcal{P}(\mathcal{P}_1(\mathbf{V}))$ that are inferior of both α and β ($\varepsilon \preceq \alpha$ and $\varepsilon \preceq \beta$) are also inferior to their union.

$$\varepsilon \preceq \alpha \text{ and } \varepsilon \preceq \beta \implies \varepsilon \preceq (\alpha \cup \beta) \qquad \text{(by Equation 10)}$$

Therefore, the union of α and β is equivalent to the meet of their corresponding atoms on the synergy lattice. □

Proof of Lemma 1 from Section 3.2:
For any set of sources $\alpha, \beta \in \mathcal{P}(\mathcal{P}_1(\mathbf{V}))$ and target variable T with state $t \in \mathcal{T}$, the function $\kappa_\sqcup$ (Equation 27) maintains the ordering from the synergy lattice under the Blackwell order.

$$\alpha \preceq \beta \implies \kappa_\sqcup(\beta, T, t) \sqsubseteq \kappa_\sqcup(\alpha, T, t) \tag{A4}$$

 doi:10.20944/preprints202403.0285.v1

Proof. We consider two cases for β:

1. If $\beta = \emptyset$, then the implication holds for any α since the bottom element $\kappa_{\sqcup}(\emptyset, T, t) = \bot_{\mathrm{BW}}$ is inferior ($\sqsubseteq$) to any other channel.
2. If $\beta \neq \emptyset$, then α is also a non-empty set since $\alpha \preceq \beta \prec \top_{\mathrm{SL}} = \emptyset$.

$$
\begin{aligned}
& & \alpha &\preceq \beta & & \\
\forall \mathbf{S}_b \in \beta, \exists \mathbf{S}_a \in \alpha. & & \mathbf{S}_b &\subseteq \mathbf{S}_a & & \text{(by Equation 10)} \\
\forall \mathbf{S}_b \in \beta, \exists \mathbf{S}_a \in \alpha. & & \kappa(\mathbf{S}_b, T, t) &\sqsubseteq \kappa(\mathbf{S}_a, T, t) & & \text{(by Equation 2)} \\
& & \bigsqcup_{\mathbf{S}_b \in \beta} \kappa(\mathbf{S}_b, T, t) &\sqsubseteq \bigsqcup_{\mathbf{S}_a \in \alpha} \kappa(\mathbf{S}_a, T, t) & & \\
& & \kappa_{\sqcup}(\beta, T, t) &\sqsubseteq \kappa_{\sqcup}(\alpha, T, t) & &
\end{aligned}
$$

Since the implication holds for both cases, the ordering is maintained. □

Corollary A2. *The defined cumulative loss measures ($i_{\cup,f}$ of Equation 29a and $I_{\cup,f}$ of Equation 30) maintain the ordering relation of the synergy lattice for any set of sources $\alpha, \beta \in \mathcal{P}(\mathcal{P}_1(\mathbf{V}))$ and target variable T with state $t \in \mathcal{T}$:*

$$
\begin{aligned}
\alpha \preceq \beta &\Longrightarrow i_{\cup,f}(\alpha, T, t) \leq i_{\cup,f}(\beta, T, t) \\
\alpha \preceq \beta &\Longrightarrow I_{\cup,f}(\alpha; T) \leq I_{\cup,f}(\beta; T)
\end{aligned}
$$

Proof. The pointwise monotonicity of the cumulative loss measure ($\alpha \preceq \beta \Longrightarrow i_{\cup,f}(\alpha, T, t) \leq i_{\cup,f}(\beta, T, t)$) is obtained from Lemma 1 and A2 with Equation 29a. Sine all cumulative pointwise losses $i_{\cup,f}$ are smaller for α than β, so will be their weighted sum ($\alpha \preceq \beta \Longrightarrow I_{\cup,f}(\alpha; T) \leq I_{\cup,f}(\beta; T)$, see Equation 30). □

Corollary A3. *The cumulative pointwise loss of the meet from two atoms is equivalent to the cumulative pointwise loss of their union for any target variable T with state $t \in \mathcal{T}$:*
$i_{\cup,f}(\alpha \wedge \beta, T, t) = i_{\cup,f}(\alpha \cup \beta, T, t)$.

Proof. The result follows from Lemma A4 and Corollary A2. □

Appendix B.2. Mapping subsets of even and odd cardinality

Let $\mathcal{P}(\mathbf{A})$ represent the power-set of a non-empty set $\mathbf{A} \neq \emptyset$ and separate the subsets of even ($\mathcal{L}_e$) and odd ($\mathcal{L}_o$) cardinality as shown below. Additionally, let $\mathcal{L}_{\leq 1}$ represent all subsets with cardinality less or equal to one and $\mathcal{L}_1$ all subsets of cardinality equal to one:

$$
\begin{aligned}
\mathcal{L}_{\leq 1} &:= \{\mathbf{B} \in \mathcal{P}(\mathbf{A}) \ : \ |\mathbf{B}| \leq 1\} \\
\mathcal{L}_1 &:= \{\mathbf{B} \in \mathcal{P}(\mathbf{A}) \ : \ |\mathbf{B}| = 1\} \\
\mathcal{L}_e &:= \{\mathbf{B} \in \mathcal{P}(\mathbf{A}) \ : \ |\mathbf{B}| \text{ even}\} \\
\mathcal{L}_o &:= \{\mathbf{B} \in \mathcal{P}(\mathbf{A}) \ : \ |\mathbf{B}| \text{ odd}\} \\
\mathcal{P}(\mathbf{A}) = \mathcal{L}_e \cup \mathcal{L}_o \ &\text{ and } \ \emptyset = \mathcal{L}_e \cap \mathcal{L}_o
\end{aligned}
\tag{A5}
$$

The number of subsets with even cardinality is equal to the number of subsets with odd cardinality as shown in Equation A6.

$$
|\mathcal{L}_e| = \sum_{i=0}^{\left\lfloor \frac{|\mathbf{A}|}{2} \right\rfloor} \binom{|\mathbf{A}|}{2i} = 2^{|\mathbf{A}|-1} = \sum_{i=0}^{\left\lfloor \frac{|\mathbf{A}|}{2} \right\rfloor} \binom{|\mathbf{A}|}{2i+1} = |\mathcal{L}_o| \tag{A6}
$$

 doi:10.20944/preprints202403.0285.v1

Consider a function $g_e : \mathcal{L}_e \to \mathcal{L}_{\leq 1}$, which takes an even subset $\mathbf{E} \in \mathcal{L}_e$ and returns a subset of cardinality $|g_e(\mathbf{E})| = \min(|\mathbf{E}|, 1)$ according to Equation A7.

$$\forall \mathbf{E} \in \mathbf{S}_e \ : \ \begin{cases} g_e(\mathbf{E}) = \emptyset & \text{if } \mathbf{E} = \emptyset \\ g_e(\mathbf{E}) = \{e\} \quad \text{s.t. } e \in \mathbf{E} & \text{otherwise} \end{cases} \tag{A7}$$

Lemma A5. *For any function $g_e \in \mathcal{G}_e$, there exists a function $G : (\mathcal{L}_e, \mathcal{G}_e) \to \mathcal{L}_o$ which satisfies the following two properties:*

a) For any subset with even cardinality, the function $g_e(\cdot)$ returns a subset of function $G(\cdot)$:

$$\forall g_e \in \mathcal{G}_e, \mathbf{E} \in \mathcal{L}_e \ : \ g_e(\mathbf{E}) \subseteq G(\mathbf{E}, g_e). \tag{A8}$$

b) The function $G(\cdot)$ which satisfies Equation A8 has an inverse on its first argument $G^{-1} : (\mathcal{L}_o, \mathcal{G}_e) \to \mathcal{L}_e$.

$$\forall g_e \in \mathcal{G}_e, \mathbf{E} \in \mathcal{L}_e, \exists G^{-1} : \ G^{-1}(G(\mathbf{E}, g_e), g_e) = \mathbf{E}. \tag{A9}$$

Proof. We construct a function G for an arbitrary g_e and demonstrate that it satisfies both properties (Equation A8 and A9) by induction on the cardinality of $\mathbf{A}$. We indicate the cardinality of $\mathbf{A}$ with $n = |\mathbf{A}|$ as subscript $\mathbf{A}_n$, $\mathcal{L}_{e,n}$, $\mathcal{L}_{o,n}$ and G_n:

1. At the base case $\mathbf{A}_1 = \{a\}$, the sets of subsets are $\mathcal{L}_{e,1} = \{\emptyset\}$ and $\mathcal{L}_{o,1} = \{\{a\}\}$. We define the function $G_1(\emptyset, g_e) := \{a\}$ for any g_e to satisfy both required properties:
   a) The constraints of Equation A7 ensures that $g_e(\emptyset) = \emptyset$. Since the empty set is the only element in $\mathbf{S}_{e,1}$, the subset relation (requirement of Equation A8) is satisfied $g_e(\emptyset) = \emptyset \subseteq \{a\} = G_1(\emptyset, g_e)$.
   b) The function $G_1 : (\mathcal{L}_{e,1}, \mathcal{G}_e) \to \mathcal{L}_{o,1}$ is a bijection from $\mathcal{L}_{e,1}$ to $\mathcal{L}_{o,1}$ and therefore has an inverse on its first argument $G_1^{-1} : (\mathcal{L}_{o,1}, \mathcal{G}_e) \to \mathcal{L}_{e,1}$ (requirement of Equation A9).
2. Assume there exists a function G_n, which satisfies both required properties (Equation A8 and A9) at sets of cardinality $1 \leq n = |\mathbf{A}_n|$.
3. For the induction step, we show the definition of a function G_{n+1} that satisfies both required properties. For sets $\mathbf{A}_{n+1} = \mathbf{A}_n \cup \{q\}$, the subsets of even and odd cardinality can be expanded as shown in Equation A10.

$$\begin{aligned} \mathcal{L}_{e,n+1} &= \mathcal{L}_{e,n} \ \cup \ \{\mathbf{O} \cup \{q\} \ : \ \mathbf{O} \in \mathcal{L}_{o,n}\}, \\ \mathcal{L}_{o,n+1} &= \mathcal{L}_{o,n} \ \cup \ \{\mathbf{E} \cup \{q\} \ : \ \mathbf{E} \in \mathcal{L}_{e,n}\}. \end{aligned} \tag{A10}$$

   We define G_{n+1} for $\mathbf{E} \in \mathcal{L}_{e,n}$ and $\mathbf{O} \in \mathcal{L}_{o,n}$ at any g_e as shown in Equation A11 using the function G_n and its inverse G_n^{-1} from the induction hypothesis. The function G_{n+1} is defined for any subset in $\mathcal{L}_{e,n+1}$ as it can be seen from Equation A10.

$$\begin{aligned} G_{n+1}(\mathbf{E}, g_e) &:= \begin{cases} \mathbf{E} \cup \{q\} & \text{if } g_e(G_n(\mathbf{E}, g_e) \cup \{q\}) \neq \{q\} \\ G_n(\mathbf{E}, g_e) & \text{if } g_e(G_n(\mathbf{E}, g_e) \cup \{q\}) = \{q\} \end{cases} \\ G_{n+1}(\mathbf{O} \cup \{q\}, g_e) &:= \begin{cases} \mathbf{O} & \text{if } g_e(\mathbf{O} \cup \{q\}) \neq \{q\} \\ G_n^{-1}(\mathbf{O}, g_e) \cup \{q\} & \text{if } g_e(\mathbf{O} \cup \{q\}) = \{q\} \end{cases} \end{aligned} \tag{A11}$$

   Figure A1 provides an intuition for the definition of G_{n+1}: the outcome of $g_e(\mathbf{O} \cup \{q\})$ determines, if the function G_{n+1} maintains or breaks the mapping of G_n.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 March 2024 doi:10.20944/preprints202403.0285.v1

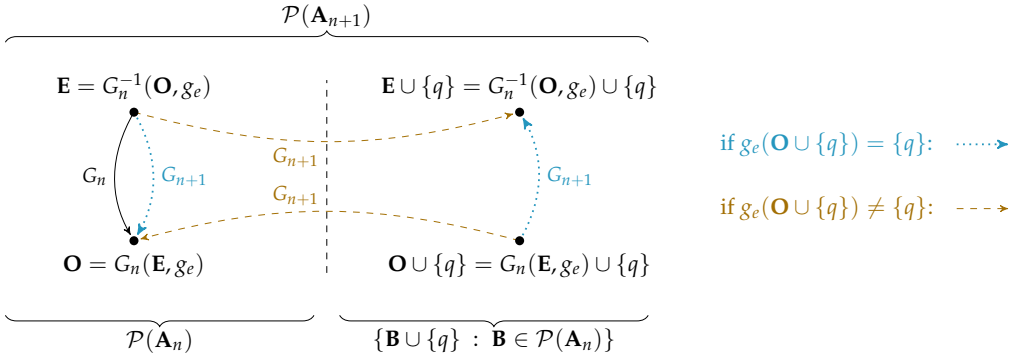


Figure A1. Intuition for the definition of Equation A11. We can divide the set $\mathcal{P}(\mathbf{A}_{n+1})$ into $\mathcal{P}(\mathbf{A}_n)$ and $\{\mathbf{B} \cup \{q\} \ : \ \mathbf{B} \in \mathcal{P}(\mathbf{A}_n)\}$. The definition of function G_{n+1} mirrors G_n if $g_e(\mathbf{O} \cup \{q\}) = \{q\}$ (blue) and otherwise breaks its mapping (orange).

The function F as defined in Equation A11 satisfies both requirements (Equation A8 and A9) for any g_e:

a) To demonstrate that the function satisfies the subset relation of Equation A8, we analyze the four cases for the return value of G_{n+1} as defined in Equation A11 individually:
   - $g_e(\mathbf{E}) \subseteq \mathbf{E} \cup \{q\}$ holds, since the function g_e always returns a subset of its input (Equation A7).
   - $g_e(\mathbf{E}) \subseteq G_n(\mathbf{E}, g_e)$ holds by the induction hypothesis.
   - if $g_e(\mathbf{O} \cup \{q\}) \neq \{q\}$ then $g_e(\mathbf{O} \cup \{q\}) \subseteq \mathbf{O}$: Since the input to function g_e is not the empty set, the function $g_e(\mathbf{O} \cup \{q\})$ returns a singleton subset of its input (Equation A7). If the element in the singleton subset is unequal to q, then it is a subset of $\mathbf{O}$.
   - if $g_e(\mathbf{O} \cup \{q\}) = \{q\}$ then $g_e(\mathbf{O} \cup \{q\}) \subseteq \{q\} \cup G_n^{-1}(\mathbf{O}, g_e)$ holds trivially.

b) To demonstrate that the function G_{n+1} has an inverse (Equation A9), we show that the function G_{n+1} is a bijection from $\mathcal{L}_{e,n+1}$ to $\mathcal{L}_{o,n+1}$. Since the function G_{n+1} is defined for all elements in $\mathcal{L}_{e,n+1}$ and both sets have the same cardinality ($|\mathcal{L}_{e,n+1}| = |\mathcal{L}_{o,n+1}|$, Equation A6), it is sufficient to show that the function G_{n+1} is distinct for all inputs.

   The return value of G_{n+1} has four cases, two of which return a set containing q (case 1 and 4 in Equation A11), while the two others do not (case 2 and 3 in Equation A11). Therefore, we have to show that both of these cases cannot coincide for any input:
   - Case 2 and 3 in Equation A11: If the return value of both cases was equal, then $\mathbf{O} = G_n(\mathbf{E}, g_e)$ and therefore $g_e(\mathbf{O} \cup \{q\}) = g_e(G_n(\mathbf{E}, g_e) \cup \{q\})$. This leads to a contradiction, since the condition of case 3 ensures $g_e(\mathbf{O} \cup \{q\}) \neq \{q\}$, while the condition of case 2 ensures $g_e(G_n(\mathbf{E}, g_e) \cup \{q\}) = \{q\}$. Hence, the return values of case 2 and 3 are distinct.
   - Case 1 and 4 in Equation A11: If the return value of both cases was equal, then $\mathbf{E} = G_n^{-1}(\mathbf{O}, g_e)$ and therefore $g_e(\mathbf{O} \cup \{q\}) = g_e(G_n(\mathbf{E}, g_e) \cup \{q\})$. This leads to a contradiction, since the condition of case 4 ensures $g_e(\mathbf{O} \cup \{q\}) = \{q\}$, while the condition of case 1 ensures $g_e(G_n(\mathbf{E}, g_e) \cup \{q\}) \neq \{q\}$. Hence, the return values of case 1 and 4 are distinct.

   Since the function G_{n+1} is a bijection, there exists an inverse G_{n+1}^{-1}.

□

Appendix B.3. The non-negativity of the decomposition

Lemma A6. *Consider a non-empty set of of binary input channel* $\mathbf{A} \neq \emptyset$ *and* $0 \leq p \leq 1$. *Quantifying an inclusion-exclusion principle on the pointwise information of their Blackwell join is larger than the pointwise information of their Blackwell meet as shown in Equation A12.*

$$i_f\left(p, \prod_{\kappa \in \mathbf{A}} \kappa\right) \leq \sum_{\emptyset \neq \mathbf{B} \subseteq \mathbf{A}} (-1)^{|\mathbf{B}|-1} i_f\left(p, \bigsqcup_{\kappa \in \mathbf{B}} \kappa\right) \tag{A12}$$

 doi:10.20944/preprints202403.0285.v1

Proof. Consider a function $g_o : \mathcal{L}_o \to \mathcal{L}_1$, where $g_o(\mathbf{O}) \subseteq \mathbf{O}$ such that the function returns a singleton subset for a set of odd cardinality. Equation A13 can be obtained from the constraints on g_e (Equation A7) and Lemma A5.

$$\forall g_e \in \mathcal{G}_e, \mathbf{E} \in \mathcal{L}_e, \exists g_o \in \mathcal{G}_o, G \; : \; \begin{cases} g_e(\emptyset) \subseteq g_o(G(\emptyset)) & \text{if } \mathbf{E} = \emptyset \\ g_e(\mathbf{E}) = g_o(G(\mathbf{E})) & \text{otherwise} \end{cases} \tag{A13}$$

Equation A14a holds since we can replace $g_e(\emptyset)$ with $g_o(G(\emptyset))$, meaning there exists a $\kappa \in \mathbf{A}$ for creating a (Minkowski) sum over the same set of channel zonogons on both sides of the quality. Equation A14b holds since Lemma A5 ensured that the existing function G is a bijection. Equation A14c holds since the intersection is a subset of each individual zonogon.

$$\forall g_e \in \mathcal{G}_e, \exists g_o \in \mathcal{G}_o, \kappa \in \mathbf{A}, G : \quad Z(\kappa) + \sum_{\mathbf{E} \in \mathcal{L}_e \setminus \emptyset} Z(g_e(\mathbf{E})) = \sum_{\mathbf{E} \in \mathcal{L}_e} Z(g_o(G(\mathbf{E}))) \tag{A14a}$$

$$\forall g_e \in \mathcal{G}_e, \exists g_o \in \mathcal{G}_o, \kappa \in \mathbf{A} : \quad Z(\kappa) + \sum_{\mathbf{E} \in \mathcal{L}_e \setminus \emptyset} Z(g_e(\mathbf{E})) = \sum_{\mathbf{O} \in \mathcal{L}_o} Z(g_o(\mathbf{O})) \tag{A14b}$$

$$\forall g_e \in \mathcal{G}_e, \exists g_o \in \mathcal{G}_o : \quad \bigcap_{\kappa \in \mathbf{A}} Z(\kappa) + \sum_{\mathbf{E} \in \mathcal{L}_e \setminus \emptyset} Z(g_e(\mathbf{E})) \subseteq \sum_{\mathbf{O} \in \mathcal{L}_o} Z(g_o(\mathbf{O})) \tag{A14c}$$

Equation A14c is parameterized by g_e and subsets are closed under set union. Therefore, we can combine all choices for g_e and g_o using the set-theoretic union as shown below. For the notation, let $m = 2^{|\mathbf{A}|-1}$ and we indicate subsets of $\mathbf{A}$ with even cardinality as $\mathbf{E}_i \in \mathcal{L}_e$, where $1 \leq i \leq m$. We use the last index for the empty set $\mathbf{E}_m = \emptyset$. The subsets of $\mathbf{A}$ with odd cardinality are correspondingly noted as $\mathbf{O}_i \in \mathcal{L}_o$. For clarity, we note binary input channels from an even subset as $\lambda \in \mathbf{E}$ and binary input channels from an odd subset as $\nu \in \mathbf{O}$.

 doi:10.20944/preprints202403.0285.v1

$$\bigcup_{\substack{\lambda_1\in\mathbf{E}_1\\ \lambda_2\in\mathbf{E}_2\\ \dots\\ \lambda_{m-1}\in\mathbf{E}_{m-1}}} \left(\bigcap_{\kappa\in\mathbf{A}} Z(\kappa) + \sum_{i=1}^{m-1} Z(\lambda_i)\right) \subseteq \bigcup_{\substack{\nu_1\in\mathbf{O}_1\\ \nu_2\in\mathbf{O}_2\\ \dots\\ \nu_m\in\mathbf{O}_m}} \left(\sum_{j=1}^{m} Z(\nu_j)\right)$$

$$\bigcap_{\kappa\in\mathbf{A}} Z(\kappa) + \sum_{i=1}^{m-1} \bigcup_{\lambda\in\mathbf{E}_i} Z(\lambda) \subseteq \sum_{j=1}^{m} \bigcup_{\nu\in\mathbf{O}_j} Z(\nu) \qquad \left(\begin{array}{c}\text{Minkowski sum dis-}\\ \text{tributes over set union}\end{array}\right)$$

$$\mathrm{Conv}\left(\bigcap_{\kappa\in\mathbf{A}} Z(\kappa) + \sum_{i=1}^{m-1} \bigcup_{\lambda\in\mathbf{E}_i} Z(\lambda)\right) \subseteq \mathrm{Conv}\left(\sum_{j=1}^{m} \bigcup_{\nu\in\mathbf{O}_j} Z(\nu)\right) \qquad \left(\begin{array}{c}\text{if } \mathbf{X}\subseteq\mathbf{Y} \text{ then}\\ \mathrm{Conv}(\mathbf{X})\subseteq\mathrm{Conv}(\mathbf{Y})\end{array}\right)$$

$$\bigcap_{\kappa\in\mathbf{A}} Z(\kappa) + \sum_{i=1}^{m-1} \mathrm{Conv}\left(\bigcup_{\lambda\in\mathbf{E}_i} Z(\lambda)\right) \subseteq \sum_{j=1}^{m} \mathrm{Conv}\left(\bigcup_{\nu\in\mathbf{O}_j} Z(\nu)\right) \qquad \left(\begin{array}{c}\text{Convex hull distributes}\\ \text{over Minkowski sum}\end{array}\right)$$

$$Z\left(\bigsqcap_{\kappa\in\mathbf{A}} \kappa\right) + \sum_{i=1}^{m-1} Z\left(\bigsqcup_{\lambda\in\mathbf{E}_i} \lambda\right) \subseteq \sum_{j=1}^{m} Z\left(\bigsqcup_{\nu\in\mathbf{O}_j} \nu\right) \qquad \text{(by Equation 7)}$$

$$i_f\left(p, \bigsqcap_{\kappa\in\mathbf{A}} \kappa\right) + \sum_{i=1}^{m-1} i_f\left(p, \bigsqcup_{\lambda\in\mathbf{E}_i} \lambda\right) \leq \sum_{j=1}^{m} i_f\left(p, \bigsqcup_{\nu\in\mathbf{O}_j} \nu\right) \qquad \text{(by Lemma A3)}$$

$$i_f\left(p, \bigsqcap_{\kappa\in\mathbf{A}} \kappa\right) + \sum_{\substack{\emptyset\neq\mathbf{B}\subseteq\mathbf{A}\\ |\mathbf{B}| \text{ even}}} i_f\left(p, \bigsqcup_{\kappa\in\mathbf{B}} \kappa\right) \leq \sum_{\substack{\emptyset\neq\mathbf{B}\subseteq\mathbf{A}\\ |\mathbf{B}| \text{ odd}}} i_f\left(p, \bigsqcup_{\kappa\in\mathbf{B}} \kappa\right) \qquad \text{(replace notation)}$$

$$i_f\left(p, \bigsqcap_{\kappa\in\mathbf{A}} \kappa\right) \leq \sum_{\emptyset\neq\mathbf{B}\subseteq\mathbf{A}} (-1)^{|\mathbf{B}|-1} i_f\left(p, \bigsqcup_{\kappa\in\mathbf{B}} \kappa\right)$$

□

Lemma A7. *The decomposition of f-information is non-negative on the pointwise and combined synergy lattice for any target variable T with state $t \in \mathcal{T}$:*

$$\forall \alpha \in \mathcal{A}(\mathbf{V}).\, 0 \leq \Delta i_{\cup,f}(\alpha, T, t),$$
$$\forall \alpha \in \mathcal{A}(\mathbf{V}).\, 0 \leq \Delta I_{\cup,f}(\alpha; T).$$

Proof. We show the non-negativity of pointwise partial information ($\Delta i_{\cup,f}(\alpha, T, t)$) in two cases . We write α^- to represent the cover set of α on the synergy lattice and use $p = P_T(t)$ as abbreviation:

1. Let $\alpha = \bot_{\mathrm{SL}} = \{\mathbf{V}\}$. The bottom element of the synergy lattice is quantified to zero (by Equation 29a, $i_{\cup,f}(\bot_{\mathrm{SL}}, T, t) = 0$) and therefore also its partial contribution will be zero ($\Delta i_{\cup,f}(\bot_{\mathrm{SL}}, T, t) = 0$), which implies Equation A15.

$$\alpha = \bot_{\mathrm{SL}} \Longrightarrow 0 \leq \Delta i_{\cup,f}(\alpha, T, t) \tag{A15}$$

2. Let $\alpha \in \mathcal{A}(\mathbf{V}) \setminus \{\bot_{\mathrm{SL}}\}$, then its cover set is non-empty ($\alpha^- \neq \emptyset$). Additionally, we know that no atom in the cover set is the empty set ($\forall \beta \in \alpha^-.\, \beta \neq \emptyset$), since the empty atom is the top element ($\top_{\mathrm{SL}} = \emptyset$).

 doi:10.20944/preprints202403.0285.v1

Since it will be required later, note that the inclusion-exclusion principle of a constant is the constant itself as shown in Equation A16 since without the empty set there exists one more subset of odd cardinality than with even cardinality (see Equation A6).

$$i_f(p, \kappa(\mathbf{V}, T, t)) = \sum_{\emptyset \neq \mathbf{B} \subseteq \alpha^-} (-1)^{|\mathbf{B}|-1} i_f(p, \kappa(\mathbf{V}, T, t)) \tag{A16}$$

We can re-write the Möbious inverse as shown in Equation A17.

$$\Delta i_{\cup,f}(\alpha, T, t) = i_{\cup,f}(\alpha, T, t) - \sum_{\beta \in \downarrow \alpha} \Delta i_{\cup,f}(\beta, T, t) \qquad \text{(by Equation 29b)}$$

$$\tag{A17a}$$

$$= i_{\cup,f}(\alpha, T, t) - \sum_{\emptyset \neq \mathbf{B} \subseteq \alpha^-} (-1)^{|\mathbf{B}|-1} \cdot i_{\cup,f}\left(\bigwedge_{\beta \in \mathbf{B}} \beta, T, t \right) \qquad \text{(by [21, p. 15])}$$

$$\tag{A17b}$$

$$= i_{\cup,f}(\alpha, T, t) - \sum_{\emptyset \neq \mathbf{B} \subseteq \alpha^-} (-1)^{|\mathbf{B}|-1} \cdot i_{\cup,f}\left(\bigcup_{\beta \in \mathbf{B}} \beta, T, t \right) \qquad \text{(by Corollary A3)}$$

$$\tag{A17c}$$

$$= - i_f(p, \kappa_\sqcup(\alpha, T, t)) + \sum_{\emptyset \neq \mathbf{B} \subseteq \alpha^-} (-1)^{|\mathbf{B}|-1} \cdot i_f(p, \kappa_\sqcup(\bigcup_{\beta \in \mathbf{B}} \beta, T, t)) \qquad \text{(by Eq. 29a, A16)}$$

$$\tag{A17d}$$

$$= - i_f(p, \kappa_\sqcup(\alpha, T, t)) + \sum_{\emptyset \neq \mathbf{B} \subseteq \alpha^-} (-1)^{|\mathbf{B}|-1} \cdot i_f(p, \bigsqcup_{\mathbf{S} \in (\bigcup_{\beta \in \mathbf{B}} \beta)} \kappa(\mathbf{S}, T, t)) \qquad (\text{by } \forall \beta \in \alpha^-. \beta \neq \emptyset)$$

$$\tag{A17e}$$

$$= - i_f(p, \kappa_\sqcup(\alpha, T, t)) + \sum_{\emptyset \neq \mathbf{B} \subseteq \alpha^-} (-1)^{|\mathbf{B}|-1} \cdot i_f(p, \bigsqcup_{\beta \in \mathbf{B}} \bigsqcup_{\mathbf{S} \in \beta} \kappa(\mathbf{S}, T, t)) \tag{A17f}$$

$$= - i_f(p, \kappa_\sqcup(\alpha, T, t)) + \sum_{\emptyset \neq \mathbf{B} \subseteq \{\kappa_\sqcup(\beta, T, t) \,:\, \beta \in \alpha^-\}} (-1)^{|\mathbf{B}|-1} \cdot i_f(p, \bigsqcup_{\kappa \in \mathbf{B}} \kappa) \tag{A17g}$$

Consider the non-empty set of channels $\mathbf{D} = \{\kappa_\sqcup(\beta, T, t) \,:\, \beta \in \alpha^-\}$, then we obtain Equation A18b from Lemma A6.

$$i_f\left(p, \prod_{\kappa \in \{\kappa_\sqcup(\beta,T,t) \,:\, \beta \in \alpha^-\}} \kappa\right) \leq \sum_{\emptyset \neq \mathbf{B} \subseteq \{\kappa_\sqcup(\beta,T,t) \,:\, \beta \in \alpha^-\}} (-1)^{|\mathbf{B}|-1} i_f\left(p, \bigsqcup_{\kappa \in \mathbf{B}} \kappa\right) \tag{A18a}$$

$$i_f\left(p, \prod_{\beta \in \alpha^-} \kappa_\sqcup(\beta, T, t)\right) \leq \sum_{\emptyset \neq \mathbf{B} \subseteq \{\kappa_\sqcup(\beta,T,t) \,:\, \beta \in \alpha^-\}} (-1)^{|\mathbf{B}|-1} i_f\left(p, \bigsqcup_{\kappa \in \mathbf{B}} \kappa\right) \tag{A18b}$$

We can construct an upper bound on $i_f(p, \kappa_\sqcup(\alpha, T, t))$ based on the cover set α^- as shown in Equation A19.

$$\forall \beta \in \alpha^-. \qquad \beta \preceq \alpha \tag{A19a}$$

$$\forall \beta \in \alpha^-. \qquad \kappa_\sqcup(\alpha, T, t) \sqsubseteq \kappa_\sqcup(\beta, T, t) \qquad \text{(by Lemma 1)} \tag{A19b}$$

$$\kappa_\sqcup(\alpha, T, t) \sqsubseteq \prod_{\beta \in \alpha^-} \kappa_\sqcup(\beta, T, t) \tag{A19c}$$

$$i_f(p, \kappa_\sqcup(\alpha, T, t)) \leq i_f\left(p, \prod_{\beta \in \alpha^-} \kappa_\sqcup(\beta, T, t)\right) \qquad \text{(by Lemma A2)} \tag{A19d}$$

 doi:10.20944/preprints202403.0285.v1

By transitivity of Equation A18b and A19d, we obtain Equation A20.

$$i_f(p, \kappa_{\sqcup}(\alpha, T, t)) \leq \sum_{\emptyset \neq \mathbf{B} \subseteq \{\kappa_{\sqcup}(\beta, T, t) \,:\, \beta \in \alpha^-\}} (-1)^{|\mathbf{B}|-1} i_f\left(p, \bigsqcup_{\kappa \in \mathbf{B}} \kappa\right) \tag{A20}$$

By Equation A17 and A20, we obtain the non-negativity of pointwise partial information as shown in Equation A21.

$$\alpha \in \mathcal{A}(\mathbf{V}) \setminus \{\bot_{\mathrm{SL}}\}. \qquad 0 \leq \Delta i_{\cup,f}(\alpha, T, t) \tag{A21}$$

From Equation A15 and A21, we obtain that pointwise partial information is non-negative for all atoms of the lattice:

$$\forall \alpha \in \mathcal{A}(\mathbf{V}).\, 0 \leq \Delta i_{\cup,f}(\alpha, T) \tag{A22}$$

If all pointwise partial components are non-negative, then their expected value will also be non-negative (see Equation 31):

$$\forall \alpha \in \mathcal{A}(\mathbf{V}).\, 0 \leq \Delta I_{\cup,f}(\alpha; T) \tag{A23}$$

□

Appendix C. Scaling f-information does not affect its transformation

Lemma A8. *The linear scaling of an f-information does not affect the transformation result and operator: Consider scaling an f-information measure $I_{a_2}(S;T) = k \cdot I_{a_1}(S;T)$ with $k \in (0, \infty)$, then their decomposition transformation to another measure $I_b(S;T)$ will be equivalent.*

Proof. Based on the definitions of Section 3.2, the loss measures scale linear with the scaling of their f-divergence. Therefore, we obtain two cumulative loss measures, where $I_{\cup,a_1}$ and $I_{\cup,a_2}$ are a linear scaling of each other (Equation A24a). They can be transformed into another measure $I_{\cup,b}$, as shown in Equation A24b.

$$I_{\cup,a_2}(\alpha; T) = k \cdot I_{\cup,a_1}(\alpha; T) \tag{A24a}$$

$$I_{\cup,b}(\alpha; T) = v_1(I_{\cup,a_1}(\alpha; T)) = v_2(I_{\cup,a_2}(\alpha; T)) \tag{A24b}$$

Equation A24b already demonstrates that their transformation results will be equivalent and that $v_1(z) = v_2(k \cdot z)$ and $k \cdot v_1^{-1}(z) = v_2^{-1}(z)$. Therefore, their operators will also be equivalent as shown below:

$$\begin{aligned} x \oplus_2 y &:= v_2\left(v_2^{-1}(x) \pm v_2^{-1}(y)\right) \\ x \oplus_1 y &:= v_1\left(v_1^{-1}(x) \pm v_1^{-1}(y)\right) \\ &= v_2\left(k v_1^{-1}(x) \pm k v_1^{-1}(y)\right) \\ &= v_2\left(v_2^{-1}(x) \pm v_2^{-1}(y)\right) \\ &= x \oplus_2 y \end{aligned}$$

□

Appendix D. Decomposition example distributions

The probability distributions used in Figure 5 can be found in Table A1. For providing an intuition of the decomposition result for $I_{\cup,\mathrm{TV}}$ at the generic example, we visualized its corresponding zonogons in Figure A2. It can be seen that the maximal zonogon height is obtained from V_1 (blue) which equals the maximal zonogon height of their joint distribution (V_1, V_2) (red). Therefore, $I_{\cup,\mathrm{TV}}$ does not attribute partial information uniquely to V_2 or their synergy by Lemma 2.

 doi:10.20944/preprints202403.0285.v1

Table A1. The used distributions from [11] and the generic example from [18]. The example names are abbreviations for: XOR-gate (XOR), Unique (Unq), Pointwise Unique (PwUnq), Redundant-Error (RdnErr), Two-Bit-copy (Tbc) and the AND-gate (AND) [11].

			Probability						
V_1	V_2	T	XOR	Unq	PwUnq	RdnErr	Tbc	AND	Generic
0	0	0	1/4	1/4	0	3/8	1/4	1/4	0.0625
0	0	1	-	-	-	-	-	-	0.3000
0	1	0	-	1/4	1/4	1/8	-	1/4	0.1875
0	1	1	1/4	-	-	-	1/4	-	0.1500
0	2	1	-	-	1/4	-	-	-	-
1	0	0	-	-	1/4	-	-	1/4	0.0375
1	0	1	1/4	1/4	-	1/8	-	-	0.0500
1	0	2	-	-	-	-	1/4	-	-
1	1	0	1/4	-	-	-	-	-	0.2125
1	1	1	-	1/4	-	3/8	-	1/4	-
1	1	3	-	-	-	-	1/4	-	-
2	0	1	-	-	1/4	-	-	-	-

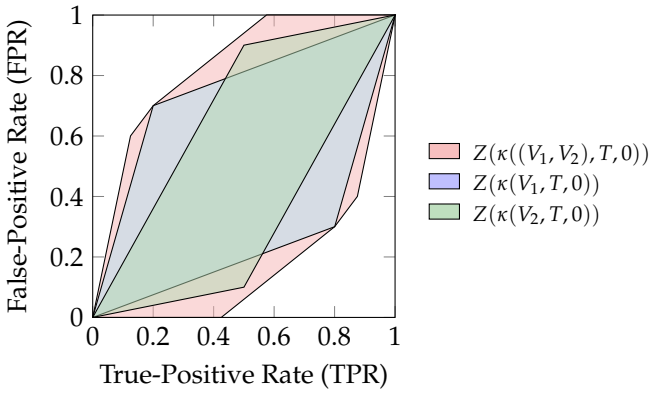


Figure A2. Visualization of the zongons from the generic example of [18] at state $t = 0$. The target variable T has two states. Therefore, the zonogons of its second state are symmetric (second column of Equation 6) and have identical heights.

Appendix E. The relation of total variation to the zonogon height

Proof of Lemma 2 a) from Section 4.1.2:
The pointwise total variation (i_{TV}) is a linear scaling of the maximal (Euclidean) height h^* that the corresponding zonogon $Z(\kappa)$ reaches above the diagonal as visualized in Figure 6 for any $0 \leq p \leq 1$.

$$i_{\mathrm{TV}}(p,\kappa) = \frac{1-p}{2} \sum_{v \in \kappa} |v_x - v_y| = (1-p)\frac{h^*}{\sqrt{2}}$$

Proof. The point of maximal height P^* that a zonogon $Z(\kappa)$ reaches above the diagonal is visualized in Figure 6 and can be obtained as shown in Equation A25, where $\Delta\vec{v}$ represents the slope of vector $\vec{v}$.

$$P^* = \sum_{\vec{v} \in \{\vec{v} \in \kappa \,:\, \Delta\vec{v} > 1\}} \vec{v} \tag{A25}$$

The maximal height (Euclidean distance) above the diagonal is calculated as shown in Equation A26, where $P^* = (P^*_x, P^*_y)$.

$$h^* = \frac{1}{2} \left\| \begin{pmatrix} P^*_x - P^*_y \\ P^*_y - P^*_x \end{pmatrix} \right\|_2 = \sqrt{(P^*_x - P^*_y)^2 + (P^*_y - P^*_x)^2} = \sqrt{2}(P^*_y - P^*_x) \tag{A26}$$

 doi:10.20944/preprints202403.0285.v1

The pointwise total variation i_{TV} can be expressed as invertible transformation of the maximal euclidean zonogon height above the diagonal as shown in Equation E, where $\vec{v} = (\vec{v}_x, \vec{v}_y)$.

$$\begin{aligned}
i_{\mathrm{TV}}(p,\kappa) &= \sum_{\vec{v}\in\kappa} \frac{1}{2}\left|\frac{\vec{v}_x}{p\vec{v}_x + (1-p)\vec{v}_y} - 1\right|(p\vec{v}_x + (1-p)\vec{v}_y) \\
&= \frac{1-p}{2}\sum_{\vec{v}\in\kappa}|\vec{v}_x - \vec{v}_y| \\
&= \frac{1-p}{2}\left(\sum_{\vec{v}\in\{\vec{v}\in\kappa\,:\,\Delta\vec{v}>1\}}(\vec{v}_y - \vec{v}_x) + \sum_{\vec{v}\in\{\vec{v}\in\kappa\,:\,\Delta\vec{v}\leq 1\}}(\vec{v}_x - \vec{v}_y)\right) \\
&= \frac{1-p}{2}\Big((P^*_y - P^*_x) + \big((1-P^*_x) - (1-P^*_y)\big)\Big) && \text{(by Equation A25)} \\
&= (1-p)(P^*_y - P^*_x) \\
&= (1-p)\frac{h^*}{\sqrt{2}} && \text{(by Equation A26)}
\end{aligned}$$

□

Proof of Lemma 2 b) from Section 4.1.2:
For a non-empty set of pointwise channel $\mathbf{A}$ *and* $0 \leq p \leq 1$*, pointwise total variation* i_{TV} *quantifies the join element to the maximum of its individual channels:*

$$i_{\mathrm{TV}}(p, \bigsqcup_{\kappa\in\mathbf{A}} \kappa) = \max_{\kappa\in\mathbf{A}} i_{\mathrm{TV}}(p,\kappa)$$

Proof. The join element $Z(\bigsqcup_{\kappa\in\mathbf{A}}\kappa)$ corresponds to the convex hull of all individual zonogons (see Equation 7). The maximal height that the convex hull reaches above the diagonal is equal to the maximum of the maximal height that each individual zonogon reaches. Since pointwise total variation is a liner scaling of the (Euclidean) zonogon height above the diagonal (Lemma 2 a) shown above), the join element is valuated to the maximum of its individual channels. □

Appendix F. Information Flow example parameters and visualization

The parameters for the Markov chain used in Section 4.2 are shown in Equation A27, where $\mathbf{M}_n = (X_n, Y_n)$, $\mathcal{X}_i = \{0,1,2\}$, $\mathcal{Y}_i = \{0,1\}$, $P_{\mathbf{M}_1}$ is the initial distribution and $P_{\mathbf{M}_{n+1}|\mathbf{M}_n}$ is the transition matrix. The visualized results for the information flow of KL-, TV-, and χ^2-information can be found in Figure 7, and the visualized results of H^2-, LC-, and JS-information in Figure A3.

$$\begin{array}{rcccccc}
\text{States } (X_1, Y_1): & (0,0) & (0,1) & (1,0) & (1,1) & (2,0) & (2,1) \\
P_{\mathbf{M}_1} = & \begin{bmatrix} 0.01 & 0.81 & 0.00 & 0.02 & 0.09 & 0.07 \end{bmatrix}
\end{array} \tag{A27a}$$

$$P_{\mathbf{M}_{n+1}|\mathbf{M}_n} = \begin{bmatrix}
0.05 & 0.01 & 0.04 & 0.82 & 0.02 & 0.06 \\
0.05 & 0.82 & 0.00 & 0.01 & 0.06 & 0.06 \\
0.04 & 0.01 & 0.82 & 0.05 & 0.04 & 0.04 \\
0.03 & 0.84 & 0.02 & 0.06 & 0.04 & 0.01 \\
0.04 & 0.03 & 0.03 & 0.02 & 0.06 & 0.82 \\
0.07 & 0.04 & 0.01 & 0.03 & 0.81 & 0.04
\end{bmatrix} \tag{A27b}$$

 doi:10.20944/preprints202403.0285.v1

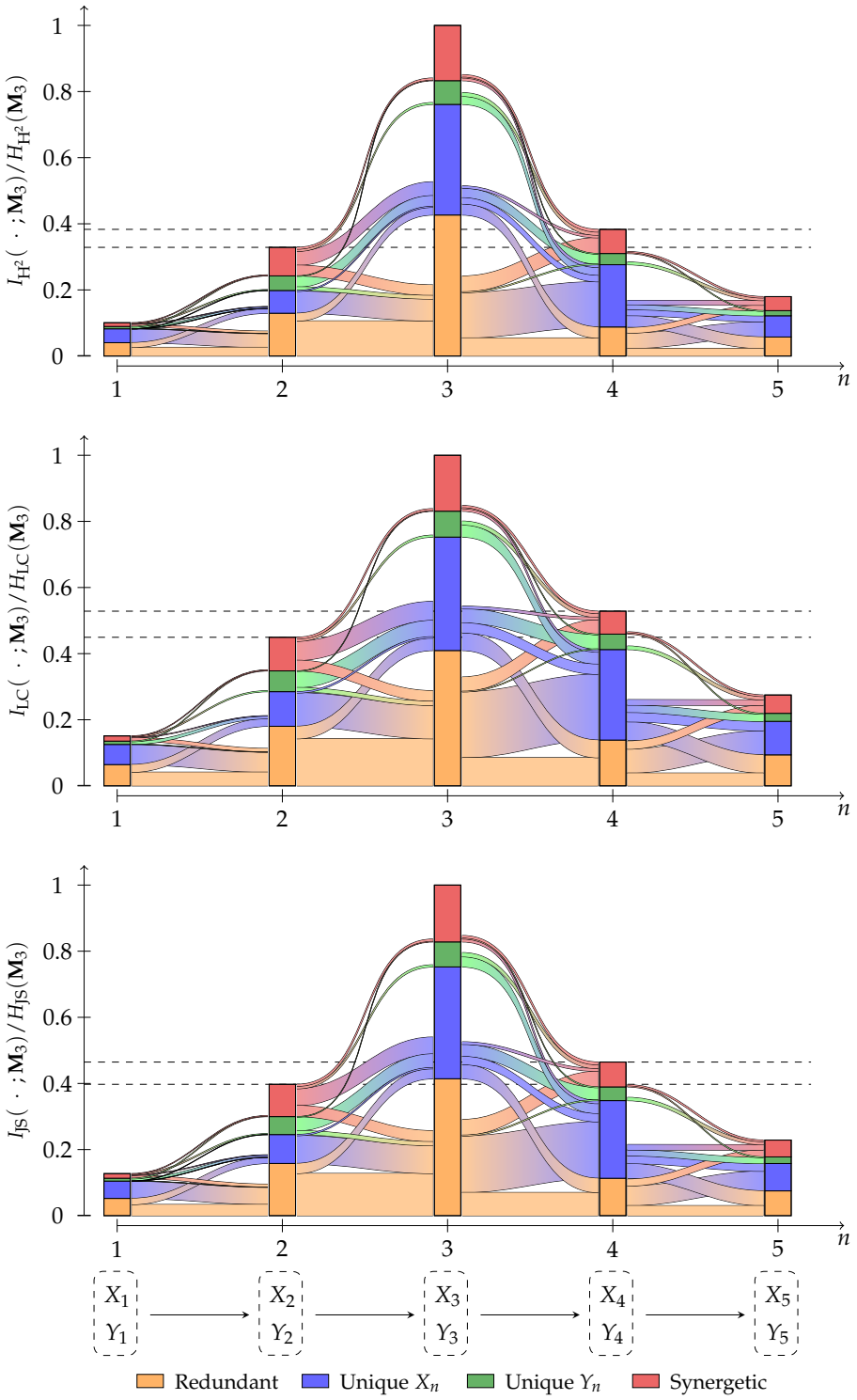


Figure A3. Analysis of the Markov chain information flow (Equation A27). Visualized results for the information measures: H^2, LC, and JS. The remaining results (KL, TV, and χ^2) can be found in Figure 7.

References

1. Williams, P.L.; Beer, R.D. Nonnegative Decomposition of Multivariate Information. arXiv 1004.2515, 2010.
2. Lizier, J.T.; Bertschinger, N.; Jost, J.; Wibral, M. Information Decomposition of Target Effects from Multi-Source Interactions: Perspectives on Previous, Current and Future Work. *Entropy* **2018**, *20*. https://doi.org/10.3390/e20040307.

 doi:10.20944/preprints202403.0285.v1

3. Griffith, V.; Chong, E.K.P.; James, R.G.; Ellison, C.J.; Crutchfield, J.P. Intersection Information Based on Common Randomness. *Entropy* **2014**, *16*, 1985–2000. https://doi.org/10.3390/e16041985.
4. Bertschinger, N.; Rauh, J.; Olbrich, E.; Jost, J. Shared Information—New Insights and Problems in Decomposing Information in Complex Systems. In Proceedings of the Proceedings of the European Conference on Complex Systems 2012; Gilbert, T.; Kirkilionis, M.; Nicolis, G., Eds., Cham, 2013; pp. 251–269.
5. Harder, M.; Salge, C.; Polani, D. Bivariate measure of redundant information. *Phys. Rev. E* **2013**, *87*, 012130. https://doi.org/10.1103/PhysRevE.87.012130.
6. Finn, C. A New Framework for Decomposing Multivariate Information. PhD thesis, University of Sydney, 2019.
7. Polyanskiy, Y.; Wu, Y. Information theory: From coding to learning. *Book draft* **2022**.
8. Mironov, I. Rényi Differential Privacy. In Proceedings of the 2017 IEEE 30th Computer Security Foundations Symposium (CSF), 2017, pp. 263–275. https://doi.org/10.1109/CSF.2017.11.
9. Bertschinger, N.; Rauh, J.; Olbrich, E.; Jost, J.; Ay, N. Quantifying Unique Information. *Entropy* **2014**, *16*, 2161–2183. https://doi.org/10.3390/e16042161.
10. Griffith, V.; Koch, C., Quantifying Synergistic Mutual Information. In *Guided Self-Organization: Inception*; Springer Berlin Heidelberg: Berlin, Heidelberg, 2014; pp. 159–190. https://doi.org/10.1007/978-3-642-53734-9_6.
11. Finn, C.; Lizier, J.T. Pointwise Partial Information Decomposition Using the Specificity and Ambiguity Lattices. *Entropy* **2018**, *20*. https://doi.org/10.3390/e20040297.
12. Ince, R.A.A. Measuring Multivariate Redundant Information with Pointwise Common Change in Surprisal. *Entropy* **2017**, *19*. https://doi.org/10.3390/e19070318.
13. Rosas, F.E.; Mediano, P.A.M.; Rassouli, B.; Barrett, A.B. An operational information decomposition via synergistic disclosure. *Journal of Physics A: Mathematical and Theoretical* **2020**, *53*, 485001. https://doi.org/10.1088/1751-8121/abb723.
14. Kolchinsky, A. A Novel Approach to the Partial Information Decomposition. *Entropy* **2022**, *24*. https://doi.org/10.3390/e24030403.
15. Bertschinger, N.; Rauh, J. The Blackwell relation defines no lattice. In Proceedings of the 2014 IEEE International Symposium on Information Theory, 2014, pp. 2479–2483. https://doi.org/10.1109/ISIT.2014.6875280.
16. Lizier, J.T.; Flecker, B.; Williams, P.L. Towards a synergy-based approach to measuring information modification. In Proceedings of the 2013 IEEE Symposium on Artificial Life (ALife), 2013, pp. 43–51. https://doi.org/10.1109/ALIFE.2013.6602430.
17. Knuth, K.H. Lattices and Their Consistent Quantification. *Annalen der Physik* **2019**, *531*, 1700370.
18. Mages, T.; Rohner, C. Decomposing and Tracing Mutual Information by Quantifying Reachable Decision Regions. *Entropy* **2023**, *25*. https://doi.org/10.3390/e25071014.
19. Blackwell, D. Equivalent comparisons of experiments. *The annals of mathematical statistics* **1953**, pp. 265–272.
20. Neyman, J.; Pearson, E.S. IX. On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* **1933**, *231*, 289–337.
21. Chicharro, D.; Panzeri, S. Synergy and Redundancy in Dual Decompositions of Mutual Information Gain and Information Loss. *Entropy* **2017**, *19*. https://doi.org/10.3390/e19020071.
22. Csiszár, I. On information-type measure of difference of probability distributions and indirect observations. *Studia Sci. Math. Hungar.* **1967**, *2*, 299–318.
23. Rényi, A. On measures of entropy and information. In Proceedings of the Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics. University of California Press, 1961, Vol. 4, pp. 547–562.
24. Sason, I.; Verdú, S. f -Divergence Inequalities. *IEEE Transactions on Information Theory* **2016**, *62*, 5973–6006. https://doi.org/10.1109/TIT.2016.2603151.
25. Kailath, T. The divergence and Bhattacharyya distance measures in signal selection. *IEEE transactions on communication technology* **1967**, *15*, 52–60.
26. Arikan, E. Channel Polarization: A Method for Constructing Capacity-Achieving Codes for Symmetric Binary-Input Memoryless Channels. *IEEE Transactions on Information Theory* **2009**, *55*, 3051–3073. https://doi.org/10.1109/TIT.2009.2021379.

 doi:10.20944/preprints202403.0285.v1

27. Bhattacharyya, A. On a measure of divergence between two statistical populations defined by their probability distribution. *Bulletin of the Calcutta Mathematical Society* **1943**, *35*, 99–110.
28. Mages, T.; Anastasiadi, E.; Rohner, C. Implementation: PID Blackwell specific information. https://github.com/uu-core/pid-blackwell-specific-information, 2024. Accessed on -.