

# The mathematical landscape of partial information decomposition: A comprehensive review of properties and measures

Alberto Liardi<sup>\*1</sup>, Keenan J. A. Down<sup>1,2,3</sup>, George Blackburne<sup>4,1</sup>, Matteo Neri<sup>5</sup>, and  
Pedro A. M. Mediano<sup>1,6</sup>

<sup>1</sup>Department of Computing, Imperial College London, UK

<sup>2</sup>Department of Psychology, University of Cambridge, Cambridge, UK

<sup>3</sup>Department of Psychology, Queen Mary University, London, UK

<sup>4</sup>Department of Experimental Psychology, University College London, London, UK

<sup>5</sup>Institut de Neurosciences de la Timone, Aix Marseille Université, Marseille, France

<sup>6</sup>Division of Psychology and Language Sciences, University College London, UK

## Abstract

Partial Information Decomposition (PID) has become one of the most prominent information-theoretic frameworks for describing the structure and quality of information in complex systems. Despite its widespread utility, there exists no unique solution constraining precisely how a PID should be constructed, leading to a multiverse of different formalisms with different mathematical commitments. In this work, we provide a comprehensive overview of the mathematical landscape of PID. By integrating existing PID measures into a common language, we systematically examine all major approaches to the PID framework that have emerged so far, determining for each measure whether or not each known property holds. In addition, we derive a web of all known theorems mapping the relationships and incompatibilities between these properties, before also revealing some novel interdependency results. In doing so, we chart a brief history of the framework, promote a unified perspective for its discussions, and offer a path towards both theoretical refinement and informed empirical applications for the future of this powerful method.

## 1 Introduction

Classical information theory, unified by Claude Shannon in 1948 [1], is one of the great mathematical triumphs of the 20th century, bringing together notions from communications theory, thermodynamics, and statistical mechanics into one all-encompassing theoretical discipline [2–7]. A key reason for its influence across disciplines and its broad applicability is that information theory provides a precise and remarkably general formalism for describing and capturing all kinds of relationships between random variables. Be they linear, nonlinear, or otherwise difficult to describe, mutual information, for example, is always equal to zero if and only if the variables are entirely independent [8]. That is to say, while correlation can detect linear relationships between variables (or at the very least, relatively monotonic relationships), mutual information and other information-based approaches can detect all possible classes of interaction phenomena.

Despite this, surprising gaps emerge in classical information theory [9, 10], especially in its treatment of higher-order dependencies in systems of many interacting components. Understanding these interdependencies is fundamental to studying the behaviour of complex systems, where precise mechanisms are often difficult to reckon with [11–17]. Fully describing these interactions using Shannon’s theory alone is challenging, since the information-sharing relationships among more than two variables can take qualitatively distinct forms that are not captured by mutual information or other classical measures of interdependence. Suppose, for example, we wish to send an encrypted letter, accompanied by a second letter that explains how to decrypt it. Neither letter alone provides any

---

\*Corresponding author: a.liardi@imperial.ac.uk

information about the intended message, each leaving the recipient uncertain of the intended message. It is only when both letters are combined that the intended information is conveyed: knowledge of both systems is required to describe the content, with the two letters acting *synergistically* to convey the message. Suppose now that we write the same unencrypted message in both letters. In this case, the message would be transmitted *redundantly*, offering the recipient two copies of the intended data. The classical theory, somewhat surprisingly, struggles to differentiate between these two effects: determining whether multiple variables contain repeated or complementary information about a target variable cannot be directly quantified using classical information theory alone.

Partial Information Decomposition (PID) is a method in contemporary information theory which addresses this particular weakness of the classical theory: the inability to separate redundant and synergistic contributions. Introduced in 2010 by Williams and Beer [10], PID aims to disentangle these components by extending the inclusion–exclusion view of information beyond the quantities described by Shannon. By defining an intersection measure  $I_{\cap}(X_1, \dots, X_n; Y)$  that captures how much information *the sources*  $X_1, \dots, X_n$  share redundantly about *a target*  $Y$ , one can derive elementary *atoms* that partition information into qualitatively distinct parts. In doing so, atoms corresponding to redundant, synergistic, and unique contributions to the mutual information emerge naturally from the proposed lattice structure.

Despite the beauty and seminal nature of Williams’ and Beer’s PID framework, the intersection measure they proposed,  $I_{\cap}^{\min}$ , was immediately met with additional scrutiny [18, 19]. Most notably, Harder *et al.* pointed out that for two independent bits  $X_1$  and  $X_2$ , with target  $Y = X_1 X_2$  (a system called the two-bit-copy, TBC),  $I_{\cap}^{\min}$  assigns to the sources  $X_1$  and  $X_2$  one bit of redundant information about  $Y$ , even though they are entirely independent [18]. Bertschinger *et al.* also noted that  $I_{\cap}^{\min}$  has the peculiar property that adding resolution to the target variable (by joining it with another variable, for example), can sometimes reduce the amount of redundancy. If *more* can be known overall about the target variable, it is certainly surprising that the sources somehow share *less* information about the target [19].

For these reasons, while the partial information lattice of Williams and Beer was (rightly) praised, their proposed intersection information measure  $I_{\cap}^{\min}$  was not, leaving a complex vacuum: an algebraic theory without an accompanying measure. As a result, a suite of new properties or “axioms” that a redundancy function should satisfy was proposed shortly after PID’s inception. This, in turn, resulted in the introduction of many different PID measures, each designed to cover properties that the original measure  $I_{\cap}^{\min}$  could not achieve. Since the genesis of the partial information decomposition method, at least 19 different approaches have been proposed, and, at the time of this review, no general consensus on the best approach has emerged [10, 18, 20–41].

Although this process enriched the conceptual landscape, it also led to a proliferation of partially overlapping and competing properties, each reflecting different conceptions of what redundant information is. To complicate the matter, several of these properties were subsequently shown to be mutually incompatible, implying that no single redundancy measure can satisfy them all [42]. Thus, the field now comprises a diverse collection of measures grounded in different, and often irreconcilable, axiomatic choices, contributing to ambiguity in both the theoretical interpretation and empirical application of PID. In this work, we endeavour to alleviate some of these outstanding shortcomings by providing a unified and comprehensive perspective on the state of the field.

## 1.1 Main contributions

In this work, we provide a unified treatment of Partial Information Decomposition (PID), generating the first comprehensive formal resource for its largely decentralised research programme.

Firstly, we present a brief mathematical introduction to those who are unfamiliar with the PID methodology, offering some motivation for the problem itself, and exploring why so much research has been done in this surprisingly challenging direction (Sec. 2). In doing so, we provide an overview of all major properties stated in the literature, along with a brief discussion of their significance and the motivations behind their original conception (Sec. 2.2 and Table 2).

Similarly, we also present a brief historical overview of each of the PID measures (Sec. 2.4), with a longer version reported in the appendix, describing the rationale underpinning their introductions. To make comparisons between measures and properties more straightforward, we provide mathematical definitions of each PID measure and axiom in a standardised language in the appendix.

We then present our first major result: a systematic verification of whether each proposed PID property is

satisfied by each existing PID measure (Table 5). For all results not already known in the literature, we provide either a proof or a counterexample for each combination (available in the appendix). To yield further insights into the latent relationships between PID measures, we hierarchically categorise PID measures based on these axiomatic profiles (Sec. 3.1 and Fig. 1), providing insights into the major philosophical branches describing *redundant* information.

Furthermore, we then collate all known theorems relating PID properties, encompassing both implications and no-go results (Sec. 3.2). In doing so, we present new and old theorems, introduce a hypergraph-based representation of these relationships (Fig. 2), and integrate them into an open-source implementation based on an automatic theorem prover, enabling automated verification of property compatibility.

Finally, we discuss the implications of this work on the Partial Information Decomposition method as a whole (Sec. 4). There, we outline useful practices for empirical implementation of the PID framework, briefly describe alternative approaches to the PID method, and discuss potential gaps in the literature and open avenues for future research.

In short, our contributions are as follows:

- We present an overview of all major PID properties stated in the literature, along with a brief description of their significance and motivation (Sec. 2 and Table 2).
- We provide a systematic verification of whether or not each known property holds for each and every measure (Table 5).
- We collate all known theorems relating PID properties into a common language, including both implications and no-go theorems, provide novel theorems (Sec. 2.2), and link them together in an automatic theorem prover [43].
- We perform a descriptive analysis of the measures and properties, resulting in a hierarchy of clustered measures and a hypergraph description of the theorems (Sec. 3).

## 2 A comprehensive guide to PID properties and their relationships

### 2.1 A partially gentle introduction to Partial Information Decomposition

Given a set of  $n$  sources  $X_1, \dots, X_n$  and a target  $Y$ , partial information decomposition aims to decompose the information that a collection of  $n$  sources  $X_1, \dots, X_n$  provides about a target  $Y$  into multiple qualitatively different parts: redundancy, unique information, and synergy. These denote the information provided about the target by both sources separately, only one source, and the two sources taken together, respectively. In the case of more than two sources, combinations of these three can also appear.

To achieve the decomposition, Williams and Beer build on the inclusion-exclusion principle (IEP) and construct a partially ordered set (poset) of redundancies, i.e. information quantities that represent the information common to a set of sources. The corresponding algebraic object is known as the *redundancy lattice*. Building the redundancy lattice involves considering the set of possible collections of sets of sources  $\mathcal{A}$  where no source is a superset of any other:

$$\mathcal{A}(X_1, \dots, X_n) \equiv \{\alpha \in \mathcal{P}(\mathcal{P}(X_1, \dots, X_n)) \mid \forall A_i, A_j \in \alpha, A_i \not\subseteq A_j\}, \quad (1)$$

where  $\mathcal{P}(\mathcal{X})$  represents the set of all non-empty subsets of  $\mathcal{X}$ . This set can then be equipped with the partial order  $\preceq$ , where  $\alpha \preceq \beta$  for  $\alpha, \beta \in \mathcal{A}$  if, for every source  $B \in \beta$  (consisting of possibly multiple variables), there is at least one source  $A \in \alpha$  such that  $A \subseteq B$ . Put simply, the lattice creates a hierarchy of information based on availability, or the level of knowledge required to access it:  $\alpha$  is “below”  $\beta$  if the information in  $\alpha$  can be fully recovered from any source present in  $\beta$ .

For each collection  $\alpha \in \mathcal{A}$ , we can then define a redundancy function  $I_\cap$ . Consider, for example, a system of three variables  $X_1, X_2, X_3$  providing information about a target  $Y$ . Then  $I_\cap(X_1, X_2 X_3; Y)$  corresponds to the information that can be known about  $Y$  from either  $X_1$  or the joint variable  $X_2 X_3$ .

The redundancy lattice provides insights into how redundant information is distributed across the sources, with higher elements in the lattice providing at least as much redundant information as lower nodes. Following this

setup, one can implicitly define the so-called PID atoms  $I_\partial$ , which signify how much information is gained at each level:

$$I_\cap(\alpha; Y) = \sum_{\beta \prec \alpha} I_\partial(\beta; Y), \quad (2)$$

where  $\alpha$  is a generic collection of sets of sources, and the value of  $I_\partial(\alpha, Y)$  is the amount of information corresponding to the *atom* represented by  $\alpha$ . Using this expression, we can isolate the atoms themselves by applying Möbius inversion to  $I_\cap$  [44, 45], obtaining each atom recursively as:

$$I_\partial(\alpha; Y) = I_\cap(\alpha; Y) - \sum_{\beta \prec \alpha} I_\partial(\beta; Y). \quad (3)$$

Throughout this work, we will, wherever possible, explicitly write the collection of sets of sources that we refer to. For example, the redundancy function of the collection  $\alpha = \{\{X_1\}, \{X_2, X_3\}\}$  will be written as  $I_\cap(X_1, X_2 X_3; Y)$  (with corresponding atom  $I_\partial(X_1, X_2 X_3; Y)$ ), so as to minimise ambiguity.

Unsurprisingly, the number of possible collections of sources  $\alpha$  can become extremely large even for a limited number of sources, growing with the Dedekind numbers [46]. Offering some further constraints on the problem, Williams and Beer introduce several axioms on the redundancy measure  $I_\cap$ . The first of these, (Weak) Symmetry (**S<sub>0</sub>**), follows the intuitive idea that the ordering of sources in  $\alpha$  should not affect the value of redundancy. Additionally, they introduce (Weak) Monotonicity (**M<sub>0</sub>**), which states that adding sources to the redundancy function should not increase  $I_\cap$ , keeping it unchanged if the added source is more informative than another already present. For reasons that will be apparent later, we refer to this last part of the property as Subset Equality (**SE**) [26].

Finally, since this is a decomposition of mutual information, a connection between these quantities and classical information-theoretic measures needs to be established. Hence, Williams and Beer introduced the Self-Redundancy (**SR**) property, imposing that the redundancy function of one source and the target is equal to the mutual information between the two. Thus, not only are (**S<sub>0</sub>**), (**M<sub>0</sub>**), and (**SR**) intuitive, but they also ground the redundancy lattice in a well-behaved framework.

Although some proposed PID measures challenge (**M<sub>0</sub>**), the remaining (**S<sub>0</sub>**), (**SR**), and (**SE**) are core properties of all PID definitions. In fact, as we will see, almost all of the proposed measures possess all three.

An example worth directly referencing is the bivariate case. Given two sources  $X_1, X_2$  and a target  $Y$ , the collection of sets of possible sources is given by  $\mathcal{A}(X_1, X_2) = \{\{\{1\}, \{2\}\}, \{1\}, \{2\}, \{1, 2\}\}$ , for a total of four possible source combinations. Correspondingly, there are four nodes in the redundancy lattice, and we can relate the intersection measure  $I_\cap$  to the individual atoms  $I_\partial$  using Eq. 2 as

$$I_\cap(X_1, X_2; Y) = I_\partial(X_1, X_2; Y) \quad (4)$$

$$I(X_1; Y) = I_\partial(X_1, X_2; Y) + I_\partial(X_1; Y) \quad (5)$$

$$I(X_2; Y) = I_\partial(X_1, X_2; Y) + I_\partial(X_2; Y) \quad (6)$$

$$I(X_1, X_2; Y) = I_\partial(X_1, X_2; Y) + I_\partial(X_1; Y) + I_\partial(X_2; Y) + I_\partial(X_1 X_2; Y), \quad (7)$$

where we employed (**S<sub>0</sub>**), (**M<sub>0</sub>**), and (**SR**) to significantly decrease the number of non-trivial redundancy terms. In the notation above,  $I_\partial(X_1, X_2; Y)$  represents redundancy,  $I_\partial(X_i; Y)$  indicates information unique to  $X_i$ , and  $I_\partial(X_1 X_2; Y)$  is synergy.

As previously mentioned, performing partial information decomposition on a system with  $n \geq 3$  sources introduces additional complexity, as different combinations of redundant, unique, and synergistic effects become possible. For example,  $I_\partial(X_1, X_2 X_3; Y)$  represents the redundant information between source  $\{X_1\}$  and  $\{X_2, X_3\}$ , with the latter term including the synergistic information  $X_2$  and  $X_3$  provide about  $Y$ . For more details on the PID lattice and the notation used in this work, we refer to Appendix A.

The conceptual clarity and mathematical simplicity of this framework has contributed to its widespread popularity, with applications spanning from artificial [26, 47–50] and biological neural networks [12, 15, 51–57], to gene regulatory systems [58, 59], cellular automata [60, 61], and labour market dynamics [62]. However, to uniquely solve the linear system of equations given by Eq. (3), an additional definition of intersection information  $I_\cap$  is required, as it is absent from the classical theory. Since specifying a redundancy function which has sufficiently many desired

properties has been the primary route forward in this area, we now explore these additional constraints. In the next section, we outline the various properties that have been proposed over the years for the measure  $I_{\cap}$ , giving a brief history of their introduction, as well as implications and incompatibilities.

## 2.2 A summary of PID properties

Throughout the development of the partial information decomposition framework, a wide range of PID properties have been suggested. These properties, often referred to as *axioms*, specify desiderata that a reasonable notion of redundancy is argued to satisfy. Generally, they are motivated by intuitive principles that one might expect redundancy to obey, supported by operational interpretations, grounded in empirical considerations, or inspired by the mathematical properties of classical information measures. Unfortunately, it was demonstrated early in the literature that some of these properties are, in fact, mutually inconsistent, so that no single redundancy measure can simultaneously satisfy all of them. As a consequence, prolific discussion about which of these properties should be favoured or abandoned has exploded, with a lack of consensus persisting to the present day.

The first properties to be introduced, Self-Redundancy (**SR**), (Weak) Symmetry (**S<sub>0</sub>**), and (Weak) Monotonicity (**M<sub>0</sub>**), were introduced by Williams and Beer for the construction of an interpretable redundancy lattice [10]. In the same work, the authors also showed that, taken together, these properties guarantee a non-negative redundancy function – a property referred to as Global Positivity, (**GP**). In addition, the authors showed that their proposed PID measure  $I_{\cap}^{\min}$  satisfied a property called Local Positivity, (**LP**), whereby all atoms obtained from the decomposition are themselves non-negative. Since then, (**LP**) has long been a heavily desired property of PID definitions, as non-negative quantities enable a clear interpretation of the resulting PID atoms, e.g. in terms of information transmission, as is possible for standard information-theoretic measures.

More recently, PID measures that reject (**LP**) in favour of a pointwise approach have been proposed [30, 35]. These suggest that negative atoms might have a natural interpretation as providing *misinformation* about the target, similar to how negative values of pointwise mutual information naturally appear in the classical theory. Alternatively, such negative atoms might indicate the presence of the so-called “mechanistic redundancy” [26], i.e. redundancy that arises purely due to the source-target interdependence, independently of the source-source correlation [18]. Within these approaches, the authors also reject the inequality condition of (**M<sub>0</sub>**) while retaining the equality statement, i.e. that  $I_{\cap}$  should remain unchanged if a source more informative than another is considered, thus introducing the Subset Equality property (**SE**) [26], later also referred to as “deterministic equality” [38].

A fundamental property which is often not explicitly stated is Equivalence-class Invariance (**EI**), first formally suggested by Griffith *et al.* [20]. Under (**EI**), redundancy functions are required to be invariant under isomorphisms, i.e. invertible transformations of the random variables. This invariance is a key principle in statistics, as it ensures that statistical quantities depend solely on the statistical structure of the variables and not on semantic details such as their particular labelling or representation. Importantly, this requirement highlights a fundamental distinction between statistical analysis and the study of underlying mechanisms, as it is known that different mechanisms can sometimes provide isomorphic probability distributions [63], therefore being indistinguishable from a purely statistical perspective [9]. For this reason, if (**EI**) is taken to hold, purely causal considerations cannot serve to guide intuitions for the construction or interpretations of PID definitions [30]. In the early works on PID, (**EI**) was majorly adopted implicitly, being a universal and well-established property in statistics. However, its application to specific, seemingly intuitive toy gates quickly contributed to the emergence of profoundly contrasting views of how redundancy should behave.

Along this line, an apparently simple yet profound system that has fundamentally shaped the history of PID is the Two-Bit Copy (TBC) gate. The TBC represents the situation in which the target is a copy of the two sources (Table 1).

Importantly, if we accept that information measures should be invariant under bijective transformations (**EI**), this system can be seen in two ways:

1. the target can be viewed as a multivariate random variable  $Y = (Y_1, Y_2)$ , where  $Y_1 = X_1$  and  $Y_2 = X_2$ ; or
2. the target can be seen as an entirely univariate variable with four states, with outcomes that can be renamed (e.g. to  $\{A, B, C, D\}$ ).

$X_1$	$X_2$	$Y$
0	0	00 $\equiv$ A
0	1	01 $\equiv$ B
1	0	10 $\equiv$ C
1	1	11 $\equiv$ D

**TAB. 1: Truth table for the Two-Bit Copy system (TBC).** The relabelling of the target is ensured by **(EI)**.

In the former instance (1), from mechanistic intuitions, one could expect that redundancy is only present if the two sources are strongly correlated. This is more apparent if we reflect on an empirical scenario, where we consider two time series  $X_1^t, X_2^t$  with high autocorrelation and low cross-correlation. If the past states  $X_1^{t-1}, X_2^{t-1}$  are considered as sources and the joint future state  $(X_1^t, X_2^t)$  as the target, one would expect low redundancy and high unique information, with redundancy increasing with the correlation between the sources. In the limit, one could expect zero redundant information for independent sources. This reasoning has led Harder *et al.* to first argue against  $I_{\square}^{\min}$  as it assesses the same *amount* of information but not the same *kind* of information, and then to introduce the Identity property **(ID)** [18].

Formally described, **(ID)** requires that the redundancy in the TBC should be equal to the mutual information between the sources [18], guaranteeing some continuity between the *sharedness* of mutual information and the *sharedness* of redundancy. However, in the second scenario above (2), it is not entirely clear what about the system should indicate sharedness of information between the sources. With this perspective, interpreting or explaining the meaning of the **(ID)** property is not completely trivial. However, if one additionally assumes **(EI)**, then relabelling outcomes should not affect the statistical structure of the system, and hence the redundancy calculation should not depend on whether one chooses perspective (1) or (2). Therefore, the TBC highlight the divergence between the two approaches even for simple systems, challenging the coexistence of **(EI)** with **(ID)**. Moreover, it can be argued that **(ID)** prevents the characterisation of “synergistic entropy”, the additional uncertainty arising over and above the uncertainty guaranteed if they were independent [26].

To partially alleviate these issues, the Independent Identity property **(IID)** has been introduced as a weaker variant of **(ID)**, prescribing that independent sources should not provide redundancy in the TBC [26]. This property maintains the intuitions developed in the example above, while avoiding the stronger implications of **(ID)** for correlated sources. Since its appearance, **(IID)** has been embraced as a key desideratum by numerous authors in the field [26, 38, 41]. However, even if one accepts **(IID)** and not **(ID)**, it has been argued that a non-vanishing redundancy should still be expected even with uncorrelated sources in the TBC since they can “happen to provide the same [probability mass] exclusions with respect to the [two-event] target distribution” [30]. Following this reasoning, **(ID)** and **(IID)** must be rejected, as accepting **(EI)** naturally implies that any conclusion should not be inferred from causal interpretations, and hence that the intuitions derived from (1) are misleading [30].

On the other hand, a complementary approach advocated by Chicharro is to drop **(EI)** in favour of **(IID)** [64], suggesting that the semantic value is meaningful as it is crucial to assess mechanistic redundancy. As we see later, this difference in stance is one of the major factors that changes the results obtained by the different PID definitions. We formalise the inconsistencies between **(ID)**, **(EI)**, and other properties in the following sections, further discussing the interpretational consequences of adopting a specific axiom.

Following a different line of reasoning, several axioms have been proposed in an effort to mirror properties of classical Shannon information measures, such as Target Monotonicity **(TM)**, Target Chain rule **(TC)** [19, 30, 35], Strong Symmetry **(S<sub>1</sub>)** [19], and Strong Monotonicity **(M<sub>1</sub>)** [20]. It soon became apparent, however, that some of these PID properties are mutually incompatible with other fundamental ones, such as **(LP)** and **(ID)** [42, 65]. Although in this work we restrict our attention to the implications of these properties for redundancy functions, we note that previous studies have analysed what it means for the PID atoms themselves to satisfy the axioms [42, 66, 67]. Along this line, Rauh *et al.* have recently advocated a more focused investigation of additional properties that had previously received little attention: Additivity **(AD)**, Continuity **(CO)**, and Differentiability **(DI)**. These are fundamental mathematical properties satisfied by all classical information measures [68].

Name	Symbol	Description
Self-redundancy	<b>(SR)</b>	$I_\cap$ of a single source equals its mutual information.
(Weak) Symmetry	<b>(S<sub>0</sub>)</b>	$I_\cap$ is invariant under any permutation of the sources.
(Weak) Monotonicity	<b>(M<sub>0</sub>)</b>	Adding a source to $I_\cap$ does not increase redundancy <sup>1</sup> .
Global Positivity	<b>(GP)</b>	$I_\cap$ is non-negative.
(Weak) Local Positivity	<b>(LP<sub>0</sub>)</b>	All $I_\partial$ atoms are non-negative for $n = 2$ sources.
Strong Local Positivity	<b>(LP<sub>1</sub>)</b>	All $I_\partial$ atoms are non-negative for $n \geq 3$ sources.
Local Positivity	<b>(LP)</b>	All $I_\partial$ atoms are non-negative $\forall n$ sources.
Independent Identity	<b>(IID)</b>	In the TBC, independent sources provide vanishing $I_\cap$ <sup>2</sup> .
Identity	<b>(ID)</b>	In the TBC, $I_\cap$ is equal to the mutual information between sources <sup>2</sup> .
Target Monotonicity	<b>(TM)</b>	Adding a target does not decrease the value of $I_\cap$ .
Target Chain rule	<b>(TC)</b>	The chain rule for $I_\cap$ holds for the target variables.
Subset Equality	<b>(SE)</b>	Removing a source more informative than another preserves $I_\cap$ .
Lower Bound	<b>(LB)</b>	A variable less informative than all sources lower-bounds $I_\cap$ .
Target Equality	<b>(TE)</b>	Adding the target to the sources does not change $I_\cap$ .
Strong Monotonicity	<b>(M<sub>1</sub>)</b>	Adding a source to $I_\cap$ does not increase redundancy <sup>1</sup> .
Strong Symmetry	<b>(S<sub>1</sub>)</b>	$I_\cap$ is invariant under any permutation of sources or target.
Assumption (*)	<b>(*)</b>	Redundancy and unique information only depend on the marginal and pairwise distributions between sources and target.
Blackwell Property	<b>(BP)</b>	Unique information vanishes iff one source Blackwell-precedes another.
Additivity	<b>(AD)</b>	$I_\cap$ of two independent subsystems equals the sum of their $I_\cap$ .
Continuity	<b>(CO)</b>	$I_\cap$ is continuous under small changes in the probability distribution.
Differentiability	<b>(DI)</b>	$I_\cap$ is differentiable with respect to the probability distribution.
Equivalence-class Invariance	<b>(EI)</b>	$I_\cap$ is invariant under the equivalence relation $\sim$ (Eq. (11)).

**TAB. 2: Summary of existing PID properties.**

Operational considerations grounded in game theory and decision-making strategies have also contributed to the formulation of PID desiderata. Investigating what it means for an agent to be more informed than another in a decision-theoretic setting led to the introduction of Assumption **(\*)** [21], which suggests that redundancy and unique information should depend on the marginal distributions between sources and targets, as opposed to the full probability distribution of the system.

Since its introduction, **(\*)** has been criticised from several perspectives. Firstly, it has been shown that it can artificially induce correlations between the sources [29]. Moreover, the decision problem that originally prompted the introduction of **(\*)** was later deemed “too restrictive” [26]: when the operational approach was generalised to a broader game-theoretic setting, explicit violations of Assumption **(\*)** were demonstrated [26].

Another central concept in statistical decision theory is the so-called Blackwell order, which provides a partial ordering relation between random variables in terms of their informational usefulness for decision-making problems. Informally, a variable is said to be Blackwell superior to another if it enables at least as good performance in any decision problem. This intuition is formalised by Blackwell’s informativeness theorem [69]. Drawing inspiration from this framework, Blackwell’s theorem was operationalised within PID through the introduction of the Blackwell Property **(BP)**, which states that a source provides zero unique information if and only if it is Blackwell inferior to all other sources [38–40]. While both **(\*)** and **(BP)** arise from operational considerations rooted in decision theory, it is important to emphasise that they are logically independent: neither implies nor is implied by the other, except for specific cases (see Appendix C of Ref. [39] for a detailed discussion).

Finally, a small subset of properties has been proposed primarily on the basis of their intuitive appeal. Target Equality **(TE)** [20, 38] states that adding the target variable itself to the set of sources should not alter the value of the redundancy  $I_\cap$ . This property is closely related to Strong Monotonicity **(M<sub>1</sub>)**: although **(TE)** was introduced

<sup>1</sup>Note that these properties also entail an equality condition, see Appendix for the precise definition.

<sup>2</sup>Note that these properties also hold for more general distributions than the TBC, see Appendix for more details.

as a separate property by Kolchinsky [38], it had already appeared as part of  $(\mathbf{M}_1)$  in the earlier work of Griffith *et al.* [20]. With the aim of constraining the range of admissible redundancy values, the Lower Bound property  $(\mathbf{LB})$  was introduced [20].  $(\mathbf{LB})$  asserts that any random variable that is less informative about the target than all sources should provide a lower bound on the redundancy of the system.

Additional properties have been suggested in the literature. Although we do not consider them in detail in this work, we briefly mention them here. Related to  $(\mathbf{AD})$  and  $(\mathbf{CO})$  is also *superadditivity*, which is often satisfied by all measures containing a min. The *Two-event partition* property was proposed by [30] as a pointwise version of  $(*)$ , for which redundancy and unique information should only depend on the marginal and pairwise pointwise distributions between each source and the target.

In Ref. [64], Chicharro proposed two additional properties: the *weak* and the *strong* axioms on stochasticity, where a generalisation of  $(\mathbf{ID})$  is formalised by setting specific constraints on the synergy atom when deterministic relations between source and targets are present. Interestingly, both conditions are stronger requirements than  $(\mathbf{ID})$ , being both sufficient, but not necessary.

Another property related to the identity axioms is the *combined secret sharing property* suggested by Rauh [70]. At its core, this property captures the intuition that PID should reflect the behaviour of ideal secret sharing schemes, in which groups of sources either fully know a secret or know nothing about it. Accordingly, redundancy is required to match exactly the amount of information that is jointly accessible to all the considered source groups [70]. In its pairwise version, this property implies  $(\mathbf{IID})$ , but not  $(\mathbf{ID})$ .

More recently, Kolchinsky introduced a property called *order equality* [38], which resembles  $(\mathbf{SE})$  but generalises it to a generic ordering of choice, e.g. Blackwell. In the case of a partial information decomposition with  $n = 2$  sources, the order equality property with Blackwell ordering corresponds to the left implication of  $(\mathbf{BP})$ .

In the same work, Kolchinsky also discusses the *inclusion-exclusion principle* (IEP). Although not strictly a PID property per se, the IEP is a fundamental assumption within PID that allows one to express synergistic effects in terms of redundancy effects. Kolchinsky argues that there is no universal reason why this property should hold, advocating for the need to define a union information quantity from which synergy can be derived [38]. In particular, Kolchinsky highlights that rejecting the IEP might remove some of the incompatibilities that we describe in the next section. Although we do not treat IEP as a property in and of itself in this work, future research might benefit strongly from carefully reconsidering its role in the field of information decomposition.

For the reader's benefit, we provide a brief description of each property in Table 5, with the full definitions for each property available in the appendix.

## 2.3 Implications and no-go theorems

As mentioned above, not all the PID properties thus far proposed are independent of each other. In fact, multiple theorems have been established, demonstrating non-trivial relations between the properties, complicating the quest for finding the minimal requirements that a redundancy function should satisfy. Several results show how certain properties can be derived from others (Table 3). For instance, examples of standard implications can be found in the original works by Williams and Beer, who showed that  $(\mathbf{GP})$  follows from the minimal axioms  $(\mathbf{S}_0)$ ,  $(\mathbf{SR})$ , and  $(\mathbf{M}_0)$  [10], or in Bertschinger *et al.* [19], who remarked that  $(\mathbf{TC})$  and  $(\mathbf{GP})$  ensure  $(\mathbf{TM})$ .

Further exacerbating the problem, it has been shown that not all the suggested properties can be simultaneously satisfied, raising the question of which axioms to drop and which ones to retain. One of the most influential inconsistency theorems was the one presented by Rauh *et al.*, who showed that the original Williams and Beer axioms  $((\mathbf{S}_0)$ ,  $(\mathbf{SR})$ , and  $(\mathbf{M}_0))$ , along with  $(\mathbf{EI})$ ,  $(\mathbf{LP}_1)$ , and  $(\mathbf{ID})$ , are not compatible [42]. This was the first demonstration that apparently intuitive and widely accepted properties could not be simultaneously reconciled, highlighting the need for a more systematic understanding of the relationships among PID axioms. Since then, many additional no-go theorems have begun to populate the PID literature, primarily regarding properties such as  $(\mathbf{S}_1)$ ,  $(\mathbf{LP})$ ,  $(\mathbf{ID})$ , and  $(\mathbf{TC})$  (Table 4).

A common denominator across the majority of these incompatibility results is the use of the so-called XOR-Source-Copy gate, first introduced by Rauh *et al.* [42] and then also used in the later works [30, 38, 41, 65]. This system is composed of three sources of the form  $X_1, X_2 \sim \text{Bern}(1/2)$  and  $X_3 = X_1 \oplus X_2$ , together with a target given by the joint variable  $Y = (X_1, X_2, X_3)$ . The fact that this gate serves as such a good counterexample for many, otherwise intuitive, properties begs the question: what is peculiar about this system that current measures struggle

Name	Implications	Source
Lemma 1	$(\mathbf{LP}) \implies (\mathbf{GP})$	[10]
Lemma 2	$(\mathbf{ID}) \implies (\mathbf{IID})$	[26]
Lemma 3	$(\mathbf{S}_1) \implies (\mathbf{S}_0)$	[19]
Lemma 4	$(\mathbf{M}_0) \implies (\mathbf{SE})$	[26]
Lemma 5	$(\mathbf{DI}) \implies (\mathbf{CO})$	[35]
Prop. 1	$(\mathbf{M}_0), (\mathbf{SR}) \implies (\mathbf{LB})$	[20]
Prop. 2	$(\mathbf{SR}), (\mathbf{LB}) \implies (\mathbf{GP})$	[10]
Prop. 3	$(\mathbf{M}_0), (\mathbf{SR}) \implies (\mathbf{GP})$	[10]
Prop. 4	$(\mathbf{LP}) \implies (\mathbf{M}_0)$	[10, 65, 71]
Prop. 5	$(\mathbf{GP}), (\mathbf{TC}) \implies (\mathbf{TM})$	[19]
Theo. 2	$(\mathbf{SR}), (\mathbf{TC}), (\mathbf{TE}) \implies (\mathbf{ID})$	new
Theo. 6	$(\mathbf{M}_0), (\mathbf{TE}) \iff (\mathbf{M}_1)$	new
Theo. 7	$(\mathbf{BP}) \implies (\mathbf{SE}), (\mathbf{TE}) \quad (n = 2)$	new
Prop. 6	$(\mathbf{ID}), (\mathbf{TM}) \implies I_{\cap}(X_1, X_2; f(X_1, X_2)) \leq I(X_1; X_2)$	[66]
Theo. 9	$(\mathbf{M}_0), (\mathbf{S}_1) \implies (\mathbf{TM})$	[19]
Theo. 10	$(\mathbf{S}_1), (\mathbf{SE}) \implies (\mathbf{TE})$	new
Theo. 11	$(\mathbf{SR}), (\mathbf{SE}), (\mathbf{S}_1) \implies (\mathbf{ID})$	new

**TAB. 3: Implications, equivalences, and inequalities among PID properties.**

to understand? A plausible explanation is that it exhibits genuinely higher-order redundancy, i.e. redundancy at the triplet level that cannot be reduced to pairwise relations. This is a consequence of  $X_1, X_2, X_3$  being pairwise independent, yet any one of them is fully determined by the other two. Such a configuration is clearly not possible in systems with only two sources. From this perspective, PID appears to lack the ability to distinguish between different kinds of redundancy. Indeed, Rauh pointed out that the XOR-Source-Copy is characterised by one bit of *redundant* information which is not *shared*, and as such it cannot be appointed to a specific atom of the PID lattice [70]. This observation suggests that intuitions based on shared information might be misleading when defining redundancy in multivariate settings. It is also worth noting that this counterexample leverages the relabelling invariance of the outcomes ( $\mathbf{EI}$ ). Therefore, an alternative solution could be to drop  $\mathbf{EI}$  in favour of  $\mathbf{IID}$  (Sec. 4.1).

Finally, we note that several results proposed in the literature were, unfortunately, later shown to be incorrect. These include the claim that  $\mathbf{TC}$  and  $\mathbf{LP}$  imply  $\mathbf{ID}$ , as stated in Ref. [19], the assertion that  $\mathbf{LP}$  and  $\mathbf{TM}$  are incompatible, reported in Ref. [66], and the fact that  $(\mathbf{S}_0), (\mathbf{M}_0), (\mathbf{SR}), (\mathbf{EI}), (\mathbf{IID})$  are incompatible for  $n \geq 3$  (Lemma 3 of Ref. [41]), as the proposed proof requires the property  $\mathbf{LP}$ , which is not stated in the hypotheses.

## 2.4 A brief history of PID measures

In this section, we provide a brief chronological overview of the introduction of the existing PID measures to date. A more comprehensive description is reported in the appendix, where we also include a visual illustration of the historical development of PID (Fig. 4).

The first redundancy measure  $I_{\cap}^{\min}$  was proposed alongside the PID framework in the seminal work by Williams and Beer [10], making use of the idea that redundant information is present in every realisation of the target. Despite being intuitive, it was then quickly criticised for overestimating redundancy by estimating the same *amount* of redundant information, rather than the same information *content* [18, 19, 22].

This concern – that quality counts as much as quantity – motivated Harder *et al.* to introduce  $I_{\cap}^{\text{red}}$  [18], a redundancy function grounded in information geometry.  $I_{\cap}^{\text{red}}$  overcomes this shortcoming of  $I_{\cap}^{\min}$  by satisfying the Identity property ( $\mathbf{ID}$ ). Despite this,  $I_{\cap}^{\text{red}}$  is majorly limited by its applicability, as it is only valid in the bivariate

Name	Incompatible properties	Source
Theo. 1	( <b>S</b> <sub>0</sub> ), ( <b>SR</b> ), ( <b>EI</b> ), ( <b>LP</b> <sub>1</sub> ), ( <b>IID</b> )	[65]
Theo. 3	( <b>SR</b> ), ( <b>EI</b> ), ( <b>LP</b> ), ( <b>IID</b> ), ( <b>TC</b> )	[30]
Theo. 4	( <b>SR</b> ), ( <b>EI</b> ), ( <b>LP</b> ), ( <b>TE</b> ), ( <b>TC</b> )	new
Theo. 5	( <b>SR</b> ), ( <b>EI</b> ), ( <b>LP</b> <sub>1</sub> ), ( <b>TC</b> )	[65]
Theo. 8	( <b>GP</b> ), ( <b>IID</b> ), ( <b>TM</b> ), mechanistic redundancy	[27, 67]
Theo. 12	( <b>M</b> <sub>0</sub> ), ( <b>SR</b> ), ( <b>LP</b> <sub>1</sub> ), ( <b>S</b> <sub>1</sub> )	[19]

**TAB. 4: Incompatibility results among PID properties.** Where applicable, (**EI**) has been explicitly stated, and (**ID**) has been relaxed into (**IID**).

PID case [18].

Shortly thereafter, Griffith *et al.* proposed an algebraic approach to defining redundancy, first with  $I_{\cap}^{\wedge}$ , inspired by the Gács-Körner common information [22], and later with  $I_{\cap}^{\alpha}$ , which follows from intuitive considerations based on Markov chains [23]. Although these measures are well-defined for any number of sources and satisfy highly desirable properties, most notably (**TM**) in the case of  $I_{\cap}^{\wedge}$ , they fail to satisfy (**LP**), tend to impose overly restrictive bounds on redundancy [23, 38], and, for  $I_{\cap}^{\wedge}$ , are insensitive to statistical correlations not captured by Gács-Körner common random variables [27].

A major step forward came with  $I_{\cap}^{\text{BROJA}}$ , simultaneously introduced by Bertschinger *et al.* alongside Griffith and Koch [21, 22].  $I_{\cap}^{\text{BROJA}}$  was the first redundancy function backed by an explicit operational approach, defining unique information by comparing the maximal expected reward functions of each source. Despite its interpretational clarity and the fact that it satisfies many of the proposed PID properties, this measure has been mildly criticised for overestimating redundancy by artificially inflating the correlation between the sources [26, 27, 29], as well as for relying on a decision-theoretic setup deemed too restrictive [26]. Additionally, this measure is also limited to the bivariate case.

Up until this point, PID measures had focused exclusively on discrete systems. This changed with the introduction of  $I_{\cap}^{\text{MMI}}$  by Barrett [24], a redundancy measure specifically designed for Gaussian systems.  $I_{\cap}^{\text{MMI}}$  was shown to yield the same decomposition as  $I_{\cap}^{\text{min}}$ ,  $I_{\cap}^{\text{red}}$ , and  $I_{\cap}^{\text{BROJA}}$  in the specific case of a bivariate Gaussian PID with univariate target, hence becoming very popular and widely applied. However,  $I_{\cap}^{\text{MMI}}$  inherited the same shortcoming of  $I_{\cap}^{\text{min}}$ , being only able to measure the same *amount* of information between sources, without discerning between qualitatively distinct components of the interaction. This limitation motivated the proposal of  $I_{\cap}^{\text{RR}}$  by Goodwell and Kumar, a modified version of  $I_{\cap}^{\text{MMI}}$  which provides zero redundant information for uncorrelated sources [25].

A parallel but conceptually distinct line of work then led to the introduction of pointwise PID measures. Ince proposed the first fully pointwise PID function,  $I_{\cap}^{\text{CCS}}$ , defining redundancy via common changes in surprisal [26]. In doing so,  $I_{\cap}^{\text{CCS}}$  departs from several PID properties that had previously been uncontroversially accepted, such as (**M**<sub>0</sub>), while retaining the essential axioms required to construct the redundancy lattice. This perspective was also pursued by Finn and Lizier with  $I_{\cap}^{\text{PM}}$  [30, 72], which distinguished informative from *misinformative* contributions and suggested a double decomposition over specificity and ambiguity lattices. As with  $I_{\cap}^{\text{CCS}}$ , this pointwise approach entails abandoning some of the original PID axioms, although they continue to hold at the local level.

Later approaches explored the role of a maximum-entropy principle in defining redundancy. Investigating the subtle differences between entropy maximisation and mutual information minimisation under Assumption (\*), James *et al.* defined  $I_{\cap}^{\text{MES}}$  [29].  $I_{\cap}^{\text{MES}}$  was never intended to be a fully fledged PID measure, as it suffers from the major shortcoming of not satisfying (**SE**).

Around the same time, inspired by works on cybernetics and reconstructability analysis [73, 74], James *et al.* also introduced  $I_{\cap}^{\text{DEP}}$ , a measure of unique information based on the construction of a constraint lattice. Originally proposed only for discrete systems, this measure was then formalised for continuous Gaussian processes [28]. As with  $I_{\cap}^{\text{BROJA}}$ ,  $I_{\cap}^{\text{DEP}}$  only provides a full decomposition in the bivariate case [27], thereby restricting its applicability.

Subsequent definitions of redundant information have continued to explore new avenues of thought. Niu and Quinn proposed  $I_{\cap}^{\text{IG}}$  [31], an information-geometry-based PID inspired by Amari’s hierarchical decomposition [75]. Originally defined for a bivariate discrete PID with full-support distributions,  $I_{\cap}^{\text{IG}}$  was later generalised to the

bivariate Gaussian case [32]. Aiming to account for the indirect mediated associations between sources and target, Sigtermans proposed the redundancy measure  $I_{\cap}^{\text{CT}}$  based on the framework of Causal Tensors [33,34].

An additional pointwise redundancy measure named  $I_{\cap}^{\text{SX}}$  was then suggested by Makkeh *et al.*, both for discrete [35] and continuous scenarios [36,37].  $I_{\cap}^{\text{SX}}$  measures redundancy via the so-called principle of *shared mass exclusion*. Similarly to  $I_{\cap}^{\text{PM}}$ ,  $I_{\cap}^{\text{SX}}$  is based on separate information and misinformation lattices, thereby renouncing some of the original PID axioms at the global level.

More recently, inspired by the algebraic measures  $I_{\cap}^{\wedge}$  and  $I_{\cap}^{\alpha}$ , Kolchinsky introduced  $I_{\cap}^{\prec}$ , a redundancy measure grounded in the notion of Blackwell order [38]. This definition admits a clear operational interpretation in decision-theoretic terms, satisfies **(IID)**, and is defined for any number of sources, but can provide negative atomic values.

With a similar intent to define a PID based on Blackwell order, Venkatesh and Schamberg proposed  $I_{\cap}^{\delta}$  for multivariate Gaussian systems [39], extending the earlier results of Barrett [24] by measuring how distant one source is from being Blackwell superior to the other. Although this approach generalises the behaviour of  $I_{\cap}^{\text{MMI}}$ , it is only defined for a bivariate PID.

Continuing along the Blackwell ordering paradigm, Mages *et al.* finally introduced  $I_{\cap}^{\text{RDR}}$ , quantifying redundancy via the set-overlap of the reachable decision region of each source [40].  $I_{\cap}^{\text{RDR}}$  yields a pointwise decomposition that nevertheless satisfies the original PID axioms for any number of sources, reducing to  $I_{\cap}^{\text{min}}$  and  $I_{\cap}^{\text{BROJA}}$  in specific circumstances.

Finally, the most recent PID measure introduced,  $I_{\cap}^{\text{do}}$  by Lyu *et al.* [41], defines unique information using an interventional approach inspired by Pearl’s *do-calculus* [76–78]. While satisfying several desirable PID properties, a major shortcoming of  $I_{\cap}^{\text{do}}$  is that it does not satisfy **(SE)**, assigning non-zero unique information even when the sources are identical. Again, this measure is also only defined for the case of two sources.

In passing, we also mention  $I_{\cap}^{\text{RAV}}$ , an unpublished proposal that defines redundancy via the maximisation of coinformation over functions of the sources [79]. Despite being unpublished, an implementation of  $I_{\cap}^{\text{RAV}}$  is available via the `dit` package, and we have hence included it in our analyses.

### 3 Results

In this section, we present our main results. We begin with a systematic overview of the PID properties satisfied by each existing PID measure. We then present a comprehensive collection of theorems and no-go theorems that elucidate the logical relationships among these properties.

#### 3.1 A comprehensive classification of PID measures

As discussed above, the proliferation of PID properties and measures has led to substantial ambiguity regarding which axioms are satisfied by which measures. We address this issue by providing a complete and unified classification (Table 5), explicitly showing for all existent PID measures whether they satisfy or not a given PID axiom. While a small subset of these results is already available in the literature, the majority of them have not been established previously. We report detailed proofs of all findings in the Appendix, also indicating where previous results can be found in the literature.

This comprehensive picture provides many different insights at a glance. First, it allows us to gauge which properties are most commonly satisfied in the PID literature, from classic axioms such as **(M<sub>0</sub>)** and **(GP)**, to others which are often not explicitly stated, like **(TE)** and **(CO)** (Fig.1a). Then, it enables drawing connections among properties, suggesting which ones are usually satisfied together. Additionally, it makes clear what a specific operational interpretation entails in terms of PID properties. Finally, it also helps us understand the theoretical relationships between the various redundancy functions.

Specifically, we make this latter observation quantitative by studying the similarity of the PID measures via a hierarchical clustering based on the properties they satisfy (Fig. 1b). Results show that the measures are clustered in groups characterised by a small subset of PID properties. The first big differentiation is among measures that do not satisfy **(GP)** ( $I_{\cap}^{\text{CCS}}$ ,  $I_{\cap}^{\text{PM}}$ ,  $I_{\cap}^{\text{SX}}$ ) against those that do. This bigger group can be further clustered first into  $I_{\cap}^{\text{do}}$  and  $I_{\cap}^{\text{MES}}$ , which do not satisfy **(M<sub>0</sub>)** nor **(LP<sub>0</sub>)**, then into the algebraic measures  $I_{\cap}^{\wedge}$ ,  $I_{\cap}^{\alpha}$ ,  $I_{\cap}^{\prec}$ , which satisfy

	$I_{\cap}^{\min}$	$I_{\cap}^{\text{MMI}}$	$I_{\cap}^{\text{RDR}}$	$I_{\cap}^{\text{BROJA}}$	$I_{\cap}^{\delta}$	$I_{\cap}^{\text{red}}$	$I_{\cap}^{\text{IG}}$	$I_{\cap}^{\text{RR}}$	$I_{\cap}^{\text{DEP}}$	$I_{\cap}^{\text{CT}}$	$I_{\cap}^{\text{RAV}}$	$I_{\cap}^{\wedge}$	$I_{\cap}^{\alpha}$	$I_{\cap}^{\prec}$	$I_{\cap}^{\text{MES}}$	$I_{\cap}^{\text{do}}$	$I_{\cap}^{\text{PM}}$	$I_{\cap}^{\text{SX}}$	$I_{\cap}^{\text{CCS}}$
(SR)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
(S <sub>0</sub> )	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
(M <sub>0</sub> )	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓ <sup>†</sup>	✓	✓	✓	✗	✗	✗ <sup>‡</sup>	✗ <sup>‡</sup>	✗
(GP)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗ <sup>‡</sup>	✗ <sup>‡</sup>	✗
(LP <sub>0</sub> )	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✗	✗	✗	✗	✗	✗ <sup>‡</sup>	✗ <sup>‡</sup>	✗
(LP <sub>1</sub> )	✓	✓	✓	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	✗	✗	✗	✗	✗	n.a.	n.a.	✗ <sup>‡</sup>	✗ <sup>‡</sup>	✗
(IID)	✗	✗	✗	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗	✓
(ID)	✗	✗	✗	✓	✓	✓	✗	✗	✓	✓	✓	✗	✗	✗	✓	✓	✗	✗	✗
(TM)	✗	✓	✗	✗	✗	✗	✓*	✗	✗	✓	✗	✓	✗	✗	✗	✗	✗	✗	✗
(TC)	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✓	✗
(SE)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗	✓	✓	✓
(LB)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓ <sup>†</sup>	✓	✓	✓	✗	✗	✗	✗	✗
(TE)	✓	✓	✓ <sup>†</sup>	✓	✓	✓	✓	✓	✓	✓	✓ <sup>†</sup>	✗	✓	✓	✓	✓	✗	✗	✗
(M <sub>1</sub> )	✓	✓	✓ <sup>†</sup>	✓	✓	✓	✓	✓	✓	✓	✓ <sup>†</sup>	✗	✓	✓	✗	✗	✗	✗	✗
(S <sub>1</sub> )	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
(*)	✓	✓	✓	✓	✓	✓	✗	✗	✗	✗	✗	✓	✗	✗	✓	✓	✗	✗	✗
(BP)	✗	✗	✓	✓	✓	✓	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗ <sup>‡</sup>	✗	✗
(AD)	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✓	✓	✓	✓	✓	✗	✗	✗
(CO) <sup>  </sup>	✓	✓	✓	✓	✓	✓ <sup>§</sup>	✓	✓	✓	✓	✗	✗	✗	✗	✓	✓	✓	✓	✗
(DI) <sup>  </sup>	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✓	✗
(EI)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

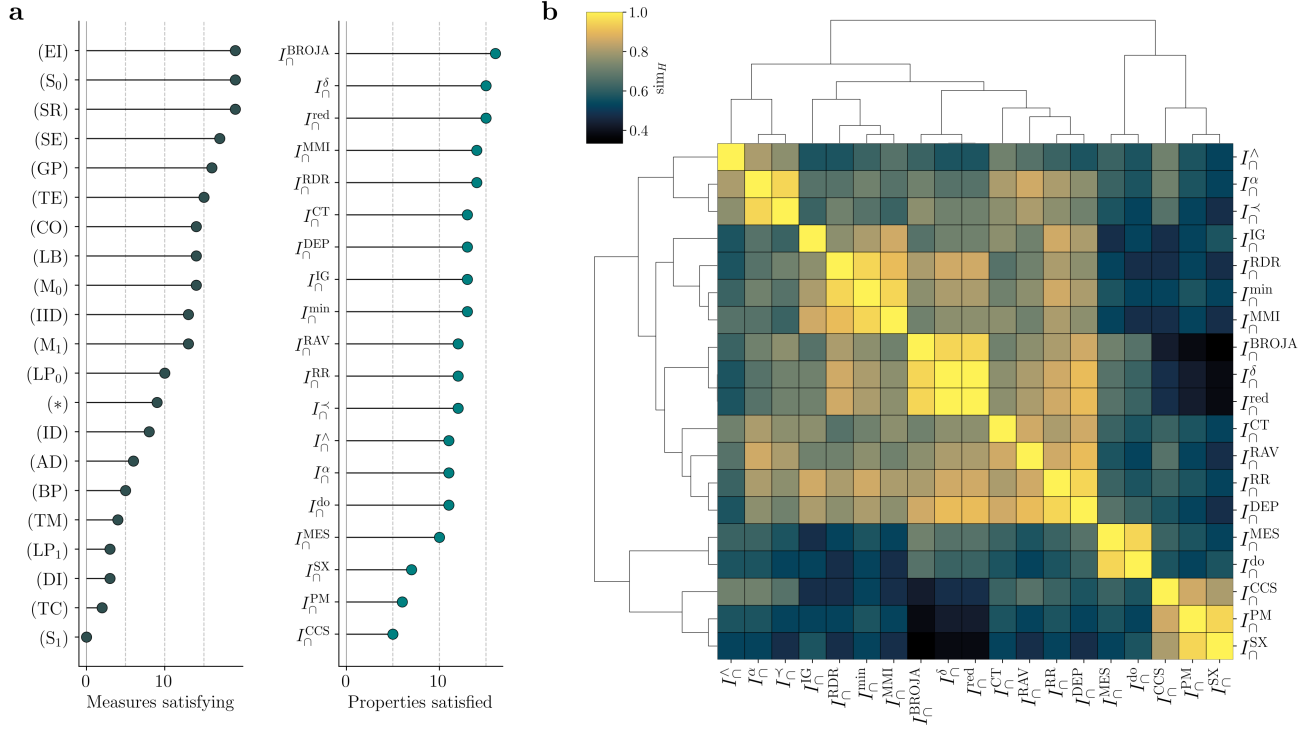
**TAB. 5: Comparison of theoretical properties across redundancy measures.** The ordering of the measures is inspired by the hierarchical clustering of Fig. 1b. \*: this property is not formally proven but is supported by empirical simulations. †: these properties only hold for  $n = 2$  sources. §: these properties only hold in the case of full-support distributions. ||: these properties are intended globally; for many measures, the almost-everywhere version holds. ‡: the pointwise version of these properties holds for both informative and misinformative lattices separately.

(M<sub>0</sub>) but not (LP<sub>0</sub>), and finally into all the other measures which satisfy both (M<sub>0</sub>) and (LP<sub>0</sub>). Within this subgroup, we notice additional clusterings due to  $I_{\cap}^{\min}$ ,  $I_{\cap}^{\text{MMI}}$ , and  $I_{\cap}^{\text{RDR}}$  satisfying both (LP<sub>1</sub>) and (\*),  $I_{\cap}^{\text{BROJA}}$ ,  $I_{\cap}^{\delta}$ , and  $I_{\cap}^{\text{red}}$  satisfying (\*) but not (LP<sub>1</sub>), and the remaining that satisfy neither. Finally, the final cluster is given by the measures that satisfy (ID) ( $I_{\cap}^{\text{RAV}}$ ,  $I_{\cap}^{\text{DEP}}$ ,  $I_{\cap}^{\text{CT}}$ ) and those that do not ( $I_{\cap}^{\text{IG}}$ ,  $I_{\cap}^{\text{RR}}$ ). In sum, although we considered twenty axioms in this analysis, only around a quarter of them are sufficient to capture the main mathematical distinctions among the PID measures proposed to date. This highlights that the various properties are not independent, and raises the question of whether there exist yet undiscovered relationships linking the key discriminating properties to the others.

### 3.2 A comprehensive view of theorems linking PID properties

In this section, we present a comprehensive account of the relations among PID properties. We review results previously established in the literature, strengthen their hypotheses and clarify explicitly assumptions that were previously left implicit, and complement them with new findings. In particular, we accompany the presentation of the results with a brief commentary on what their consequences imply, and how these incompatibility issues could be overcome. The mathematical formulation of the results, as well as their proofs, is reported in the Appendix.

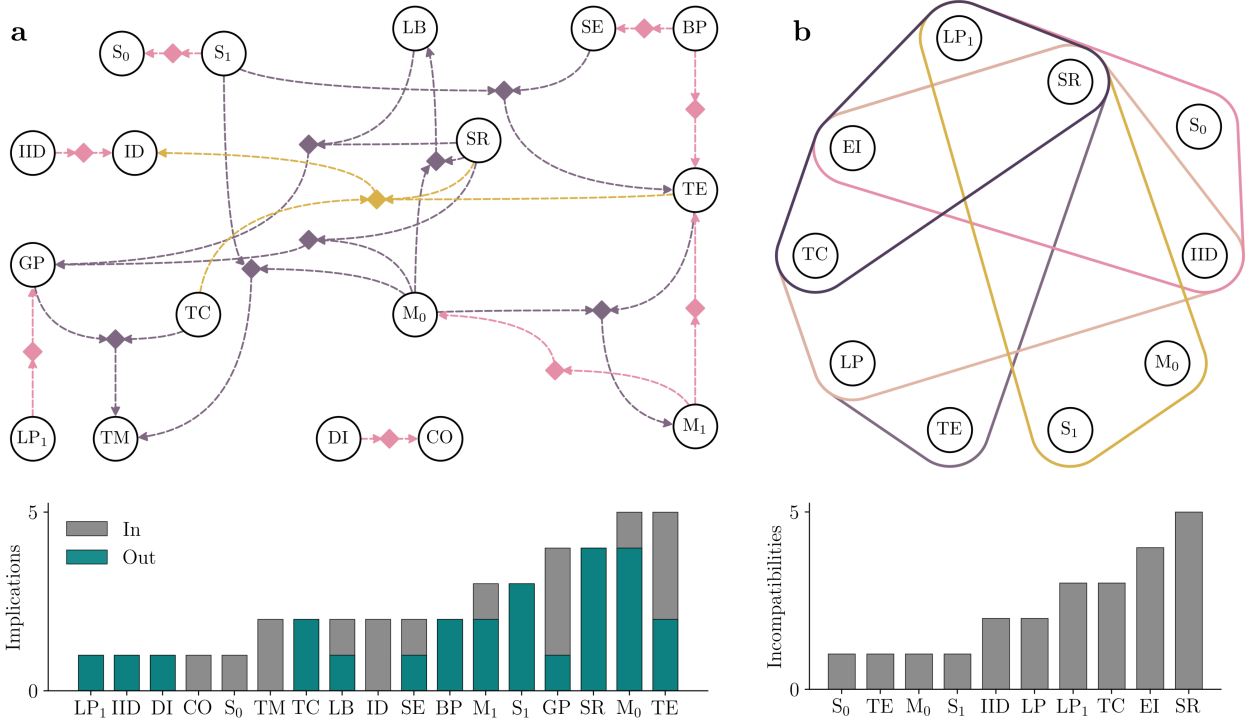
The first demonstration that not all the PID properties can simultaneously coexist was proposed by Rauh *et al.* [42], who showed that (S<sub>0</sub>), (M<sub>0</sub>), (SR), (EI), (LP<sub>1</sub>) and (ID) are incompatible. Although not explicitly



**FIG. 1: Hierarchical organisation of PID measures induced by properties.** **a)** Number of measures satisfying each mathematical property (left), and number of properties satisfied by each measure (right), from Table 5. **(SR)**, **(EI)** and **(S<sub>0</sub>)** are satisfied by all measures, whereas **(S<sub>1</sub>)** is currently satisfied by no measures.  $I_{\emptyset}^{\text{BROJA}}$  satisfies the most properties (16), whereas  $I_{\emptyset}^{\text{CCS}}$  satisfies the least (5). **b)** Hierarchical clustering of the PID measures based on mathematical properties.  $I_{\emptyset}^{\text{CCS}}$ ,  $I_{\emptyset}^{\text{PM}}$  and  $I_{\emptyset}^{\text{SX}}$  clearly emerge as a distant community of measures likely due to their unique lack of possession of **(GP)**.  $I_{\emptyset}^{\text{MES}}$  and  $I_{\emptyset}^{\text{do}}$  occupy another cluster in between this and the main community of measures, likely due to the absence of **(M<sub>0</sub>)**, **(M<sub>1</sub>)**, **(LB)**, which they share with  $I_{\emptyset}^{\text{CCS}}$ ,  $I_{\emptyset}^{\text{PM}}$  and  $I_{\emptyset}^{\text{SX}}$ . The main cluster comprises  $I_{\emptyset}^{\text{RDR}}$ ,  $I_{\emptyset}^{\text{min}}$ ,  $I_{\emptyset}^{\text{MMI}}$ ,  $I_{\emptyset}^{\text{red}}$ ,  $I_{\emptyset}^{\text{BROJA}}$ ,  $I_{\emptyset}^{\delta}$ ,  $I_{\emptyset}^{\text{IG}}$ ,  $I_{\emptyset}^{\text{RR}}$ ,  $I_{\emptyset}^{\text{RAV}}$ ,  $I_{\emptyset}^{\text{DEP}}$ , and  $I_{\emptyset}^{\text{CT}}$ . A particularly compact subgroup is unsurprisingly formed by  $I_{\emptyset}^{\text{RDR}}$ ,  $I_{\emptyset}^{\text{min}}$  and  $I_{\emptyset}^{\text{MMI}}$  due to satisfying **(LP<sub>1</sub>)** and their lack of **(IID)**. Similarly,  $I_{\emptyset}^{\text{red}}$ ,  $I_{\emptyset}^{\text{BROJA}}$ ,  $I_{\emptyset}^{\delta}$  form a subcluster as a consequence of satisfying **(BP)**, **(\*)**, and **(ID)**. Finally, the last community comprises  $I_{\emptyset}^{\angle}$ ,  $I_{\emptyset}^{\alpha}$ , and  $I_{\emptyset}^{\angle}$ , which violate **(LP<sub>0</sub>)** while retaining **(IID)**.

stated in the hypotheses, we underline the necessity of **(EI)** required for the relabelling of the outcomes within the XOR-Source-Copy gate, as later clarified by Matthias *et al.* [65]. In that work, the authors further refine Rauh *et al.*'s result by showing that **(M<sub>0</sub>)** is not necessary (Theorem 1 of [65]). Here, we additionally observe that the requirement of **(ID)** can be relaxed in favour of the weaker **(IID)**, as the same counterexample holds. Hence, in its minimal assumptions, the incompatibility involves **(S<sub>0</sub>)**, **(SR)**, **(EI)**, **(LP<sub>1</sub>)**, and **(IID)** (Theo. 1).

Interestingly, we also note that this same result has also been indirectly derived from different premises. Kolchinsky assumed **(S<sub>0</sub>)**, **(M<sub>0</sub>)**, **(SR)**, **(EI)**, **(IID)**, and showed that the union information ( $I_{\cup}$ ) can exceed the joint mutual information ( $I$ ), leveraging this result to argue against the inclusion-exclusion principle (IEP) (Lemma 1 of Ref. [38]). However, while having  $I_{\cup} > I$  can be intuitively unpleasant, it only becomes a formal mathematical contradiction if one also assumes **(LP)**, as  $I_{\cup} > I$  together with IEP indeed implies the presence of negative atoms (see Prop. 7). Hence, if one were to retain IEP at all costs, this argument leads to the rejection of **(LP)**, therefore recovering Theo. 1. Using a different approach based on subsystem consistencies, Lyu *et al.* proposed that **(S<sub>0</sub>)**, **(M<sub>0</sub>)**, **(SR)**, **(EI)**, **(IID)** were incompatible for  $n \geq 3$  sources (Lemma 3 of [41]). However, although not stated in their hypotheses, their proof implicitly relies on **(LP)**, thus corresponding once again to Theo. 1.



**FIG. 2: Formal relationships among PID properties.** **a)** Directed hypergraph representing logical implications of PID properties from Table 3, hyperedges are coloured by order of implication (upper). In- and out-degree distributions for property implications reveal that  $(M_0)$  and  $(SR)$  imply the most properties (4), whereas  $(TE)$  and  $(GP)$  are implied by the most properties (3). **b)** Undirected hypergraph represented logical incompatibilities from Table 4. Degree distributions for property incompatibilities reveal that despite implying the most properties,  $(SR)$  is also incompatible with the most properties (4), tied with  $(EI)$ .

On a parallel line of work, a property which has long been highly desired is Target Chain rule  $(TC)$ , considered by many to be the *holy grail* of the PID axioms. Unfortunately, besides not being satisfied by the majority of the existing PID measures, this property has also been shown to be incompatible with  $(SR)$ ,  $(EI)$ ,  $(LP_0)$ , and  $(IID)$  for any number of sources (Theorem 6 of [30], here Theo. 3).

Notably, the incompatibilities above all involve the controversial property  $(IID)$ . Hence, if one wants to retain the PID lattice structure, the easiest solution would seem to drop  $(IID)$  (and thus  $(ID)$ ) altogether. However, here we present a novel finding which directly relates  $(IID)$  to other intuitive properties. In fact, the combination of the desirable  $(TC)$  with the intuitive  $(TE)$  and the fundamental  $(SR)$  promptly leads to the controversial  $(ID)$  (and hence  $(IID)$ ) (Theo. 2). This result has interesting consequences. If we additionally require  $(GP)$ , then we further obtain  $(TM)$  (see Prop. 5), which was thought to be hard to reconcile with  $(ID)$  [42]. Moreover, although gaining  $(ID)$  is not necessarily problematic per se, it has a more dire consequence when combined with Theo. 3.

In fact, by merging Theos. 2-3 we get that  $(SR)$ ,  $(EI)$ ,  $(LP_0)$ ,  $(TE)$  and  $(TC)$  are not compatible for any number of sources (Theo. 4). Hence, if one wants to retain  $(LP)$ , either  $(TC)$  or  $(TE)$  necessarily need to be dropped. At least for  $n \geq 3$  sources, this last finding can be improved, as it has recently been shown that  $(SR)$ ,  $(EI)$ ,  $(LP_1)$ , and  $(TC)$  are incompatible, even without requiring  $(TE)$  (Theorem 2 of [65], here Theo. 5). Therefore, these findings prevent any redundancy function from satisfying the typical properties of classical information measures.

Thus, if abandoning  $(IID)$  does not resolve these issues, how else can one avoid such undesired behaviours? Taken together, these results suggest that the properties that classical information measures benefit from, such as  $(TC)$  and  $(LP)$ , should not be expected to hold for PID measures. This perspective can be more naturally

accommodated by redundancy measures that give operational meaning and clear interpretation to negative atoms, such as  $I_{\cap}^{\text{CCS}}$ ,  $I_{\cap}^{\text{PM}}$ , and  $I_{\cap}^{\text{SX}}$ , though at the cost of losing  $(\mathbf{M}_0)$ . Moreover, another possible point of reflection is that the mechanistic intuition underlying  $(\mathbf{IID})$  might be fundamentally incompatible with the purely statistical perspective embodied by  $(\mathbf{EI})$ . Consequently, despite the natural appeal of  $(\mathbf{EI})$ , relaxing this property could offer a viable route to alleviate these issues. We refer to Sec. 4.1 for further discussion on these points.

In closing, we also mention a couple of novel results regarding the PID properties  $(\mathbf{TE})$  and  $(\mathbf{BP})$ .  $(\mathbf{TE})$  is justified by the intuitive idea that adding the target to the set of sources should not affect redundancy. Interestingly, combining  $(\mathbf{TE})$  with the fundamental  $(\mathbf{M}_0)$  provides a necessary and sufficient condition for  $(\mathbf{M}_1)$ , supporting the claim that  $(\mathbf{M}_1)$  is a useful desideratum [20] (Theo. 6). Finally, we note that the decision-making-inspired  $(\mathbf{BP})$  is a stronger condition of both  $(\mathbf{TE})$  and  $(\mathbf{SE})$ , at least for  $n = 2$  sources (Theo. 7). Hence, if one accepts the operational idea underpinning  $(\mathbf{BP})$ , then  $(\mathbf{TE})$  follows directly.

We represent all these results and additional existing relations and incompatibilities in a hypergraph (Fig. 2), also reporting which properties are the most commonly involved.

To further investigate the logical relationships among the properties, we employed the Z3 automatic theorem prover, a satisfiability modulo theories program that allows us to consistently check implications and incompatibilities by deciding satisfiability of combinations of axioms [43]. With this tool at hand, we verified that the properties satisfied by each measure (Table 5) are indeed compatible with the known relationships among the PID axioms.

Furthermore, we used Z3 to determine the maximal sets of properties that can coexist without leading to any incompatibility. In other words, these sets represent the largest collections of mutually compatible properties, meaning that no additional property can be added without introducing a conflict. Our results indicate that, as expected, the maximal compatible sets contain 20 properties, achievable simply by abandoning either  $(\mathbf{LP}_1)$  or  $(\mathbf{EI})$ . If both are retained, however, the maximal set is reduced to 17 properties, at the expense of  $(\mathbf{ID})$ ,  $(\mathbf{IID})$ ,  $(\mathbf{TC})$ , and  $(\mathbf{S}_1)$ . Importantly, these conclusions hold under the assumption that the fundamental axioms  $(\mathbf{S}_0)$ ,  $(\mathbf{M}_0)$ , and  $(\mathbf{SR})$  are satisfied.

This analysis carries two important implications. First, it encourages the development of new PID measures capable of satisfying a larger number of axioms, enabling researchers to determine the maximal set of compatible properties given certain desirable features. Second, the fact that no existing PID measure achieves the maximal set suggests the existence of yet undiscovered incompatibility relationships among the properties.

## 4 Discussion

### 4.1 Identity property, Equivalence-class Invariance, and Mechanistic redundancy

As we have seen above, if one wants to retain the fundamental  $(\mathbf{S}_0)$ ,  $(\mathbf{SR})$ , and the convenient  $(\mathbf{LP})$ , then  $(\mathbf{EI})$  is not compatible with  $(\mathbf{ID})$  or  $(\mathbf{IID})$ , at least for  $n \geq 3$  sources (Theo. 1). The conceptual inconsistency between  $(\mathbf{EI})$  and  $(\mathbf{IID})$  already transpired from the conflicting interpretations briefly outlined in Sec. 2.2. Here we further discuss the implications.

A concept strictly related to  $(\mathbf{ID})$ , and  $(\mathbf{EI})$  is that of *mechanistic redundancy*. Introduced by Harder *et al.* [18], mechanistic redundancy indicates the redundant information that is generated by the mechanism, as opposed to the *source redundancy* which stems from the correlation between the sources. The existence of mechanistic redundancy entails that redundant information can be non-zero even if the sources are independent. However, as seen above, if  $(\mathbf{EI})$  is adopted, then the prospect of measuring mechanistic redundancy vanishes, since different mechanisms can lead to the same probability distribution, up to an isomorphism [63]. Thus, although  $(\mathbf{EI})$  has not received much attention in the PID literature, we believe it is fundamental to investigate its significance and ponder on which scenarios it should or should not be expected to hold. To the best of our knowledge, the only work which explicitly argues in favour of  $(\mathbf{IID})$  as opposed to  $(\mathbf{EI})$  is that of Chicharro [64]. In this work, the author specifically investigates the property of PID being invariant under isomorphisms of the target, concluding that if mechanistic redundancy is to hold, then redundancy depends on the composition of the target and one should drop the Target to Source Copy (TSC) isomorphism, i.e. a specific instance of  $(\mathbf{EI})$  applied to the target. In fact, contrary to the strong axiom, the weak version is compatible with the violation of  $(\mathbf{EI})$ , since it allows for dependencies on the semantic properties of the outcomes [64]. Moreover, he also derives that the negativity of atoms follows from the

existence of deterministic target-source dependencies and **(EI)** [64]. Hence, accepting the semantic value of specific realisations can also resolve the non non-negativity of the decomposition.

## 4.2 Practical guidance for empirical applications

With the many approaches available in the PID literature, each satisfying different properties, the endeavour of choosing a PID measure for empirical purposes might seem an overwhelming task. In this section, we aim to alleviate this burden by providing explicit, practical recommendations tailored to the specific application under consideration and to the nature of the sources and target.

A possible starting point could be to reflect upon what kind of information independent sources are expected to provide. For example, PID is often employed in the study of dynamical processes evolving over time, with the sources being the states of two elements of the system in the past, and the target being the joint state of those components in a future timestep, rather than another variable [80, 81]. In this respect, under a mechanistic perspective, one could expect that independent sources provide unique information to the target and zero redundancy or synergy, thus satisfying **(IID)**. This reasoning alone already helps restrict the space of possible PIDs to choose from.

Another key point to consider is the relevance of Local Positivity **(LP)**, which is closely tied to the interpretation one wants to assign to the PID atoms [64]. If PID atoms are interpreted operationally in the classical communication-theoretic sense, **(LP)** becomes essential, as the atoms represent quantities of information that can be transmitted or exploited, with negative values lacking an operational meaning. Within this view, the informational architecture is necessarily grounded in the statistical structure of the variables alone, without reference to semantic content, labels, or specific realisations, hence in accordance with **(EI)**. Belonging to this paradigm is the operational interpretation of PID atoms in terms of secret-key agreement rates, where information is quantified in terms of the ability of parties to establish shared keys under public communication constraints [29]. Within this view, one can distinguish between the so-called *camel* and *elephant* intuitions. The former considers the sources as contributors that jointly construct the outcome of the target, whereas the latter interprets the sources as partial and potentially overlapping observations of the target. Importantly, these two perspectives induce different directional interpretations of redundancy, and certain apparent shortcomings of PID measures arise only when a decomposition is assessed through one intuition rather than the other [29, 82].

Conversely, if information is interpreted as a measure of statistical dependence, negative local contributions correspond to misinformation, i.e. a higher uncertainty about the target after observing the sources. Along this line, pointwise PID approaches relax **(LP)** deliberately, favouring an interpretation of PID atoms in terms of local dependencies and belief updates, rather than communication capacity [26, 30, 35]. Moreover, within this perspective, since PID atoms are associated with changes in the target distribution putatively induced by the sources, a semantic dependency based on the specific value of the variables might be contemplated, objecting **(EI)** to retain mechanistic interpretations. Therefore, whether **(LP)** should be pursued depends on the intended interpretation and operational significance of the decomposition.

Finally, additional considerations relating **(AD)** and **(CO)** can be relevant in empirical settings characterised by noise and multiscale organisation, such as many biological systems [83, 84]. In these contexts, **(AD)** can be desirable when combining information contributions across different scales, e.g. by performing PID analyses considering microscopic and macroscopic states, as it allows to disentangle independent contributions [85]. Likewise, **(CO)** plays an important practical role in the presence of sampling noise and estimation errors, guaranteeing that small perturbations in the underlying distributions lead to small changes in the resulting PID atoms. This robustness is particularly valuable when dealing with high-dimensional data with a limited sample size, where sharp discontinuities in the PID of choice may hinder interpretability or comparability across conditions.

An additional consideration that further restricts the set of admissible measures is their domain of definition. Some PID measures are defined only for discrete variables, whereas others are formulated for continuous Gaussian settings, and only a small subset extends to generic continuous distributions [37]. Moreover, several discrete PID definitions require the underlying distribution to have full support, imposing additional constraints on their applicability.

### 4.3 Other approaches to information decomposition

In this paper, we centred our analysis on the Williams and Beer redundancy lattice, as it is the original and most widely used formalism for studying synergy and redundancy in recent work [10]. Accordingly, our results concern information decomposition measures based on this lattice. Nevertheless, we acknowledge the existence of alternative approaches that investigate information-sharing relationships through different means. In particular, some frameworks decompose mutual information using alternative lattice structures; others focus on decomposing functionals different from mutual information (e.g., entropy-based decompositions); and others characterise synergy and redundancy without performing an explicit decomposition, relying instead on linear combinations of Shannon measures or other statistical quantities.

Approaches that decompose mutual information based on alternative lattice structures or different theoretical frameworks have been proposed for different purposes. Some are motivated by the goal of addressing specific limitations of the original PID lattice, while others aim to provide more general formalisms or to bridge disparate theoretical perspectives. For instance, the constraint lattice originally introduced by Zwick in reconstructability analysis [73] was later employed by James *et al.* to define  $I_{\cap}^{\text{DEP}}$  [27], and has since become widely adopted for information decomposition purposes. In fact, Ay *et al.* proposed an extended version of the lattice by rooting in cooperative game theory, addressing the incompatibility between **(LP)** and **(ID)** [86]. While this construction resolves the conflict between these two central properties, it achieves so by collapsing several redundant atoms of the PID lattice, thereby reducing the level of detail captured by the decomposition. Similarly, Rosas *et al.* leveraged this extended constraint lattice to introduce a synergy-first decomposition, based on an operational notion of synergy motivated by recent developments in the literature of data privacy [87]. Due to the superexponential growth of the number of atoms, they further proposed a backbone lattice, which summarises the decomposition by collapsing multiple terms into a one-dimensional representation. This backbone construction was subsequently employed by Varley, who introduced a scalable synergy-based decomposition in which synergy is quantified in terms of the system’s resilience to the removal of its components [88].

Other proposals investigate the simultaneous use of multiple lattices, exploiting their complementarities by mapping their relationships [38,89]. Within this line of work, Chicharro *et al.* proposed a synergy-first decomposition of mutual information, introducing a lattice based on mutual information loss dual to the original redundancy one, where synergy and redundancy contributions are inverted [89]. A similar approach has been explored by Kolchinsky, who posits the idea that the “inclusion-exclusion principle”, upon which the PID lattice is constructed, should not be expected to hold, thus suggesting a double decomposition of union and intersection information [38]. Beyond the inversion of synergy and redundancy, Pica *et al.* proposed analysing all possible PID decompositions obtained by swapping sources and target, then extracting invariant informational components shared across the different combinations [90]. From a different perspective, Perrone *et al.* introduced an information-geometric approach to capture the hierarchical organisation of interactions, offering an alternative viewpoint on how synergy can be quantified across different orders of dependencies [91]. More recently, an alternative lattice that removes singleton nodes has been introduced by Lyu *et al.* [92]. This reduced lattice focuses on higher-order synergistic and unique contributions, at the cost of not resolving redundancy terms, resulting in a coarser decomposition.

Other frameworks depart from decomposing mutual information altogether. To investigate how information is structured across multiple sources and targets, Mediano *et al.* developed the  $\Phi$ ID formalism [93].  $\Phi$ ID extends the PID redundancy lattice to a double-redundancy lattice, introducing additional atoms that describe novel modes of informational contributions. However, while this framework allows a finer-grained decomposition of information, it also requires the definition of a double-redundancy function. A different direction consists of entropy-based decompositions such as Partial Entropy Decomposition (PED), which avoid the explicit distinction between sources and targets and instead decompose the overall informational structure of a set of random variables [51,94]. Along similar lines, a more general proposal that embeds both entropy and mutual-information decompositions is the Generalised Information Decomposition (GID) proposed by Varley [95]. GID provides a unifying perspective by performing a decomposition based on the Kullback–Leibler divergence, enabling a comparison of how different lattice-based approaches relate to one another.

Finally, several approaches characterise synergy and redundancy without relying on lattices at all. While these approaches offer advantages, either in not relying on specific definitions of redundancy or being less computationally expensive, they also do not allow for decomposing information into all the atoms defined by PID. Quax *et al.* intro-

duced an operational definition of synergy based on synergistic auxiliary variables, providing a constructive way to isolate purely synergistic information [96]. Olbrich *et al.* proposed a geometric approach that decomposes synergistic contributions by projecting probability distributions onto exponential families [97]. Other methods quantify synergy and redundancy using only linear combinations of Shannon entropies or mutual information, thereby avoiding the need to define a redundancy function or a decomposition lattice. These approaches have the advantage of requiring no additional assumptions beyond classical information theory. Prominent examples include the O-information [98], the redundancy–synergy index [99], and the coinformation [100], which identify whether synergy or redundancy is the dominant mode of information sharing, and the Total Correlation hierarchical decomposition [75], the neural information decomposition [101], the connected information [102–104] and the symmetric decomposition [105], which instead aim to disentangle dependencies of different orders. A recent unifying framework formalises this class of measures by defining Shannon invariants as all combinations of informational atoms that can be directly computed as linear combinations of entropy or mutual information [106].

An additional approach developed to study synergistic and redundant effects is the Logarithmic Decomposition (LD) by Down and Mediano [107, 108]. LD extends the Möbius inversion to create a signed-measure space for classical information measures, where variables correspond to measurable sets on which standard set-theoretic operations can be performed. Using LD, the authors showed that the dyadic and triadic systems of James and Crutchfield [9] – famously thought to be classically indistinguishable – can in fact be algebraically distinguished by interrogating classical information measures carefully [107].

#### 4.4 Limitations and future work

Despite the systematic and comprehensive nature of our work, we acknowledge some limitations. While we presented the most complete map of relationships among PID axioms to date, collecting all the properties mentioned in the literature and introducing new theorems, it remains possible that additional implications or inconsistencies exist and have not been identified by our analysis. Moreover, our analysis is primarily conceptual and theoretical, leaving for future work a systematic empirical evaluation across real-world systems. In fact, while we focused on comparing formal properties and logical relationships between different measures, we did not assess their practical behaviour or performance on empirical data when addressing concrete scientific questions. Future studies might further investigate the relationship between the different informational approaches of this study and the notions of causality and emergence, which are central to the study of complex systems. In addition, although we concentrated on the standard PID lattice, future work could extend this analysis to alternative lattice structures and to decompositions of information-theoretic quantities beyond those considered here.

## 5 Conclusion

In this work, we provided a unified and systematic analysis of Partial Information Decomposition approaches grounded in the Williams and Beer redundancy lattice, organising the existing PID measures and properties and clarifying the relationships among them. By collecting existing results, identifying previously implicit connections, and introducing new theorems, we offer a coherent map of the conceptual landscape underlying the quantification of redundancy, synergy, and unique information within this widely used formalism. Beyond serving as a reference for existing PID measures, this analysis highlights the structural constraints that shape possible decompositions and clarifies which desiderata can, or cannot, be simultaneously satisfied. We hope that this perspective will help guide both the development of new information-decomposition frameworks and the informed application of existing ones, contributing to a clearer understanding of multivariate information sharing in complex systems.

#### Acknowledgments

We thank Hardik Rajpal, Gustavo Menesse, Fernando Rosas, Davide Orsenigo, Maximilian Kathofer, and all the members of the Imperial College MIND Lab for the useful discussions and precious comments. We also thank Aaron J. Gutknecht, Philip H. Matthias, and Artemy Kolchinsky for suggestions that helped improve this work. M.N. has received funding from the French government through the ‘France 2030’ investment plan managed by the

French National Research Agency (Agence Nationale de la Recherche; reference: ANR-16-CONV000X/ANR-17-EURE-0029) and from the Excellence Initiative of Aix-Marseille University-A\*MIDEX (AMX-19-IET-004). M.N. has received funding from the Boehringer Ingelheim Fonds (BIF). G.B. is supported by the Leverhulme Trust.

### **Code availability**

The code used to perform the analyses with the Z3 automatic theorem prover is publicly available in the open-source repository at <https://github.com/Imperial-MIND-lab/pid-prover>.

## A Notation and mathematical background

In this work, we make use of the following notation. We indicate a generic probability distribution for the random variables  $X_1, \dots, X_n$  as  $P(X_1, \dots, X_n)$ , or equivalently as  $P_{X_1 \dots X_n}$ . On the other hand, we denote the probability associated with a specific outcome  $X_1 = x_1, \dots, X_n = x_n$  as  $p(x_1, \dots, x_n)$ ,  $p(X_1 = x_1, \dots, X_n = x_n)$ , or  $P(X_1 = x_1, \dots, X_n = x_n)$ . Throughout the work, we use these notations interchangeably. Unless otherwise specified, within PID the sources will be indicated as  $X_1, \dots, X_n$ , while the (possibly multivariate) target is  $Y \equiv (Y_1, \dots, Y_n)$ . We indicate both the mutual information and the coinformation with the letter  $I(\cdot)$ , since the latter reduces to the former in the case of two arguments. The mutual information can be both expressed in terms of the probability of each outcome, or in terms of entropies:

$$\begin{aligned} I(X; Y) &= H(X) - H(X|Y) = H(X) + H(Y) - H(X, Y) \\ &= \sum_{xy} p(x, y) \log \frac{p(x, y)}{p(x)p(y)}. \end{aligned} \quad (8)$$

For the coinformation, we write it following the convention of interaction information, i.e.

$$I(X_1; \dots; X_{n+1}) = I(X_1; \dots; X_n) - I(X_1; \dots; X_n | X_{n+1}). \quad (9)$$

For three arguments, this reduces to

$$I(X_1; X_2; X_3) = I(X_1; X_3) + I(X_2; X_3) - I(X_1, X_2; X_3). \quad (10)$$

If we interpret  $(X_1, X_2)$  as sources and  $X_3$  as a target of a PID, Williams and Beer showed that the coinformation corresponds the difference between redundancy and synergy, thus explaining the until then unexplained negative values obtained by the coinformation [10]. To distinguish between multivariate and multiple arguments, we respectively make use of the comma , and the semicolon ;. For instance,  $I(X_1, X_2, X_3; Y)$  is the mutual information between  $(X_1, X_2, X_3)$  and  $Y$ , while  $I(X_1, X_2, X_3; Y)$  is the coinformation between  $(X_1, X_2)$ ,  $X_3$  and  $Y$ .

### A.1 Equivalence relations and partial ordering

In Ref. [109], an equivalence relation between random variables is proposed. Intuitively, this predicates that two random variables are informationally equivalent if and only if they can be mapped to each other through a one-to-one correspondence. Clearly, this induces an equivalence relation:

$$X \sim Y \iff \exists, f \text{ s.t. } X = f(Y), \text{ and } Y = f^{-1}(X). \quad (11)$$

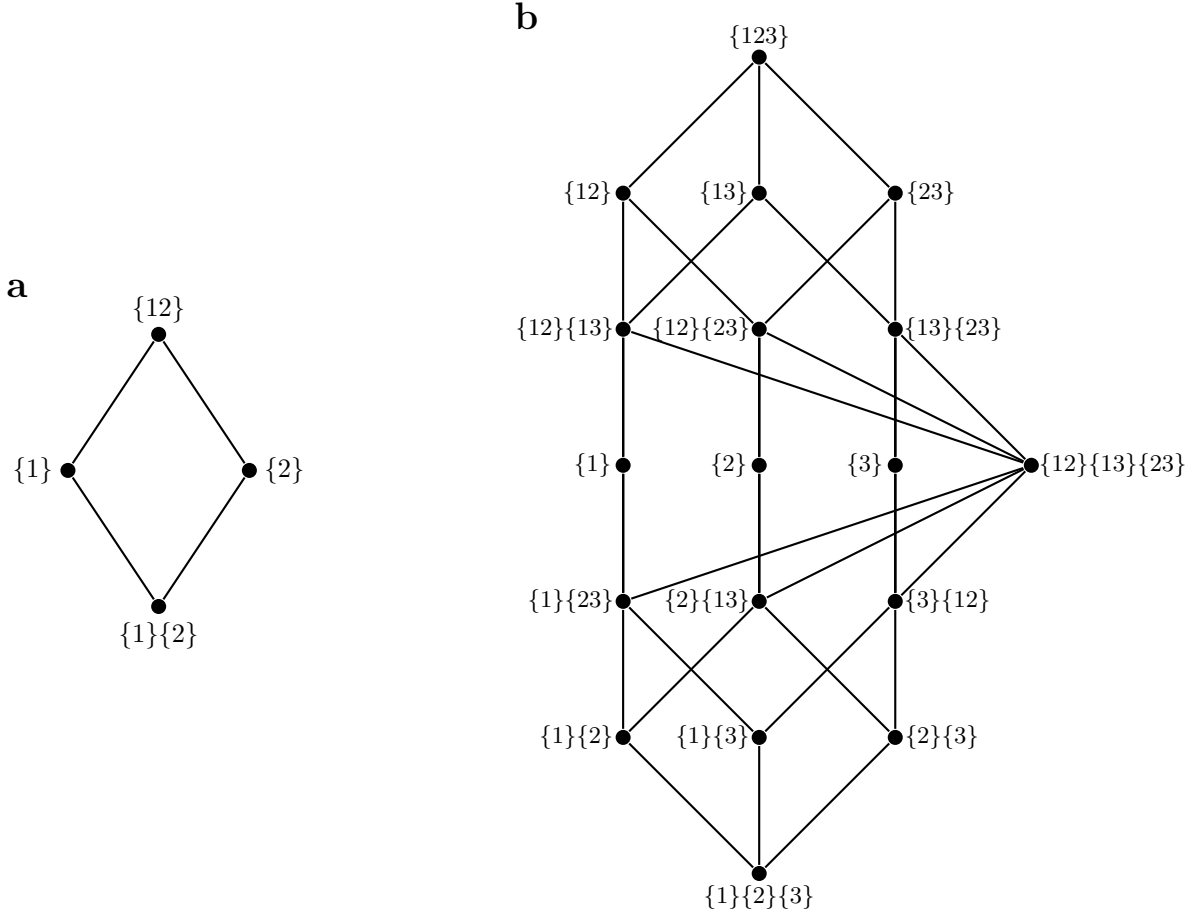
A similar idea was already introduced by Shannon to define an *ordering* between random variables [110]. Within this approach, a random variable  $X$  is more informative than  $Y$  if there exists a function  $f$  such that  $Y = f(X)$ . We indicate this as  $Y \leq_I X$ :

$$Y \leq_I X \iff \exists f \text{ s.t. } Y = f(X). \quad (12)$$

If  $f$  is bijective and can be inverted, then equality holds and we obtain the equivalence relation of Eq. (11).

However, there are many other ways to define an ordering between random variables [23, 111, 112]. Another fundamental approach is the one based on Blackwell sufficiency and operationalised by Blackwell's theorem [69]. Blackwell sufficiency is a key concept of statistical decision theory [69, 113], which formalises the idea that one random variable is at least as informative as another for any decision-making problem with respect to a given target, in the sense that any strategy based on the latter can be replicated, possibly after randomisation, using the former. This setup leads to Blackwell's theorem [69, 114], which provides the following definition: a random variable  $X_1$  is Blackwell sufficient for  $X_2$  relative to the target  $Y$  if there exists a stochastic channel  $k$  such that

$$p(x_2|y) = \sum_{x_1} k(x_2|x_1)p(x_1|y). \quad (13)$$



**FIG. 3:** Partial information decomposition lattices for **a)**  $n = 2$  and **b)**  $n = 3$  sources.

Hence, intuitively, since the channel  $p(x_2|y)$  can be obtained as a linear combination of  $p(x_1|y)$ , this operationalises the idea that  $X_2$  does not provide any more information than  $X_1$  about the target  $Y$ . We indicate this relation as  $X_2 \preceq_Y X_1$ . In other contexts, this condition is sometimes phrased by saying that the channel  $p(x_1|y)$  is stochastically degraded w.r.t.  $p(x_2|y)$  [39], or that  $X_2$  is a (degraded) garbling of  $X_1$  relative to  $Y$ . This operational reasoning has been the foundation of the Blackwell property (**BP**) and various PID definitions [21, 38–40, 67].

## A.2 Redundancy lattice and PID atoms

Partial information decomposition (PID) is a framework designed to refine the notion of multivariate information by distinguishing how multiple sources jointly inform a target. Rather than treating the mutual information  $I(X_1, \dots, X_n; Y)$  as a single quantity, PID separates it into components that reflect information shared across sources, exclusive to individual sources, or that emerges from the joint consideration of several sources. These correspond to the so-called redundancy, unique information, and synergy. In large systems, these contributions lead to a combinatorial structure of informational terms.

Williams and Beer introduced a sophisticated way to organise these components by exploiting an ordering of the redundancy relations inspired by the inclusion–exclusion principle (IEP). Given two sets  $X_1$  and  $X_2$ , the IEP states that the sizes of their union and intersection are not independent, being related as

$$|X_1 \cup X_2| = |X_1| + |X_2| - |X_1 \cap X_2|. \quad (14)$$

More generally, for  $n$  sets, the IEP reads

$$\left| \bigcup_{i=1}^n X_i \right| = \sum_{i=1}^n |X_i| - \sum_{1 \leq i < j \leq n} |X_i \cap X_j| + \sum_{1 \leq i < j < k \leq n} |X_i \cap X_j \cap X_k| - \cdots + (-1)^{n-1} |X_1 \cap \cdots \cap X_n|. \quad (15)$$

This setup enables a set-theoretic interpretation of information-theoretic quantities [115, 116] and, within PID, it defines a direct link between redundancy and synergy [10].

Hence, the key idea is to enumerate all meaningful ways in which information can be common to groups of sources, then arrange them within a partially ordered set. This construction gives rise to the so-called redundancy lattice, whose elements correspond to distinct notions of shared information.

Formally, the lattice is constructed from the following collections of sources:

$$\mathcal{A}(X_1, \dots, X_n) \equiv \alpha \in \mathcal{P}(\mathcal{P}(X_1, \dots, X_n)) \mid \forall A_i, A_j \in \alpha; A_i \not\subseteq A_j \quad (16)$$

where  $\mathcal{P}(\mathcal{X})$  denotes the set of all non-empty subsets of  $\mathcal{X}$ . Each element  $\alpha \in \mathcal{A}$  specifies a collection of source groups that define a candidate redundancy relationship, which we indicate as  $I_\cap(\alpha; Y)$ . The set  $\mathcal{A}$  is also in correspondence with all so-called *antichains* of the set  $X_1, \dots, X_n$  [117]. As such, for each antichain there exists a corresponding piece in the Partial Information Decomposition.

Furthermore, this set of antichains can then be endowed with a partial order given by

$$\alpha \preceq \beta \iff \forall B \in \beta \exists A \in \alpha \text{ s.t. } A \subseteq B, \quad (17)$$

with  $\alpha, \beta \in \mathcal{A}$ . Intuitively, this ordering reflects the idea that redundancy defined over smaller or more specific source groups refines redundancy defined over larger ones.

Throughout this work, we denoted with  $I_\cap$  the redundancy function, and as  $I_\partial$  the PID atoms obtained by the decomposition. Given a collection of sources  $\alpha$ , the fundamental construction of the redundancy lattice imposes that:

$$I_\cap(\alpha; Y) = \sum_{\beta \preceq \alpha} I_\partial(\beta; Y), \quad (18)$$

which then leads to the following recursive relation to find the PID atoms  $I_\partial$ :

$$I_\partial(\alpha; Y) = I_\cap(\alpha; Y) - \sum_{\beta \prec \alpha} I_\partial(\beta; Y). \quad (19)$$

To fix ideas, consider the collection  $\alpha = \{\{X_1\}, \{X_2, X_3\}, \{X_4, X_5\}\}$ , with a small abuse of notation, we also indicate this as  $\alpha = \{\{1\}, \{2, 3\}, \{4, 5\}\}$ . The corresponding redundancy function will be  $I_\cap(X_1, X_2X_3, X_4X_5; Y)$ , which indicates the shared information among the sets  $\{X_1\}, \{X_2, X_3\}$ , and  $\{X_4, X_5\}$ , where  $\{X_i, X_j\}$  also include the synergistic dependencies between  $X_i$  and  $X_j$ . Hence, the only purely redundant atom is the one following from the set  $\alpha = \{\{X_1\}, \dots, \{X_n\}\}$ , whose redundancy function is  $I_\cap(X_1, \dots, X_n; Y)$ .

We remark the difference between  $I_\cap(X_i, X_j; Y)$  and  $I_\cap(X_iX_j; Y)$ : the first is the redundancy function between the sources  $X_i$  and  $X_j$ , while the second is the redundancy between the composite variable  $(X_1X_2)$  and the target, and is equal to the mutual information  $I(X_1, X_2; Y)$  (by **(SR)**).

In the simplest case of  $n = 2$  sources, the PID lattice is composed of only four atoms (Fig. 3a), and the underlying equations directly read:

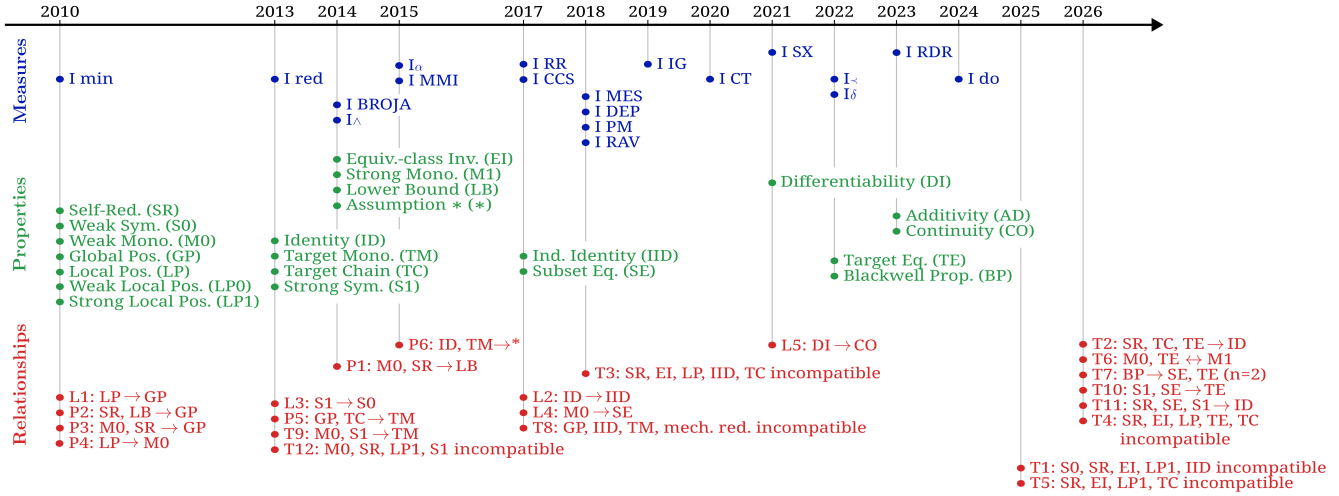
$$I_\cap(X_1, X_2; Y) = I_\partial(X_1, X_2; Y) \quad (20)$$

$$I(X_1; Y) = I_\partial(X_1, X_2; Y) + I_\partial(X_1; Y) \quad (21)$$

$$I(X_2; Y) = I_\partial(X_1, X_2; Y) + I_\partial(X_2; Y) \quad (22)$$

$$I(X_1, X_2; Y) = I_\partial(X_1, X_2; Y) + I_\partial(X_1; Y) + I_\partial(X_2; Y) + I_\partial(X_1X_2; Y). \quad (23)$$

Although PID is mathematically well-defined for any number of sources, performing the decomposition on systems with multiple sources becomes quickly infeasible, with the number of atoms increasing with the Dedekind number. For instance, already for  $n = 3$  sources there are 18 atoms (Fig. 3b), and 166 for  $n = 4$ .



**FIG. 4: Temporal timeline of the introduction of PID measures, properties, and their relationships.** \* refers to the inequality:  $I_{\cap}(X_1, X_2; f(X_1, X_2)) \leq I_{\cap}(X_1, X_2)$  (see Prop. 6).

As discussed in detail throughout this work, there are many approaches to defining the  $I_{\cap}(X_1, \dots, X_n; Y)$ , each characterised by distinct properties. In turn, such properties are linked via specific relationships and incompatibilities. We provide a visual overview of the historical development of PID in Fig. 4, indicating when each measure, property, and mutual relationship was introduced. Further information about the measures, properties, and interrelationships can be found in the following sections.

## B PID properties

- **(SR)**: Self-redundancy [10].

The redundancy of a single source equals its mutual information:

$$I_{\cap}(X_i; Y) = I(X_i; Y). \quad (24)$$

- **(S<sub>0</sub>)**: Weak Symmetry [10].

Redundancy is invariant under any permutation of the sources:

$$I_{\cap}(X_1, \dots, X_n; Y) = I_{\cap}(\sigma(X_1), \dots, \sigma(X_n); Y). \quad (25)$$

for any permutation  $\sigma$ .

- **(M<sub>0</sub>)**: Weak Monotonicity [10].

Adding a source to the redundancy function does not increase redundancy, and removing a source more informative than another preserves redundancy:

$$I_{\cap}(X_1, \dots, X_n; Y) \leq I_{\cap}(X_1, \dots, X_{n-1}; Y), \quad (26)$$

with equality if there exists  $Z \in \{X_1, \dots, X_{n-1}\}$  such that  $Z = f(X_n)$ , i.e.  $Z \leq_I X_n$ .

- **(GP)**: Global Positivity [10].

Redundancy is non-negative:

$$I_{\cap}(X_1, \dots, X_n; Y) \geq 0. \quad (27)$$

- **(LP<sub>0</sub>)**: Weak Local Positivity [10].

The atoms  $I_\partial$  are non-negative for a bivariate PID:

$$I_\partial(\alpha; Y) \geq 0, \quad (28)$$

where  $\alpha$  indicates any collection of subsets of  $n = 2$  sources, i.e.  $\alpha \in \mathcal{A}(X_1, X_2) = \{\{\{1\}, \{2\}\}, \{1\}, \{2\}, \{1, 2\}\}$ .

- **(LP<sub>1</sub>)**: Strong Local Positivity [10].

The atoms  $I_\partial$  are non-negative for a PID with  $n > 2$  sources:

$$I_\partial(\alpha; Y) \geq 0, \quad (29)$$

where  $\alpha$  indicates any collection of subsets of  $n \geq 3$  sources. If both **(LP<sub>0</sub>)** and **(LP<sub>1</sub>)** hold, then we denote it as just **(LP)**.

- **(IID)**: Independent Identity [26].

Independent sources provide vanishing redundancy about a target that is a copy of them:

$$I_\cap(X_1, X_2; X_1 X_2) = 0, \quad (30)$$

for  $X_1 \perp\!\!\!\perp X_2$ . If  $X_1$  and  $X_2$  are binary random variables, this corresponds to the TBC gate.

- **(ID)**: Identity [18].

The redundancy between two sources and a target that is a copy of them is equal to the mutual information between the sources:

$$I_\cap(X_1, X_2; X_1 X_2) = I(X_1; X_2). \quad (31)$$

If  $X_1$  and  $X_2$  are binary random variables, this corresponds to the TBC gate.

- **(TM)**: Target Monotonicity [19].

Adding a target does not decrease the value of redundancy:

$$I_\cap(X_1, \dots, X_n; Y_1 \dots Y_n) \geq I_\cap(X_1, \dots, X_n; Y_1 \dots Y_{n-1}). \quad (32)$$

- **(TC)**: Target Chain rule [19].

The chain rule property holds for the target variables in the redundancy function:

$$I_\cap(X_1, \dots, X_n; Y_1 Y_2) = I_\cap(X_1, \dots, X_n; Y_1) + I_\cap(X_1, \dots, X_n; Y_2 | Y_1), \quad (33)$$

with  $I_\cap(X_1, \dots, X_n; Y_2 | Y_1) = \sum_{y_1} p(y_1) I_\cap(X_1, \dots, X_n; Y_2 | y_1)$ .

- **(SE)**: Subset Equality [26], also denoted as *deterministic equality* by Kolchinsky [38].

Removing a source more informative than another preserves redundancy:

$$I_\cap(X_1, \dots, X_n; Y) = I_\cap(X_1, \dots, X_{n-1}; Y), \quad (34)$$

if there exists  $Z \in \{X_1, \dots, X_{n-1}\}$  such that  $Z = f(X_n)$ , i.e.  $Z \leq_I X_n$ .

- **(LB)**: Lower Bound [20].

A variable less informative than all sources is a lower bound for redundancy:

$$I_\cap(X_1, \dots, X_n; Y) \geq I(Q; Y) \quad \text{for } Q \leq_I X_i \quad \forall i = 1, \dots, n. \quad (35)$$

- **(TE)**: Target Equality [20, 38].

Adding the target to the sources does not change redundancy:

$$I_\cap(X_1, \dots, X_n; Y) = I_\cap(X_1, \dots, X_n, Y; Y). \quad (36)$$

- **(M<sub>1</sub>)**: Strong Monotonicity [20].

Adding a source to the redundancy function does not increase redundancy, and removing a source more informative than another source or the target preserves redundancy:

$$I_{\cap}(X_1, \dots, X_n; Y) \leq I_{\cap}(X_1, \dots, X_{n-1}; Y), \quad (37)$$

with equality if there exists  $Z \in \{X_1, \dots, X_{n-1}, Y\}$  such that  $Z = f(X_n)$ , i.e.  $Z \leq_I X_n$ .

- **(S<sub>1</sub>)**: Strong Symmetry [19].

Redundancy is invariant under any permutation of sources or target:

$$I_{\cap}(X_1, \dots, X_n; Y) = I_{\cap}(\sigma(X_1), \dots, \sigma(X_n); \sigma(Y)), \quad (38)$$

for any permutation  $\sigma$ .

- **(\*)**: Assumption (\*) of Bertschinger *et al.* [21].

Redundancy and unique information only depend on the marginal and pairwise distributions between sources and the target:

$$I_{\cap}(X_1, \dots, X_n; Y) \text{ is a function of only } p(Y), p(X_i, Y) \forall i = 1, \dots, n. \quad (39)$$

- **(BP)**: Blackwell Property [38, 39, 67].

The unique information of a source vanishes if and only if the source is Blackwell inferior to all others:

$$I_{\partial}(X_i; Y) = 0 \iff X_i \preceq_Y X_j \quad \forall X_i \in \{X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n\}. \quad (40)$$

- **(AD)**: Additivity [68].

The redundancy of two independent subsystems is equal to the sum of their redundancies:

$$\begin{aligned} & I_{\cap}(X_1, \dots, X_n, X'_1, \dots, X'_n, Z_1, \dots, Z_n, Z'_1, \dots, Z'_n; Y_1, \dots, Y_n, Y'_1, \dots, Y'_n) = \\ & = I_{\cap}(X_1, \dots, X_n, Z_1, \dots, Z_n; Y_1, \dots, Y_n) + I_{\cap}(X'_1, \dots, X'_n, Z'_1, \dots, Z'_n; Y'_1, \dots, Y'_n), \end{aligned} \quad (41)$$

where  $(X_1, \dots, X_n, Z_1, \dots, Z_n; Y_1, \dots, Y_n) \perp\!\!\!\perp (X'_1, \dots, X'_n, Z'_1, \dots, Z'_n; Y'_1, \dots, Y'_n)$ .

- **(CO)**: Continuity [68].

Redundancy is continuous under small changes in the probability distribution:

$$\|p(x_1, \dots, x_n) - p'(x_1, \dots, x_n)\|_p < \delta \implies |I_{\cap}^p(X_1, \dots, X_n; Y) - I_{\cap}^{p'}(X_1, \dots, X_n; Y)| < \varepsilon, \quad (42)$$

where we indicated with  $\|\cdot\|_p$  the standard  $p$ -norm.

- **(DI)**: Differentiability [35].

Redundancy is differentiable with respect to the probability distribution:

$$\lim_{\|p' - p\|_p \rightarrow 0} \frac{I_{\cap}^{p'}(X_1, \dots, X_n; Y) - I_{\cap}^p(X_1, \dots, X_n; Y)}{\|p' - p\|_p} = \nabla_p I_{\cap}^p(X_1, \dots, X_n; Y)^{\top} (p' - p) < \infty, \quad (43)$$

where  $\nabla_p I_{\cap}^p(X_1, \dots, X_n; Y)$  denotes the gradient of the redundancy with respect to the probability distribution  $p(x_1, \dots, x_n)$ .

- **(EI)**: Equivalence-class Invariance [20].

Redundancy is invariant under the substitution of a random variable with an informationally equivalent one. In other words, it is invariant under the equivalence relation  $\sim$  (Eq. (11)).

## C Relations between PID properties

In this section, we present a comprehensive catalogue of implications and incompatibilities among PID properties, together with their precise mathematical formulations and proofs. With the exception of Prop. 3, Theo. 11, and Theo. 4, all results reported here are minimal, in the sense that additional relations can be derived by combining them. We nevertheless state these findings explicitly, due to their historical significance and practical importance.

### C.1 Basic implications

**Lemma 1.**  $(LP) \implies (GP)$ .

**Lemma 2.**  $(ID) \implies (IID)$ .

**Lemma 3.**  $(S_1) \implies (S_0)$ .

**Lemma 4.**  $(M_0) \implies (SE)$ .

**Lemma 5.**  $(DI) \implies (CO)$ .

**Proposition 1** ([20]).  $(M_0), (SR) \implies (LB)$

*Proof.* If there exists  $Q$  s.t.  $Q \leq_I X_i \ \forall i = 1, \dots, n$ , then by applying the equality condition of  $(M_0)$  recursively we get

$$I_{\cap}(X_1, \dots, X_n; Y) \geq I_{\cap}(X_1, \dots, X_n, Q; Y) = \dots = I_{\cap}(Q; Y) = I(Q; Y), \quad (44)$$

where the last step follows from  $(SR)$ . ■

**Proposition 2** ([10]).  $(SR), (LB) \implies (GP)$ .

*Proof.* The proof directly follows from choosing  $f_i(X_i) = C \ \forall i = 1, \dots, n$  as constant functions in the definition of  $(LB)$ , which then implies

$$I_{\cap}(X_1, \dots, X_n; Y) \geq I_{\cap}(C; Y) = I(C; Y) = 0. \quad (45)$$

**Proposition 3** ([10]).  $(M_0), (SR) \implies (GP)$ .

*Proof.* This follows from Props. 1 and 2. ■

**Proposition 4** ([10, 65, 71]).  $(LP) \implies (M_0)$ .

*Proof.* It follows directly from the definition of the PID lattice, since:

$$I_{\cap}(\alpha; Y) = \sum_{\beta \preceq \alpha} I_{\cap}(\beta; Y), \quad (46)$$

and from  $(LP)$

$$I_{\cap}(\alpha; Y) \geq I_{\cap}(\beta; Y) \ \forall \alpha, \beta. \quad (47)$$

**Proposition 5** ([19]).  $(GP), (TC) \implies (TM)$

*Proof.* It follows immediately from

$$I_{\cap}(X_1, \dots, X_n; Y_1 Y_2) = I_{\cap}(X_1, \dots, X_n; Y_1) + I_{\cap}(X_1, \dots, X_n; Y_2 | Y_1) \geq I_{\cap}(X_1, \dots, X_n; Y_1). \quad (48)$$

■

## C.2 Main implications and incompatibilities

**Theorem 1** ([65]).  $(S_0)$ ,  $(SR)$ ,  $(EI)$ ,  $(LP_1)$ , and  $(IID)$ <sup>2</sup> are incompatible. ■

*Proof.* Theo. 1 of Ref. [65]. ■

**Theorem 2.**  $(SR)$ ,  $(TC)$ ,  $(TE) \implies (ID)$ .

*Proof.*

$$\begin{aligned}
 I_{\cap}(X_1, X_2; X_1 X_2) &= I_{\cap}(X_1, X_2; X_1) + I_{\cap}(X_1, X_2; X_2|X_1) \\
 &= I_{\cap}(X_2; X_1) + I_{\cap}(X_1; X_2|X_1) \\
 &= I(X_2; X_1) + I(X_1; X_2|X_1) \\
 &= I(X_1; X_2),
 \end{aligned} \tag{49}$$

where in the first, second, and third steps we used  $(TC)$ ,  $(TE)$ , and  $(SR)$ , respectively. ■

**Theorem 3** ([30]).  $(SR)$ ,  $(EI)$ ,  $(LP_0)$ ,  $(IID)$ <sup>3</sup>, and  $(TC)$  are incompatible.

*Proof.* Theo. 6 or Ref. [30]. ■

**Theorem 4.**  $(SR)$ ,  $(EI)$ ,  $(LP_0)$ ,  $(TE)$  and  $(TC)$  are not compatible.

*Proof.* The result directly follows from Theos. 2-3. ■

**Theorem 5** ([65]).  $(SR)$ ,  $(EI)$ ,  $(LP_1)$ , and  $(TC)$  are incompatible.

*Proof.* Theo. 2 of Ref. [65]. ■

**Theorem 6.**  $(M_0)$ ,  $(TE) \iff (M_1)$

*Proof.*  $(\implies)$  The inequality condition follows from  $(M_0)$ . For the equality, we only have to check the case where  $Y \leq_I W$ , as the rest is covered by  $(M_0)$ : ■

$$I_{\cap}(X_1, \dots, X_n, W; Y) = I_{\cap}(X_1, \dots, X_n, Y; Y) = I_{\cap}(X_1, \dots, X_n; Y) \tag{50}$$

where the first step follows from  $(M_0)$  and  $Y \leq_I W$ , and the second from  $(TE)$ .

$(\impliedby)$   $(M_0)$  follows directly from the definition of  $(M_1)$ .  $(TE)$  follows from  $(M_1)$  since the source  $Y$  is a trivial function of the target  $Y$ . ■

**Theorem 7.** For  $n = 2$  sources,  $(BP) \implies (SE)$ ,  $(TE)$ .

*Proof.*  $((BP) \implies (SE))$ :

Consider two random variables  $X_1, X_2$ . If  $X_1 = f(X_2)$  then for any  $Y$  we have  $X_1 \preceq_Y X_2$ , and from  $(BP)$  it follows  $I_{\partial}(X_1; Y) = 0$ , i.e.  $I_{\cap}(X_1, X_2; Y) = I_{\cap}(X_1; Y)$ , which is  $(SE)$ .

$((BP) \implies (TE))$ :

Similar to the reasoning above, if  $X_2 = Y$  then  $X_1 \preceq_Y X_2$ , and thus from  $(BP)$  we have  $I_{\partial}(X_1; Y) = 0$ , i.e.  $I_{\cap}(X_1, X_2; Y) = I_{\cap}(X_1; Y)$ , which is  $(TE)$ . ■

**Proposition 6** ([66]).  $(ID)$ ,  $(TM) \implies I_{\cap}(X_1, X_2; f(X_1, X_2)) \leq I_{\cap}(X_1; X_2)$

*Proof.*

$$I_{\cap}(X_1, X_2; f(X_1, X_2)) \leq I_{\cap}(X_1, X_2; X_1, X_2) = I(X_1; X_2), \tag{51}$$

where the first inequality comes from  $(TM)$  and choosing the function  $f$  to concatenate the sources. ■

**Theorem 8** ([27, 67]).  $(GP)$ ,  $(IID)$ ,  $(TM)$  and *mechanistic redundancy* are incompatible.

<sup>2</sup>Although the original stated  $(ID)$  in the hypotheses, their counterexample is more general and only requires  $(IID)$ .

<sup>3</sup>Although the original stated  $(ID)$  in the hypotheses, their counterexample is more general and only requires  $(IID)$ .

*Proof.* Consider two independent sources  $X_1 \perp\!\!\!\perp X_2$  and a redundancy measure that satisfies **(GP)**, **(IID)**, and **(TM)**. In this case, we can still use Prop. 6 for independent sources to obtain

$$I_{\cap}(X_1, X_2; f(X_1, X_2)) \leq I(X_1; X_2) = 0. \quad (52)$$

Because of **(GP)**, also the LHS needs to vanish for any  $f$  and therefore there is no mechanistic redundancy. ■

### C.3 Strong symmetry

**Theorem 9** ([19]).  $(M_0), (S_1) \implies (TM)$

*Proof.*

$$I(X_1, \dots, X_n; Y_1 Y_2) = I(X_1, \dots, X_{n-1}, Y_1, Y_2; X_n) \geq I(X_1, \dots, X_{n-1}, Y_1; X_n) = I(X_1, \dots, X_n; Y_1), \quad (53)$$

where the first and last step follow from **(S<sub>1</sub>)** and the second from **(M<sub>0</sub>)**. ■

**Theorem 10.**  $(S_1), (SE) \implies (TE)$ .

*Proof.*

$$\begin{aligned} I_{\cap}(X_1, \dots, X_n, Y; Y) &= I_{\cap}(X_1, \dots, X_{n-1}, Y, Y; X_n) \\ &= I_{\cap}(X_1, \dots, X_{n-1}, Y; X_n) \\ &= I_{\cap}(X_1, \dots, X_n; Y), \end{aligned} \quad (54)$$

where in the first and last step we used **(S<sub>1</sub>)**, and in the second **(SE)**. ■

**Theorem 11.**  $(SR), (SE), \text{ and } (S_1) \implies (ID)$ .

*Proof.*

$$\begin{aligned} I_{\cap}(X_1, X_2; X_1 X_2) &= I_{\cap}(X_1, X_1, X_2; X_2) \\ &= I_{\cap}(X_1, X_2; X_2) \\ &= I_{\cap}(X_1; X_2) \\ &= I(X_1; X_2), \end{aligned} \quad (55)$$

where, in order, we used **(S<sub>1</sub>)**, **(SE)**, **(TE)** (since it follows from Theo. 10), and **(SR)**. ■

**Theorem 12** ([19]).  $(M_0), (SR), (LP_1), \text{ and } (S_1)$  are incompatible.

*Proof.* See Theo. 1 of Ref. [19]. This can also be derived by combining Theo. 1 and Theo. 11. ■

## D PID measures

In this section, we provide more insights into the existing PID measures. We first present a more complete overview of the PID measures that have been proposed to date. Rather than following a chronological order, we now discuss the measures based on the philosophies behind their introductions and their similarities. Then, we report the mathematical definitions of the measures in a unified notation with a brief description for each.

### D.1 A long history of PID measures

Alongside the introduction of PID and the redundancy lattice, Williams and Beer proposed  $I_{\cap}^{\min}$ , the first redundancy function and one of the most widely known to date [10].  $I_{\cap}^{\min}$  builds on an intuitive notion of redundancy, for which shared information is the minimum information provided by the sources for each outcome of the target. Closely related to this philosophy is  $I_{\cap}^{\text{MMI}}$ , where the minimum is taken across all realisations [118].  $I_{\cap}^{\text{MMI}}$  was the first measure specifically designed for continuous (Gaussian) systems, which also reconciled  $I_{\cap}^{\min}$ ,  $I_{\cap}^{\text{red}}$ , and  $I_{\cap}^{\text{BROJA}}$  by showing that they yield the same decomposition in the particular case of a bivariate Gaussian PID with

univariate target. Due to its computational efficiency and simple interpretation,  $I_{\cap}^{\text{MMI}}$  has since become one of the most widely used redundancy measures, with applications ranging from neuroscience [12] to complex systems in general [93]. Interestingly, the first appearance of  $I_{\cap}^{\text{MMI}}$  was in the work by Bertschinger *et al.* [19] (there denoted by  $I_I$ ). Although it does not introduce any new PID measure, we mention this study as an influential analysis that shaped subsequent developments. In it, the authors proposed new PID axioms, including **(TM)** and **(TC)** (called left target monotonicity and left chain rule), and discussed the properties of the measures available at the time. However, a major shortcoming of both  $I_{\cap}^{\text{min}}$  and  $I_{\cap}^{\text{MMI}}$ , already noticed by Bertschinger *et al.* [19], is that they tend to overestimate redundancy by quantifying the same *amount* of information between the sources, as opposed to the same information [18, 19, 22]. With a specific empirical application to environmental time-series in mind, Goodwell and Kumar partially addressed this issue by proposing  $I_{\cap}^{\text{RR}}$  as a modified version of  $I_{\cap}^{\text{MMI}}$ , which provides zero redundant information for uncorrelated sources [25].

Motivated by this issue, Harder *et al.* introduced the identity property **(ID)**, resolving the apparent shortcoming of  $I_{\cap}^{\text{min}}$  and  $I_{\cap}^{\text{MMI}}$  on the Two-Bit Copy system. Consequently, they introduced the PID function  $I_{\cap}^{\text{red}}$ , estimating redundancy geometrically by defining redundancy as the projected information onto specific convex closures [18]. Following these developments, another early measure that satisfied **(ID)** was  $I_{\cap}^{\text{BROJA}}$ . Independently introduced by Bertschinger *et al.* and by Griffith and Koch [21, 22],  $I_{\cap}^{\text{BROJA}}$  was the first redundancy function backed by an explicit operational approach, explicitly acknowledging a specific functional dependency of redundant and unique information via Assumption **(\*)**, and that satisfied many of the PID axioms present in the literature at the time. Within  $I_{\cap}^{\text{BROJA}}$ , unique information defines it via a minimisation of the conditional mutual information over a constrained set of probability distributions, leveraging concepts of decision theory [21]. However, despite its interpretational clarity, this measure has been argued to overestimate redundancy by artificially inflating the correlation between the sources [26, 27, 29], as well as for relying on a decision-theoretic setup that was later argued to be too restrictive [26]. On a similar line of work, intending to define a maximum-entropy based redundancy function that satisfies Assumption **(\*)**, James *et al.* investigated  $I_{\cap}^{\text{MES}}$ , a redundancy measure closely related to  $I_{\cap}^{\text{BROJA}}$ , underlining the conceptual differences between performing maximum-entropy projections and directly minimising mutual information [29]. Although  $I_{\cap}^{\text{MES}}$  fails to satisfy the equality condition of **(M<sub>0</sub>)**, i.e. **(SE)**, thus providing highly counterintuitive results, it highlights how source–target marginals can still implicitly encode higher-order interactions. In this sense, it provides an interesting discussion for understanding how synergistic information may persist even under maximum-entropy projections. Exploiting maximum entropy models and studies on cybernetics and reconstructability analysis [73, 74], another PID measure that satisfies **(ID)** is  $I_{\cap}^{\text{DEP}}$  by James *et al.* [27].  $I_{\cap}^{\text{DEP}}$  is a measure of unique information that assesses the statistical dependencies via the construction of the so-called constraint lattice. In this framework, each level of the lattice introduces an additional constraint between the sources. The unique information of a source is then defined as the smallest change in mutual information between sources and target across all lattice edges involving the addition of the corresponding source-target constraint. Originally proposed only for discrete systems, this measure has then been defined also for Gaussian processes [28]. Although  $I_{\cap}^{\text{red}}$ ,  $I_{\cap}^{\text{BROJA}}$ ,  $I_{\cap}^{\text{MES}}$ , and  $I_{\cap}^{\text{DEP}}$  all satisfy **(ID)**, hence following the mechanistic intuition behind the TBC system [18], they are all limited to the bivariate case, and cannot be easily extended to antichains with three or more sources [18, 27, 29, 119].

A different line of work has seen the introduction of PID measures based on algebraic considerations. Within these approaches, redundancy is defined via the mutual information between the meet of the sources and the target variable. However, this approach entails choosing a notion of ordering between random variables, which might depend on the empirical application or operational interpretation behind the measure, hence leading to the introduction of several measures. Griffith *et al.* first proposed  $I_{\cap}^{\wedge}$ , a PID measure based on the Gács–Körner common information [120], where they employed the more informative partial relation  $\leq_I$  [22]. Although  $I_{\cap}^{\wedge}$  was the first measure that satisfied the highly desired **(TM)**, it was criticised for providing an overly strict bound on redundancy [23, 38], vanishing for all distributions with full support, and being insensitive to statistical correlations not captured by a common random variable [27]. Building on this approach, Griffith *et al.* then subsequently proposed  $I_{\cap}^{\alpha}$  [23], relaxing the constraints underlying  $I_{\cap}^{\wedge}$  by employing an ordering based on Markov chains and conditional independence [23, 38], at the cost of sacrificing **(TM)**. Along this line, Kolchinsky more recently proposed  $I_{\cap}^{\prec}$ , defining redundancy using the same algebraic foundation but employing under Blackwell order [38]. Through Blackwell’s theorem [69], this definition admits a clear operational interpretation in decision-theoretic terms, improves the previous algebraic formulations by using the “more informative [Blackwell] ordering relation”, and provides an explicit

intuition in set-theoretic terms [38]. Moreover, in this work, the author advocates against the inclusion-exclusion principle, introducing the union information as the complementary notion of redundancy, arguing that it is required to define synergy. These algebraic formulations provide an elegant mathematical groundwork to define the notion of shared information, bridging concepts of information theory and set theory. Although this approach ensures desirable properties such as **(IID)**, it also leads to the rejection of **(LP)**, complicating the interpretation of negative PID atoms.

Proceeding along the Blackwell ordering perspective, Venkatesh and Schamberg proposed  $I_{\cap}^{\delta}$ , a PID formulation designed for multivariate Gaussian systems [39]. Arguing that the result obtained by Barrett [24] with  $I_{\cap}^{\text{MMI}}$  in bivariate Gaussian distributions with univariate target is a trivial case of Blackwell ordering, they introduced a measure of how distant each source is from being Blackwell superior to the other in the general multivariate case. Naming such a distance *deficiency*, they then employed it to define redundancy in a manner analogous to  $I_{\cap}^{\text{MMI}}$ . Although  $I_{\cap}^{\delta}$  reduces to  $I_{\cap}^{\text{MMI}}$  in the case of a univariate target, it is interesting that it satisfies **(ID)**, unlike  $I_{\cap}^{\text{MMI}}$ . Despite this approach being more general than the one obtained with  $I_{\cap}^{\text{MMI}}$ , it is only defined for a bivariate PID. Finally, continuing along the Blackwell ordering paradigm, Mages *et al.* proposed  $I_{\cap}^{\text{RDR}}$ , a redundancy measure grounded on the set-overlap of the reachable decision regions of the sources [40]. Such an overlap can be obtained by transforming the problem into a decision-making task for each realisation of the target.  $I_{\cap}^{\text{RDR}}$  yields a pointwise decomposition that nevertheless satisfies **(M<sub>0</sub>) (GP)** and also **(LP)** for any number of sources, hence violating **(ID)** in the process. Moreover, in the specific scenarios where one of the sources is pointwise Blackwell superior to the other, or if the target is binary,  $I_{\cap}^{\text{RDR}}$  reduces to  $I_{\cap}^{\text{min}}$  and  $I_{\cap}^{\text{BROJA}}$ , respectively.

A conceptually distinct line of work has focused on the definition of pointwise PID measures. Starting from a fundamentally different premise, Ince proposed the first fully pointwise PID decomposition  $I_{\cap}^{\text{CCS}}$ , in which redundant information is defined at the local level via the pointwise coinformation [26]. Specifically, redundancy is identified with the *common change in surprisal* induced by the sources about the target: when the pointwise information contributions from all sources share the same sign, the pointwise coinformation captures the overlap between the individual pointwise information values, providing a local, set-theoretic notion of redundancy. In this process,  $I_{\cap}^{\text{CCS}}$  departs from several global PID properties, such as **(M<sub>0</sub>)** and **(GP)**, while instead introducing and satisfying **(IID)** and **(SE)**. Shortly after, Finn and Lizier proposed another pointwise PID definition,  $I_{\cap}^{\text{PM}}$ , based on a double decomposition over specificity and ambiguity lattices [30,72].  $I_{\cap}^{\text{PM}}$  focuses on distinguishing *informative* information, which reduces uncertainty about the target, from *misinformative* information, which instead increases it. This approach renounces global **(M<sub>0</sub>)** and **(GP)** in favour of properties such as **(TC)**, with **(M<sub>0</sub>)**, **(GP)**, and **(LP)** still holding locally and separately for the specificity and ambiguity lattice. Following a similar philosophy, Makkeh *et al.* introduced the measure  $I_{\cap}^{\text{SX}}$ , first in discrete systems [35], and subsequently in continuous scenarios [36,37].  $I_{\cap}^{\text{SX}}$  measures redundancy by assessing the portion of probability mass associated with a particular target realisation that can be jointly excluded by all sources together. Similarly to  $I_{\cap}^{\text{PM}}$ ,  $I_{\cap}^{\text{SX}}$  is based on separate informative and misinformative lattices, both singularly satisfying **(M<sub>0</sub>)**, **(GP)** and **(LP)**, though it loses these properties upon recombination into the global redundancy measure. Although negative PID atoms are often avoided – making **(LP)** one of the most sought-after properties – pointwise approaches not only argue that negative values are not problematic, but they actively embrace them. In fact, the presence of negative atoms would indicate the presence of mechanistic redundancy within  $I_{\cap}^{\text{CCS}}$ , [26], or scenarios in which the sources are providing misinformative information about the target for  $I_{\cap}^{\text{PM}}$  and  $I_{\cap}^{\text{SX}}$  [30,35].

Additional definitions of redundant information have then exploited unexplored ideas. Inspired by the hierarchical decomposition of Amari [75], Niu and Quinn proposed  $I_{\cap}^{\text{IG}}$ , a redundancy function based on concepts of information geometry. Within  $I_{\cap}^{\text{IG}}$ , unique information and synergy are obtained by projecting the full probability distribution onto specific submanifolds characterised by vanishing interaction terms [31].  $I_{\cap}^{\text{IG}}$  was originally defined for a bivariate discrete PID with full-support distributions, and was later generalised to the continuous Gaussian case by Kay [32]. Sigtermans proposed the redundancy measure  $I_{\cap}^{\text{CT}}$  based on the framework of Causal Tensors – multivariate stochastic functions that map sources into the target via distinct transmission paths [33,34]. Unlike previous redundancy measures,  $I_{\cap}^{\text{CT}}$  has been advanced to specifically account for the indirect mediated associations, explicitly representing how information is transmitted through cascades of stochastic channels. A different approach led to the introduction of  $I_{\cap}^{\text{do}}$  by Lyu *et al.* [41]. This framework adopts an explicitly interventional perspective inspired by causal inference, where unique information is defined employing Pearl’s *do-calculus* [76–78]. Specifically, unique information is expressed as a conditional mutual information evaluated under a do-operation.

Although  $I_{\cap}^{\text{do}}$  is among the few measures that satisfy both **(CO)** and **(AD)**, it does not satisfy **(M<sub>0</sub>)** since it lacks **(SE)**, leading to a non-vanishing unique information even in the case when the sources are identical.

Finally, in this work, we also considered  $I_{\cap}^{\text{RAV}}$ , an unpublished PID measure that quantifies redundancy by maximising the coinformation between sources, target, and an arbitrary function of the sources, taken over all possible such functions [79].

## D.2 Mathematical definitions

Here we list the formal PID definitions considered in this work, ordered chronologically by their appearance in the literature.

### Williams and Beer’s Minimum Information ( $I_{\cap}^{\text{min}}$ )

This was the first PID measure proposed [10]. It defines redundancy as the minimum information any individual source provides about the target, calculated pointwise for each outcome of the target. It has been criticised for not distinguishing between “whether different random variables carry the same information or just the same amount of information”, thus sometimes overestimating redundancy [18, 19, 23].

$$I_{\cap}^{\text{min}}(X_1, \dots, X_n; Y) = \sum_y p(y) \min_{i=1, \dots, n} I(X_i; Y = y). \quad (56)$$

### Harder *et al.*’s Projected Information ( $I_{\cap}^{\text{red}}$ )

$I_{\cap}^{\text{red}}$  defines redundancy in terms of information geometry [18]. Redundant information is the part of the information that can be represented by projecting one source’s conditional distributions onto the convex closure of the other’s. It is specifically defined for the bivariate case, and it avoids some of the overestimation issues of  $I_{\cap}^{\text{min}}$ . It was the first redundancy measure that satisfied **(ID)**.

$$\pi_A(p) = \arg \min_{r \in A} D_{\text{KL}}(p \| r), \quad (57)$$

$$C_{\text{cl}}(A) = \{\lambda p + (1 - \lambda)q \mid p, q \in A, \lambda \in [0, 1]\} \quad (58)$$

$$\langle X_i \rangle_Y = \{p(Y|x_i) \text{ s.t. } x_i \in \mathcal{X}_i\} \quad (59)$$

$$p_{(x_i \searrow X_j)}(Y) = \pi_{C_{\text{cl}}(\langle X_j \rangle_Y)}(p(Y|x_i)), \quad (60)$$

$$I_Y^{\pi}(X_1 \searrow X_2) = \sum_{x_1, y} p(x_1, y) \log \frac{p_{(x_1 \searrow X_2)}(y)}{p(y)}, \quad (61)$$

$$I_{\cap}^{\text{red}}(X_1, X_2; Y) = \min \{I_Y^{\pi}(X_1 \searrow X_2), I_Y^{\pi}(X_2 \searrow X_1)\}, \quad (62)$$

where  $\pi_{C_{\text{cl}}(\langle X_1 \rangle_Y)}(q)$  denotes the information projection of a distribution  $q$  onto the convex closure of the set of conditional distributions of  $Y$  given  $X_1$ .

### Griffith *et al.*’s Common Information Redundancy ( $I_{\cap}^{\wedge}$ )

The “intersection information”  $I_{\cap}^{\wedge}$  defines redundancy as the maximum mutual information obtainable from a variable  $Q$  that can be obtained from any source via a deterministic function [20]. It was the first proposed measure that satisfied **(TC)**, serving as “a lower bound on any reasonable redundancy measure” [20]:

$$I_{\cap}^{\wedge}(X_1, \dots, X_n; Y) = \max_Q I(Q; Y) \quad \text{such that } H(Q|X_i) = 0 \forall i \in 1, \dots, n. \quad (63)$$

### Griffith *et al.*'s Relaxed Common Information Redundancy ( $I_{\cap}^{\alpha}$ )

$I_{\cap}^{\alpha}$  defines redundancy as the maximal information obtainable from a variable  $Q$  that does not increase the mutual information between each source and the target [23]. This formalism loosens the constraints given by  $I_{\cap}^{\wedge}$  and provides a tighter bound on a plausible measure of redundancy, as  $I_{\cap}^{\wedge} \leq I_{\cap}^{\alpha}$ .

$$I_{\cap}^{\alpha}(X_1, \dots, X_n; Y) = \max_Q I(Q; Y) \quad \text{such that } I(Q, X_i; Y) = I(X_i; Y) \quad \forall i \in 1, \dots, n. \quad (64)$$

### BROJA/Griffith and Koch's Constrained Optimisation ( $I_{\cap}^{\text{BROJA}}$ )

$I_{\cap}^{\text{BROJA}}$  defines redundancy via constrained optimisation over joint distributions [21, 22]. The idea is to find all distributions that preserve each source-target marginal, then define redundancy as the maximum coinformation in these distributions. It was independently proposed by Bertschinger *et al.* [21] and Griffith and Koch [22], and admits an operational interpretation from decision theory. It is widely used in empirical applications due to computability and intuitiveness, though it is only defined for  $n = 2$  sources.

$$I_{\cup}^{\text{BROJA}}(X_1, X_2; Y) = I(X_1, X_2; Y) - \min_{Q \in \Delta_P} I_Q(X_1, X_2; Y) \quad (65)$$

$$I_{\cap}^{\text{BROJA}}(X_1, X_2; Y) = \max_{Q \in \Delta_P} I_Q(X_1, X_2; Y), \quad (66)$$

where  $\Delta_P = \{Q \in \Delta \mid P(X_i; Y) = Q(X_i; Y), i = 1, 2\}$ , and  $I_Q$  refers to the mutual information evaluated on the distribution  $Q$ .

### Barrett's Minimum Mutual Information ( $I_{\cap}^{\text{MMI}}$ )

Minimal Mutual Information ( $I_{\cap}^{\text{MMI}}$ ) defines redundancy as the minimum mutual information between any individual source and the target [24]. Therefore, it captures the smallest guaranteed contribution from any source, analogous to  $I_{\cap}^{\min}$ , but applied directly at the global level of mutual information, rather than pointwise. As such, it serves as the upper bound for any redundancy measure. Its simplicity makes it easy to implement and particularly popular in empirical applications.

$$I_{\cap}^{\text{MMI}}(X_1, \dots, X_n; Y) = \min_{i=1, \dots, n} I(X_i; Y). \quad (67)$$

### Goodwell and Kumar's Rescaled Redundancy ( $I_{\cap}^{\text{RR}}$ )

Rescaled Redundancy ( $I_{\cap}^{\text{RR}}$ ) was introduced as a rescaled version of  $I_{\cap}^{\text{MMI}}$  so that it would provide zero redundancy for independent sources [25]. It achieves this by interpolating between the  $I_{\cap}^{\text{MMI}}$  and a positive coinformation, weighted by the dependence between sources.  $I_{\cap}^{\text{RR}}$  was designed for the study of dynamical processes where it is desired that independent sources should not carry redundancy [25].

$$R_{\min}(X_1, X_2, Y) = \max(0, I(X_1; X_2; Y)) \quad (68)$$

$$R_{\text{MMI}}(X_1, X_2, Y) = \min(I(X_1; Y), I(X_2; Y)) \quad (69)$$

$$I_s(X_1, X_2) = \frac{I(X_1; X_2)}{\min(H(X_1), H(X_2))} \quad (70)$$

$$I_{\cap}^{\text{RR}}(X_1, X_2; Y) = R_{\min}(X_1, X_2, Y) + I_s(X_1, X_2)(R_{\text{MMI}}(X_1, X_2, Y) - R_{\min}(X_1, X_2, Y)). \quad (71)$$

### Ince's Common Change in Surprisal ( $I_{\cap}^{\text{CCS}}$ )

Common Change in Surprisal ( $I_{\cap}^{\text{CCS}}$ ) defines redundancy based on coherent changes in surprisal [26]. A set of sources is redundant if they all increase or decrease the surprisal of the target in the same direction for a given outcome on a constrained maximum entropy distribution. Hence, it captures redundancy as the pointwise coinformation

between all sources. The overall redundancy is then computed via the expectation value of the pointwise change in surprisal.

$$\hat{P}(X_1, \dots, X_n, Y) = \arg \max_{Q \in \Delta_P} H_Q(X_1, \dots, X_n, Y) \quad (72)$$

$$\Delta_P = \{Q \in \Delta \mid Q(X_i, Y) = P(X_i, Y), i = 1, \dots, n, \text{ and } Q(X_1, \dots, X_n) = P(X_1, \dots, X_n)\} \quad (73)$$

$$\Delta_s h^{\text{com}}(x_1, \dots, x_n, y) = \begin{cases} c_{\hat{p}}(x_1, \dots, x_n, y), & \text{if } \text{sgn } \Delta_s h(x_1) = \dots = \text{sgn } \Delta_s h(x_n) \\ & = \text{sgn } \Delta_s h(x_1, \dots, x_n) = \text{sgn } c(x_1, \dots, x_n, y), \\ 0, & \text{otherwise.} \end{cases} \quad (74)$$

$$I_{\cap}^{\text{CCS}}(X_1, \dots, X_n; Y) = \sum_{x_1, \dots, x_n, y} \hat{p}(x_1, \dots, x_n, y) \Delta_s h^{\text{com}}(x_1, \dots, x_n, y), \quad (75)$$

where  $c_q$  indicates the pointwise coinformation evaluated on the distribution  $q$ .

### James *et al.*'s Unique Information via Dependency Constraints ( $I_{\cap}^{\text{DEP}}$ )

Unique information via Dependency Constraints ( $I_{\cap}^{\text{DEP}}$ ) defines the unique information of each source using maximum entropy projections [27]. Specifically,  $I_{\cap}^{\text{DEP}}$  requires the construction of a constraint lattice where each edge represents the addition of a constraint between a source and the target. Unique information is then computed as the minimal information gained when the source-target constraint is added. Since this decomposition stems from unique information, it provides a complete definition of the lattice only for  $n = 2$  sources.

$$I_{\partial}(X_i; Y) = \min\left(I_{X_i Y: X_j Y}(X_i; Y|X_j), I_{X_i X_j: X_i Y: X_j Y}(X_i; Y|X_j)\right), \quad (76)$$

where  $I_{A:B:C}(\cdot)$  indicates that the mutual information is calculated on the maximum entropy distribution with constraint  $p(A) = q(A)$ ,  $p(B) = q(B)$ ,  $p(C) = q(C)$ . A continuous version of  $I_{\cap}^{\text{DEP}}$  was subsequently introduced for Gaussian systems by Kay and Ince [28].

### James *et al.*'s Maximum Entropy Star ( $I_{\cap}^{\text{MES}}$ )

Maximum Entropy Star ( $I_{\cap}^{\text{MES}}$ ) defines redundancy as the coinformation in the maximum entropy distribution that preserves each source's marginal with the target [29]. Its definition follows  $I_{\cap}^{\text{BROJA}}$ 's approach, where, instead of minimising the mutual information, entropy is maximised under the constraints given by Assumption (\*) [21].

$$Q = \arg \max_{Q' \in \Delta_P} H_{Q'}(X_1, X_2, Y) \quad (77)$$

$$I_{\cup}^{\text{MES}}(X_1, X_2; Y) = I(X_1, X_2; Y) - I_Q(X_1, X_2; Y) \quad (78)$$

$$I_{\cap}^{\text{MES}}(X_1, X_2; Y) = I_Q(X_1; X_2; Y), \quad (79)$$

where  $\Delta_P = \{Q \in \Delta \mid P(X_i; Y) = Q(X_i; Y), i = 1, 2\}$ .

### Finn and Lizier's Specificity and Ambiguity ( $I_{\cap}^{\text{PM}}$ )

Plus-Minus ( $I_{\cap}^{\text{PM}}$ ) redundancy operates at the level of individual realisations of the sources and target. It separates the positive and negative contributions to local surprisal, defining a specificity and ambiguity lattice based on informative and misinformative exclusions [72]. Specifically, redundant specificity and ambiguity are associated with the source event which induces the smallest total or misinformative exclusions, respectively [30]. Finally, the

overall redundancy can then be computed via an expectation value.

$$i_{\cap}^{\text{PM}+}(x_1, \dots, x_n; y) = \min_{i=1, \dots, n} h(x_i) \quad (80)$$

$$i_{\cap}^{\text{PM}-}(x_1, \dots, x_n; y) = \min_{i=1, \dots, n} h(x_i|y) \quad (81)$$

$$i_{\cap}^{\text{PM}}(x_1, \dots, x_n; y) = i_{\cap}^{\text{PM}+}(x_1, \dots, x_n; y) - i_{\cap}^{\text{PM}-}(x_1, \dots, x_n; y), \quad (82)$$

$$I_{\cap}^{\text{PM}}(X_1, \dots, X_n; Y) = \sum_{x_1, \dots, x_n, y} p(x_1, \dots, x_n, y) i_{\cap}^{\text{PM}}(x_1, \dots, x_n; y). \quad (83)$$

### Niu and Quinn's Information-Geometric Redundancy ( $I_{\cap}^{\text{IG}}$ )

Information Geometry ( $I_{\cap}^{\text{IG}}$ ) uses an information-geometric approach inspired by Amari's Total Correlation hierarchical decomposition [75], providing a directed decomposition of mutual information [31]. In this scenario, synergy is defined by the distance between the joint distribution and the geodesic that passes through the points of conditional independence of the original distribution. Unique information is then defined as the distance between such a projection and the points of conditional independence. The decomposition is only defined for  $n = 2$  sources.

$$\bar{P}_1 := \arg \min_{Q \in S_1} D(P_{X_1 X_2 Y} \| Q) = \frac{P_{X_1 Y} P_{X_1 X_2}}{P_{X_1}} \quad (84)$$

$$\bar{P}_2 := \arg \min_{Q \in S_2} D(P_{X_1 X_2 Y} \| Q) = \frac{P_{X_2 Y} P_{X_1 X_2}}{P_{X_2}}. \quad (85)$$

$$t^* = \arg \min_{t \in \mathbb{R}} D(P_{X_1 X_2 Y} \| \mathcal{N} \bar{P}_1^t \bar{P}_2^{(1-t)}) \quad (86)$$

$$P^* = \mathcal{N} \bar{P}_1^{t^*} \bar{P}_2^{(1-t^*)} \quad (87)$$

$$I_{\partial}(X_1 X_2; Y) = D(P_{X_1 X_2 Y} \| P^*) \quad (88)$$

$$\begin{aligned} I_{\cap}^{\text{IG}}(X_1, X_2; Y) &= D(\bar{P}_1 \| P_{X_1 X_2 Y}) - D(P^* \| \bar{P}_2), \\ &= D(\bar{P}_2 \| P_{X_1 X_2 Y}) - D(P^* \| \bar{P}_1) \end{aligned} \quad (89)$$

where  $S_i$  is a submanifold with specific coupling parameters set to zero, and  $\mathcal{N}$  is a normalisation constant. Parametrising the full manifold as  $M = \{P_{X_1 X_2 Y} \text{ s.t. } P_{X_1 X_2 Y} = \exp(\theta_1 F_1(x_1) + \theta_2 F_2(x_2) + \theta_3 F_3(y) + \theta_{12} F_{12}(x_1, x_2) + \theta_{13} F_{13}(x_1, y) + \theta_{23} F_{23}(x_2, y) + \theta_{123} F_{123}(x_1, x_2, y) - \psi(\theta))\}$ , we define  $S_i = \{P \in M \text{ s.t. } \theta_{j3} = \theta_{123} = 0\}$ ,  $i, j = 1, 2$ .

A continuous version of  $I_{\cap}^{\text{IG}}$  was subsequently introduced for (bivariate) Gaussian systems by Kay [32].

### Sigtermans' Redundancy via Causal Tensors ( $I_{\cap}^{\text{CT}}$ )

Causal Tensor ( $I_{\cap}^{\text{CT}}$ ) defines redundancy using the framework of Causal Tensors, i.e. multilinear stochastic maps that transform source data into destination data, enabling one to express an indirect association in terms of the direct, constituent associations [33,34].  $I_{\cap}^{\text{CT}}$  compares the information transmitted from a source to the target across all possible transmission paths (cascades), and defines redundancy as the weakest indirect path that includes all sources and ends at the target. For two variables, this corresponds to the minimum of the two directional cascades:

$$I_{P_{X_1-Y}}\{X_1\}\{X_2\}\{Y\} = \sum_{ijk} p^{ijk} \log \frac{\sum_l A_i^l B_l^k}{p^k} = \sum_{x_1 x_2 y} p(x_1, x_2, y) \log \frac{\sum_{x_2'} p(x_2'|x_1) p(y|x_2')}{p(y)} \quad (90)$$

$$I_{\cap}^{\text{CT}}(X_1, X_2; Y) = \min\left(I_{P_{X_1-Y}}\{X_1\}\{X_2\}\{Y\}, I_{P_{X_2-Y}}\{X_2\}\{X_1\}\{Y\}\right), \quad (91)$$

where the subscript  $P_{X_i-Y}$  indicates that the path information is calculated on the probability distribution induced by the cascade  $X_i \rightarrow X_j \rightarrow Y$ .

### Makkeh *et al.*'s Shared Exclusion ( $I_{\cap}^{\text{SX}}$ )

Shared Exclusion ( $I_{\cap}^{\text{SX}}$ ) is a pointwise PID measure that assigns redundancy to the shared exclusion of probability mass [35–37]. In other words, redundancy is defined as the amount of information that can be inferred when the joint realisation of all sources excludes a particular realisation of the target.  $I_{\cap}^{\text{SX}}$  decomposes the information into an informative and a misinformative part, which can then be combined to obtain the overall redundancy.

$$i_{\cap}^{\text{SX}+}(x_1, \dots, x_n; y) = \log_2 \frac{1}{P(x_1 \cup x_2 \cup \dots \cup x_n)}, \quad (92)$$

$$i_{\cap}^{\text{SX}-}(x_1, \dots, x_n; y) = \log_2 \frac{P(y)}{P(y \cap (x_1 \cup x_2 \cup \dots \cup x_n))} \quad (93)$$

$$i_{\cap}^{\text{SX}}(x_1, \dots, x_n; y) = i_{\cap}^{\text{SX}+}(x_1, \dots, x_n; y) - i_{\cap}^{\text{SX}-}(x_1, \dots, x_n; y), \quad (94)$$

$$I_{\cap}^{\text{SX}}(X_1, \dots, X_n; Y) = \sum_{x_1, \dots, x_n, y} p(x_1, \dots, x_n, y) i_{\cap}^{\text{SX}}(x_1, \dots, x_n; y). \quad (95)$$

Originally defined only for discrete distributions, this approach has then been generalised to continuous settings [36, 37].

### Kolchinsky's Blackwell Redundancy ( $I_{\cap}^{\prec}$ )

$I_{\cap}^{\prec}$  defines redundancy as the maximum information obtainable from a variable  $Q$  that is Blackwell inferior to all the sources [38]. As such, it is a more liberal definition than  $I_{\wedge, \alpha}$  since  $I_{\cap}^{\wedge} \leq I_{\cap}^{\alpha} \leq I_{\cap}^{\prec} \leq I_{\cap}^{\text{MMI}}$ .  $I_{\cap}^{\prec}$  is based on an operational interpretation based on utility maximisation in decision theory. Importantly, in the proposed decomposition, the author rejects the inclusion-exclusion principle and defines the union information needed to complete the lattice decomposition [38].

$$I_{\cap}^{\prec}(X_1, \dots, X_n; Y) = \min_Q I(Q; Y) \quad \text{such that } Q \succeq_Y X_i \forall i \in 1, \dots, n \quad (96)$$

$$I_{\cap}^{\prec}(X_1, \dots, X_n; Y) = \max_Q I(Q; Y) \quad \text{such that } Q \preceq_Y X_i \forall i \in 1, \dots, n. \quad (97)$$

### Venkatesh and Schamberg's Deficiency Redundancy ( $I_{\cap}^{\delta}$ )

$I_{\cap}^{\delta}$  employs statistical deficiencies to generalise  $I_{\cap}^{\text{MMI}}$ 's definition and provide a Blackwellian PID for multivariate Gaussian systems [39]. This is achieved by removing from the marginal mutual information the extent to which one source is distant from being Blackwell inferior to the other source, i.e. the *deficiency*  $\delta$ . This approach is only defined for  $n = 2$  source variables.

$$(P_{A|B} \circ P_{B|C})(a|c) = \int db P(A = a|B = b)P(B = b|C = c) \quad (98)$$

$$\delta(Y; X_i \setminus X_j) = \inf_{P(X_i|X_j)} \mathbb{E}_{P_Y} [D(P_{X_i|Y} || P_{X_i|X_j} \circ P_{X_j|Y})] \quad (99)$$

$$I_{\cap}^{\delta}(X_1, X_2; Y) = \min \left( I(X_1; Y) - \delta(Y; X_1 \setminus X_2), I(X_2; Y) - \delta(Y; X_2 \setminus X_1) \right). \quad (100)$$

### Mages and Rohner's Decision Reachability Redundancy ( $I_{\cap}^{\text{RDR}}$ )

Reachable Decision Region ( $I_{\cap}^{\text{RDR}}$ ) is a redundancy definition based on a direct decision-making viewpoint [40]. For every possible outcome of the target  $y$ , it defines a one-vs-rest variable  $Y^y = y$ , and for every source, it considers its “reachable decision region”, i.e. the entire set of performance trade-offs you can achieve using that source alone. Hence, since these regions contain all true/false-positive pairs that any decision rule based on that source can reach, they capture the pointwise uncertainty of each source about  $Y = y$ . Geometrically, these structures correspond to the zonotopes [121]. Using these regions,  $I_{\cap}^{\text{RDR}}$  then expresses redundancy as their overlap, i.e. the part of the

decision region that is reachable using any of the sources on their own.

$$Y^y = \begin{cases} y & \text{if } Y = y \\ 1 - y & \text{otherwise} \end{cases} \quad (101)$$

$$i_{\cap}^{\text{RDR}}(X_i; Y^y) = \sum_{x_i} P(X_i = x_i | Y^y = y) \log \frac{P(X_i = x_i | Y^y = y)}{P(Y^y = y) P(X_i = x_i | Y^y = y) + (P(Y^y \neq y)) P(X_i = x_i | Y^y \neq y)} \quad (102)$$

$$i_{\cap}^{\text{RDR}}(X_i \cup \beta; Y^y) = i_{\cap}^{\text{RDR}}(X_i; Y^y) + i_{\cap}^{\text{RDR}}(\beta; Y^y) - i_{\cap}^{\text{RDR}}(\{X_i \sqcup_{Y^y} X_j | X_j \in \beta\}; Y^y) \quad (103)$$

$$I_{\cap}^{\text{RDR}}(\alpha; Y) = \sum_y p(y) i_{\cap}^{\text{RDR}}(\alpha; Y^y), \quad (104)$$

where  $\alpha, \beta$  are collection of subsets of sources, and  $X_i \sqcup_{Y^y} X_j$  indicates the union under Blackwell order w.r.t. the target  $Y^y$ .

### Lyu *et al.*'s do-calculus Redundancy ( $I_{\cap}^{\text{do}}$ )

$I_{\cap}^{\text{do}}$  ([41]) is grounded on the interventional approach of the do-calculus proposed by Pearl [76, 77, 122]. Within this framework, unique information is defined as the conditional mutual information between an intervened source variable and the target. Since the decomposition is based on a definition of unique information, this approach fully determines all PID atoms only for  $n = 2$  sources.

$$p(X'_i = x_i, Y = y | X_j = x_j) = p(X_i = x_i | Y = y) p(Y = y | X_j = x_j) \quad (105)$$

$$I_{\partial}(X_i; Y) = I(X'_i; Y | X_j) \quad (106)$$

$$I_{\cap}^{\text{do}}(X_1, X_2; Y) = I(X'_1; X_2) = I(X'_2; X_1). \quad (107)$$

### Redundancy via Auxiliary Variable ( $I_{\cap}^{\text{RAV}}$ )

Redundancy via Auxiliary Variable ( $I_{\cap}^{\text{RAV}}$ ) defines shared information as the maximal coinformation achievable by the sources, the target, and any function of the sources. It captures redundancy as information that can be extracted collectively from sources, regardless of whether it is present individually. This measure is not published, but it is part of the `dit` package [79]. Although it can be easily defined for any number of source variables, we underline that  $I_{\cap}^{\text{RAV}}$  satisfies many of the PID properties only in the case  $n = 2$ , which is hence the case we focused on.

$$I_{\cap}^{\text{RAV}}(X_1, \dots, X_n; Y) = \max_f I(X_1; \dots; X_n; Y; f(X_1, \dots, X_n)). \quad (108)$$

## E Properties of PID measures

In this section, we provide the justifications of the PID properties for each measure, as well as the proofs needed for the construction of Table 5.

### E.1 List of properties

#### Williams and Beer's Minimum Information ( $I_{\cap}^{\text{min}}$ )

- (SR), (S<sub>0</sub>), (M<sub>0</sub>), (GP), (LP<sub>0</sub>) follow directly from the definition [10];
- (ID) and (IID) are violated, as shown by direct checks on the TBC (e.g., Table 1 of [18]);
- (TM) is violated: a counterexample is provided in the table in Sec. 5 of Ref. [19];
- (TC) fails as a consequence of (GP) together with the violation of (TM) (see Prop. 5);
- (S<sub>1</sub>) is violated: see Table 1 of Ref. [20];
- (M<sub>1</sub>) is satisfied: see Ref. [23] (although it is not reported in Table 1 of Ref. [20]);

**(LP<sub>1</sub>)** is satisfied: see Refs. [10,20];  
**(BP)** is violated, whereas **(TE)** and **(\*)** are satisfied: see Ref. [38];  
**(AD)** is violated, while **(CO)** holds: see Ref. [68];  
**(DI)** fails globally but holds almost everywhere, as  $I_{\cap}^{\min}$  is a minimum of differentiable functions;  
**(LB)** follows from **(M<sub>0</sub>)** and **(SR)** (see Prop. 1);  
**(EI)** is satisfied: see Ref. [20];  
**(SE)** follows from **(M<sub>0</sub>)**.

#### Harder *et al.*'s Projected Information ( $I_{\cap}^{\text{red}}$ )

**(SR)**, **(S<sub>0</sub>)**, **(M<sub>0</sub>)**, **(GP)**, **(LP<sub>0</sub>)**, **(ID)**, and **(IID)** come from its definition [18];  
**(TM)** is violated: see Refs. [19,66];  
**(TC)** fails as a consequence of **(GP)** together with the violation of **(TM)** (see Prop. 5);  
**(S<sub>1</sub>)** is violated: see Table 1 of Ref. [20];  
**(M<sub>1</sub>)** is satisfied: see Table 1 of Ref. [66] (although it is not reported in Table 1 of Ref. [20]);  
**(LP<sub>1</sub>)** is violated, as the  $I_{\cap}^{\text{red}}$  decomposition is only defined for  $n = 2$  sources;  
**(BP)**, **(TE)**, and **(\*)** are satisfied: see Ref. [38];  
**(AD)** and **(CO)** are violated: see Ref. [68];  
**(DI)** fails globally but holds almost everywhere, as  $I_{\cap}^{\text{red}}$  is a minimum of differentiable functions;  
**(LB)** follows from **(M<sub>0</sub>)** and **(SR)** (see Prop. 1);  
**(EI)** is satisfied: see Ref. [20];  
**(SE)** follows from **(M<sub>0</sub>)**.

Note that **(CO)** would hold in the case of full-support distributions.

#### Griffith *et al.*'s Common Information Redundancy ( $I_{\cap}^{\wedge}$ )

**(SR)**, **(S<sub>0</sub>)**, **(M<sub>0</sub>)**, and **(GP)** follow from the definition, whereas **(LP<sub>0</sub>)** and **(ID)** are violated [20];  
**(IID)** is satisfied: see Table 1 of Ref. [38];  
**(TM)** follows from its definition [20];  
**(TC)** is violated, since  $I_{\cap}^{\wedge}(X_1, \dots, X_n; Y_1 Y_2) = \max_Q I(Q; Y_1 Y_2) = \max_Q (I(Q; Y_1) + I(Q; Y_2 | Y_1)) \leq \max_Q I(Q; Y_1) + \max_Q I(Q; Y_2 | Y_1) = I_{\cap}^{\wedge}(X_1, \dots, X_n; Y_1) + I_{\cap}^{\wedge}(X_1, \dots, X_n; Y_2 | Y_1)$ ;  
**(S<sub>1</sub>)** is violated by definition [20];  
**(M<sub>1</sub>)** does not hold: see Table 1 of Refs. [20,66];  
**(LP<sub>1</sub>)** fails as a consequence of the violation of **(LP<sub>0</sub>)**;  
**(BP)** and **(TE)** are violated, whereas **(\*)** is satisfied: see Ref. [38];  
**(AD)** is satisfied, while **(CO)** is violated: see Ref. [68];  
**(DI)** is violated as **(CO)** does not hold;  
**(LB)** follows from **(M<sub>0</sub>)** and **(SR)** (see Prop. 1);  
**(EI)** comes from the definition [20];  
**(SE)** follows from **(M<sub>0</sub>)**.

#### Griffith *et al.*'s Relaxed Common Information Redundancy ( $I_{\cap}^{\alpha}$ )

**(SR)**, **(S<sub>0</sub>)**, **(M<sub>0</sub>)**, and **(GP)** follow from the definition, whereas **(LP<sub>0</sub>)** is violated [23];  
**(ID)** is violated: see Table 1 of Ref. [66];  
**(IID)** is satisfied: see Table 1 of Ref. [38];  
**(TM)** is violated: see Refs. [38,66] (despite being claimed without proof in Ref. [23]);  
**(TC)** fails as a consequence of **(GP)** together with the violation of **(TM)** (see Prop. 5);  
**(S<sub>1</sub>)** is violated due to the clear asymmetry between sources and target in the maximisation;  
**(M<sub>1</sub>)** is satisfied: see Table 1 of Ref. [66];  
**(LP<sub>1</sub>)** is violated as a consequence of the violation of **(LP<sub>0</sub>)**;  
**(BP)** and **(\*)** are violated, whereas **(TE)** is satisfied: see Ref. [38];  
**(AD)** is satisfied, while **(CO)** is violated: see Ref. [68];

(**DI**) is violated as (**CO**) does not hold;  
(**LB**) follows from (**M<sub>0</sub>**) and (**SR**) (see Prop. 1);  
(**EI**) comes from the invariance of mutual information under equivalence classes [109];  
(**SE**) follows from (**M<sub>0</sub>**).

### BROJA/Griffith and Koch's Constrained Optimisation ( $I_{\cap}^{\text{BROJA}}$ )

(**SR**), (**S<sub>0</sub>**), (**M<sub>0</sub>**), (**GP**), (**LP<sub>0</sub>**), (**ID**), and (**IID**) follow directly from the definition [21];  
(**TM**) is violated: see Ref. [42];  
(**TC**) fails as a consequence of (**GP**) together with the violation of (**TM**) (see Prop. 5) [23, 66];  
(**S<sub>1</sub>**) is violated due to the clear asymmetry between sources and target induced by the constraints defining  $\Delta_P$ ;  
(**M<sub>1</sub>**) is satisfied: see Table 1 of Ref. [66];  
(**LP<sub>1</sub>**) is violated, as the  $I_{\cap}^{\text{BROJA}}$  decomposition is only defined for  $n = 2$  sources;  
(**BP**), (**TE**), and (**\***) are satisfied: see Ref. [38];  
(**AD**) and (**CO**) are satisfied: see Ref. [68];  
(**DI**) does not hold: see Refs. [68, 123].  
(**LB**) follows from (**M<sub>0</sub>**) and (**SR**) (see Prop. 1);  
(**EI**) comes from the invariance of mutual information under equivalence classes [109];  
(**SE**) follows from (**M<sub>0</sub>**).

### Barrett's Minimum Mutual Information ( $I_{\cap}^{\text{MMI}}$ )

(**SR**), (**S<sub>0</sub>**), (**M<sub>0</sub>**), (**GP**), and (**LP<sub>0</sub>**) follow directly from the definition [24];  
(**ID**) and (**IID**) are violated, as shown by direct checks on the TBC;  
(**TM**) follows directly from the definition, since  $I(X_1, \dots, X_n; Y_1, Y_2) \geq I(X_1, \dots, X_n; Y_i)$ ,  $i = 1, 2$ ;  
(**TC**) is violated, since  $I_{\cap}^{\text{MMI}}(X_1, X_2; Y_1 Y_2) = \min\left(I(X_1; Y_1) + I(X_1; Y_2 | Y_1), I(X_2; Y_1) + I(X_2; Y_2 | Y_1)\right) \leq \min\left(I(X_1; Y_1) + I(X_2; Y_1)\right) + \min\left(I(X_1; Y_2 | Y_1) + I(X_2; Y_2 | Y_1)\right) = I_{\cap}^{\text{MMI}}(X_1, X_2; Y_1) + I_{\cap}^{\text{MMI}}(X_1, X_2; Y_2 | Y_1)$ ;  
(**S<sub>1</sub>**) is violated, as shown by direct evaluation on the TBC with  $X_1 \perp\!\!\!\perp X_2$ , for which  $I_{\cap}^{\text{MMI}}(X_1, X_2; Y) = \min\{H(X_1), H(X_2)\} \neq I_{\cap}^{\text{MMI}}(X_1, Y; X_2) = 0$ ;  
(**M<sub>1</sub>**) follows from (**M<sub>0</sub>**) and (**TE**) (see Theo. 6);  
(**LP<sub>1</sub>**) is satisfied: see Prop. 8;  
(**BP**) is violated, whereas (**TE**) and (**\***) are satisfied: see Ref. [38] (note that (**BP**) holds in the Gaussian case of two sources and a univariate target [39]);  
(**AD**) is violated, while (**CO**) is satisfied: see Ref. [68];  
(**DI**) fails globally but holds almost everywhere, as  $I_{\cap}^{\text{MMI}}$  is a minimum of differentiable functions;  
(**LB**) follows from (**M<sub>0</sub>**) and (**SR**) (see Prop. 1);  
(**EI**) comes from the invariance of mutual information under equivalence classes [109];  
(**SE**) follows from (**M<sub>0</sub>**).

### Goodwell and Kumar's Rescaled Redundancy ( $I_{\cap}^{\text{RR}}$ )

(**SR**), (**S<sub>0</sub>**), and (**M<sub>0</sub>**) follow directly from the definition [25];  
(**GP**) comes from Prop. 9;  
(**LP<sub>0</sub>**) follows from Prop. 10;  
(**ID**) is violated, as shown by direct evaluation on the TBC;  
(**IID**) follows directly from the definition [25];  
(**TM**) is violated: for instance, taking  $X_1 \perp\!\!\!\perp X_2$  and assuming  $I(X_1; X_2; Y_1 Y_2) \geq 0$  yields  $I_{\cap}^{\text{RR}}(X_1, X_2; Y_1 Y_2) = I(X_1; X_2; Y_1 Y_2) \not\geq I(X_1; X_2; Y_1) = I_{\cap}^{\text{RR}}(X_1; X_2; Y_1)$ ;  
(**TC**) is violated by the same reasoning as for  $I_{\cap}^{\text{MMI}}$ , together with (**GP**) and the violation of (**TM**) (see Prop. 5);  
(**S<sub>1</sub>**) is violated, as shown by direct evaluation on a system with  $X_1 = Y$ , for which  $R_{\min} = 0$  and  $I_{\cap}^{\text{MMI}}(X_1, X_2; Y) = I_{\cap}^{\text{MMI}}(X_1, Y; X_2)$ , but  $I_s(X_1, X_2) \neq I_s(X_1, Y)$ ;  
(**M<sub>1</sub>**) follows from (**M<sub>0</sub>**) and (**TE**) (see Theo. 6);

**(LP<sub>1</sub>)** is violated, as the  $I_{\cap}^{\text{RR}}$  decomposition is only defined for  $n = 2$  sources;  
**(BP)** is violated: see Prop. 11;  
**(TE)** is satisfied: see Prop. 12;  
**(\*)** is violated, since both  $R_{\min}$  and  $I_s(X_1, X_2)$  depend explicitly on  $I(X_1; X_2)$ , and thus on  $p(x_1, x_2)$ ;  
**(AD)** is violated, since  $I_{\cap}^{\text{RR}} \propto I_{\cap}^{\text{MMI}}$  for  $I(X_1; X_2; Y) \geq 0$  and  $I_{\cap}^{\text{MMI}}$  is not additive;  
**(CO)** comes from the continuity of  $I_{\cap}^{\text{MMI}}$  and the coinformation;  
**(DI)** fails globally but holds almost everywhere, as  $I_{\cap}^{\text{RR}}$  contains minima and maxima of differentiable functions;  
**(LB)** follows from **(M<sub>0</sub>)** and **(SR)** (see Prop. 1);  
**(EI)** comes from the invariance of mutual information under equivalence classes [109];  
**(SE)** follows from **(M<sub>0</sub>)**.

### Ince's Common Change in Surprisal ( $I_{\cap}^{\text{CCS}}$ )

**(SR)** and **(S<sub>0</sub>)** are satisfied, whereas **(M<sub>0</sub>)** and **(GP)** are violated by definition [26];  
**(LP<sub>0</sub>)** fails as a consequence of the violation of **(GP)**;  
**(ID)** is violated and **(IID)** is satisfied, as given by the definition [26];  
**(TM)**, **(TC)**, and **(S<sub>1</sub>)** are violated as observed from empirical tests;  
**(M<sub>1</sub>)** does not hold due to the violation of **(M<sub>0</sub>)**;  
**(LP<sub>1</sub>)** is violated because **(GP)** does not hold;  
**(BP)**, **(TE)**, and **(\*)** are violated: see Ref. [38];  
**(AD)** and **(CO)** are violated: see Ref. [68];  
**(DI)** does not hold as **(CO)** fails;  
**(LB)** fails as a consequence of the violation of **(GP)** (see Prop. 2);  
**(EI)** is satisfied due to invariance of local mutual information under equivalence classes [109];  
**(SE)** is satisfied by definition [26].

### James *et al.*'s Unique Information via Dependency Constraints ( $I_{\cap}^{\text{DEP}}$ )

**(SR)**, **(S<sub>0</sub>)**, **(M<sub>0</sub>)**, **(GP)**, **(LP<sub>0</sub>)**, and **(ID)** are satisfied by definition [27];  
**(IID)** follows from **(ID)**;  
**(TM)** is violated by definition [27];  
**(TC)** fails due to the validity of **(GP)** and the violation of **(TM)** (see Prop. 5);  
**(S<sub>1</sub>)** is violated as observed from empirical tests;  
**(M<sub>1</sub>)** is satisfied, as it follows from **(M<sub>0</sub>)** and **(TE)** (see Theo. 6);  
**(LP<sub>1</sub>)** is violated, as the  $I_{\cap}^{\text{DEP}}$  decomposition is only defined for  $n = 2$  sources;  
**(BP)** is violated: see Ref. [38];  
**(TE)** is satisfied: see Prop. 14 (despite being reported otherwise in Ref. [38]);  
**(\*)** is violated: see Ref. [38];  
**(AD)** is violated while **(CO)** is satisfied: see Ref. [42];  
**(DI)** fails globally but holds almost everywhere, as  $I_{\cap}^{\text{DEP}}$  is a minimum of differentiable functions;  
**(LB)** follows from **(M<sub>0</sub>)** and **(SR)** (see Prop. 1);  
**(EI)** is satisfied due to invariance of mutual information under equivalence classes [109];  
**(SE)** follows from **(M<sub>0</sub>)**.

### James *et al.*'s Maximum Entropy Star ( $I_{\cap}^{\text{MES}}$ )

**(SR)** and **(S<sub>0</sub>)** follows directly from the definition of  $I_{\cap}^{\text{MES}}(X_1, X_2; Y)$  as a mutual information (see Prop. 15);  
**(M<sub>0</sub>)** fails due to the violation of **(SE)** (see Prop. 17);  
**(GP)** is satisfied by definition of  $I_{\cap}^{\text{MES}}$  being a mutual information (see Prop. 15);  
**(LP<sub>0</sub>)** is violated, as synergy can be negative as seen from empirical tests;  
**(IID)** is follows from **(ID)**;  
**(ID)** is satisfied, since  $I_{\cap}^{\text{MES}}(X_1, X_2; Y) = I_Q(X_1; X_2)$ , and for the TBC the maximum entropy distribution  $Q$  equals  $P$  (see Prop. 15 and Lemma 8);

**(TM)** is violated as observed from empirical tests;  
**(TC)** fails due to the validity of **(GP)** and the violation of **(TM)** (see Prop. 5);  
**(S<sub>1</sub>)** does not hold due to the clear asymmetry between sources and target given by the constraints;  
**(M<sub>1</sub>)** is violated as **(M<sub>0</sub>)** fails (see Theo. 6);  
**(LP<sub>1</sub>)** fails as  $I_{\cap}^{\text{MES}}$  is only defined for  $n = 2$  sources;  
**(BP)** is violated: see Prop. 19;  
**(TE)** is satisfied, since  $I_{\cap}^{\text{MES}}(X_1, Y; Y) = I_Q(X_1; Y; Y) = I_Q(X_1; Y) = I_{\cap}^{\text{MES}}(X_1; Y)$ ;  
**(\*)** holds by definition [29];  
**(AD)** is satisfied: see Prop. 20;  
**(CO)** holds: see Prop. 18;  
**(DI)** is violated similarly to  $I_{\cap}^{\text{BROJA}}$ : see Ref. [123];  
**(LB)** is violated: see Prop. 17;  
**(EI)** is satisfied due to invariance under equivalence classes [109];  
**(SE)** is violated: see Prop. 17.

### Finn and Lizier's Specificity and Ambiguity ( $I_{\cap}^{\text{PM}}$ )

**(SR)** and **(S<sub>0</sub>)** hold, whereas **(M<sub>0</sub>)** and **(GP)** are violated by definition [30];  
**(LP<sub>0</sub>)** does not hold as a consequence of the violation of **(GP)**;  
**(ID)** and **(IID)** do not hold, as confirmed by direct checks on the TBC [30];  
**(TM)** is violated and **(TC)** is satisfied by definition [30];  
**(S<sub>1</sub>)** is violated due to the clear asymmetry between sources and target;  
**(M<sub>1</sub>)** does not hold due to the violation of **(M<sub>0</sub>)**;  
**(LP<sub>1</sub>)** does not hold because **(LP<sub>0</sub>)** fails;  
**(BP)**, **(TE)**, and **(\*)** do not hold: see Ref. [38];  
**(AD)** is violated, since the min function is superadditive;  
**(CO)** holds, as  $I_{\cap}^{\text{PM}}$  is a composition of continuous functions;  
**(DI)** fails globally but holds almost everywhere, as  $I_{\cap}^{\text{PM}}$  contains a minimum of differentiable functions;  
**(LB)** does not hold, since  $i^{\pm}(q; y) \leq i^{\pm}(x_i; y)$  does not imply  $I(Q; Y) \leq I^{\pm}(X_i; Y)$  (see also Prop. 2);  
**(EI)** holds due to invariance of local mutual information under equivalence classes [109];  
**(SE)** follows from the informative and misinformative **(M<sub>0</sub>)** (since  $i_{\cap}^{\pm}(X_1, \dots, X_n; Y) = i^{\pm}(X_1, \dots, X_{n-1}; Y)$  when  $X_i = f(X_n)$  for some  $i = 1, \dots, n-1$ , and hence  $i_{\cap}(X_1, \dots, X_n; Y) = i_{\cap}^{+}(X_1, \dots, X_n; Y) - i_{\cap}^{-}(X_1, \dots, X_n; Y) = i_{\cap}(X_1, \dots, X_{n-1}, Y)$ ).  
 We remark that **(M<sub>0</sub>)**, **(GP)**, and **(LP)** hold separately at the pointwise level for the informative and misinformative lattices.

### Niu and Quinn's Information-Geometric Redundancy ( $I_{\cap}^{\text{IG}}$ )

**(SR)** holds by construction;  
**(S<sub>0</sub>)** follows from Prop. 21;  
**(M<sub>0</sub>)** follows from Prop. 22;  
**(GP)** comes from its original definition [31];  
**(LP<sub>0</sub>)** is satisfied since **(GP)** and both unique information and synergy are KL divergences [31];  
**(ID)** and **(IID)** are violated, as confirmed by direct application of the independent TBC;  
**(TM)** holds from empirical tests;  
**(TC)** is violated as observed from empirical tests;  
**(S<sub>1</sub>)** is violated since there is a clear asymmetry between sources and targets in the geometric projections;  
**(M<sub>1</sub>)** is satisfied since **(M<sub>0</sub>)** and **(TE)** hold (see Theo. 6);  
**(LP<sub>1</sub>)** is violated since the  $I_{\cap}^{\text{IG}}$  decomposition is only valid for  $n = 2$  sources;  
**(BP)** is violated as observed from empirical tests;  
**(TE)** follows from Prop. 23;  
**(\*)** fails as proved in Prop. 24;

(**AD**) is violated and (**CO**) is satisfied: see Ref. [68];  
(**DI**) holds as  $I_{\cap}^{\text{IG}}$  is a composition of smooth functions [68];  
(**LB**) since (**M<sub>0</sub>**) and (**SR**) (see Prop. 1);  
(**EI**) comes from the invariance of probability distribution under equivalence class [109];  
(**SE**) follows from (**M<sub>0</sub>**).

### Sigtermans' Redundancy via Causal Tensors ( $I_{\cap}^{\text{CT}}$ )

(**SR**), (**S<sub>0</sub>**), (**M<sub>0</sub>**), and (**GP**) are satisfied by definition [33];  
(**LP<sub>0</sub>**) is violated as synergy can be negative [33];  
(**ID**) follows from its definition [33];  
(**IID**) is satisfied as a consequence of (**ID**);  
(**TM**) is satisfied by definition [33];  
(**TC**) is violated: similarly to  $I_{\cap}^{\text{MMI}}$ , the chain rule holds for mutual information, but taking the minimum breaks it;  
(**S<sub>1</sub>**) does not hold: see Theorem 3 of Ref. [33];  
(**M<sub>1</sub>**) is satisfied, following from (**M<sub>0</sub>**) and (**TE**) (see Theo. 6);  
(**LP<sub>1</sub>**) is violated by definition [33];  
(**BP**) does not hold: see Prop. 26;  
(**TE**) is satisfied, since paths including  $Y$  as sources cannot be considered as they are directed, and hence  $I_{\cap}^{\text{CT}}(X_1, \dots, X_n, Y; Y) = I_{\cap}^{\text{CT}}(X_1, \dots, X_n; Y)$ ;  
(\*) is violated, since altering the distribution between  $X_1$  and  $X_2$  evidently changes  $I_{\cap}^{\text{CT}}(X_1, X_2; Y)$  (see Prop. 25);  
(**AD**) does not hold as observed from empirical tests;  
(**CO**) is satisfied, as  $I_{\cap}^{\text{CT}}$  is the minimum of mutual information terms;  
(**DI**) fails globally but holds almost everywhere, as  $I_{\cap}^{\text{CT}}$  is a minimum of differentiable functions;  
(**LB**) follows from (**M<sub>0</sub>**) and (**SR**) (see Prop. 1);  
(**EI**) is satisfied, since causal tensors can be rewritten in terms of conditional probabilities, which are invariant under equivalence classes [109];  
(**SE**) follows from (**M<sub>0</sub>**).

### Makkeh *et al.*'s Shared Exclusion ( $I_{\cap}^{\text{SX}}$ )

(**SR**) and (**S<sub>0</sub>**) are satisfied by definition [35];  
(**M<sub>0</sub>**), (**GP**), and (**LP<sub>0</sub>**) do not hold due to the pointwise nature of the measure [35];  
(**ID**) and (**IID**) are violated, as it can be verified on the independent TBC;  
(**TM**) is not satisfied as observed from empirical tests;  
(**TC**) follows from its original definition [35];  
(**S<sub>1</sub>**) does not hold due to the clear asymmetry between sources and target;  
(**M<sub>1</sub>**) is violated, because (**M<sub>0</sub>**) does not hold;  
(**LP<sub>1</sub>**) does not hold, as (**LP<sub>0</sub>**) fails;  
(**BP**) is violated as (**TE**) fails (see Theo. 7);  
(**TE**) does not hold as  $i_{\cap}^{\text{SX}-}(X_1, \dots, X_n, Y; Y) = 0$  but  $i_{\cap}^{\text{SX}+}(X_1, \dots, X_n, Y; Y) \neq 0$ ;  
(\*) is not satisfied due to the pointwise unique behaviour of the measure [35];  
(**AD**) is violated as observed from empirical tests;  
(**CO**) and (**DI**) are satisfied as  $I_{\cap}^{\text{SX}}$  is a composition of smooth functions [35];  
(**LB**) does not hold, since (**GP**) fails (see Prop. 2);  
(**EI**) is satisfied due to invariance of probability distributions under equivalence classes [109];  
(**SE**) follows from the informative and misinformative (**M<sub>0</sub>**) (since  $i_{\cap}^{\pm}(X_1, \dots, X_n; Y) = i^{\pm}(X_1, \dots, X_{n-1}; Y)$  when  $X_i = f(X_n)$  for some  $i = 1, \dots, n-1$ , and hence  $i_{\cap}(X_1, \dots, X_n; Y) = i_{\cap}^{+}(X_1, \dots, X_n; Y) - i_{\cap}^{-}(X_1, \dots, X_n; Y) = i_{\cap}(X_1, \dots, X_{n-1}, Y)$ ).

We remark that (**M<sub>0</sub>**), (**GP**), and (**LP**) are satisfied separately at the pointwise level for the informative and misinformative lattices.

### Kolchinsky’s Blackwell Redundancy ( $I_{\cap}^{\prec}$ )

- (**SR**), (**S<sub>0</sub>**), and (**M<sub>0</sub>**) hold by definition (see Theo. 1 in Ref. [38]);
- (**GP**) is satisfied, as it is a supremum of mutual information [38];
- (**LP<sub>0</sub>**) is violated as observed from empirical tests;
- (**ID**) does not hold and (**IID**) is satisfied by definition [38];
- (**TM**) is not satisfied: see Prop. 27;
- (**TC**) fails as a consequence of (**GP**) and the violation of (**TM**) (see Prop. 5);
- (**S<sub>1</sub>**) is violated due to Blackwell order asymmetry between sources and targets;
- (**M<sub>1</sub>**) holds following from (**TE**) and (**M<sub>0</sub>**) (see Theo. 6);
- (**LP<sub>1</sub>**) does not hold, since (**LP<sub>0</sub>**) is violated;
- (**BP**), (**TE**), and (**\***) are satisfied: see Ref. [38];
- (**AD**) is satisfied while (**CO**) does not hold (though it is locally continuous under specific assumptions [38]): see Ref. [68];
- (**DI**) is violated as (**CO**) does not hold;
- (**LB**) follows from (**M<sub>0</sub>**) and (**SR**) (see Prop. 1);
- (**EI**) holds due to invariance under equivalence classes [109];
- (**SE**) is satisfied, as  $I_{\cap}^{\prec}$  obeys the more general “order equality” and follows from (**M<sub>0</sub>**) [38].

### Venkatesh and Schamberg’s Deficiency Redundancy ( $I_{\cap}^{\delta}$ )

- (**SR**) is satisfied by construction;
- (**S<sub>0</sub>**) holds by definition [39];
- (**M<sub>0</sub>**) follows from Prop. 28;
- (**GP**) holds since  $0 \leq \delta(Y; X_1 \setminus X_2) \leq I(X_1; Y)$ ;
- (**LP<sub>0</sub>**) is satisfied according to empirical tests [39];
- (**ID**) follows from Prop. 30;
- (**IID**) holds as a consequence of (**ID**);
- (**TM**) is violated as observed from empirical tests;
- (**TC**) fails due to (**GP**) combined with the violation of (**TM**) (see Prop. 5);
- (**S<sub>1</sub>**) is violated: for univariate target  $Y$ ,  $I_{\cap}^{\delta}$  reduces to  $I_{\cap}^{\text{MMI}}$  which does not satisfy (**S<sub>1</sub>**);
- (**M<sub>1</sub>**) follows from (**M<sub>0</sub>**) and (**TE**) (see Theo. 6);
- (**LP<sub>1</sub>**) does not hold, as the decomposition is defined only for  $n = 2$  sources;
- (**BP**) is satisfied by definition [39];
- (**TE**) follows from Prop. 29;
- (**\***) is satisfied, since  $\delta$  depends only on the marginal distributions  $P(X_1; Y)$  and  $P(X_2; Y)$  (see also Appendix C in Ref. [39]);
- (**AD**) is violated, since in the univariate case  $I_{\cap}^{\delta}$  reduces to  $I_{\cap}^{\text{MMI}}$ , which does not satisfy (**AD**);
- (**CO**) is satisfied, as  $I_{\cap}^{\delta}$  is a composition of KL divergences and expectations;
- (**DI**) fails globally but holds almost everywhere, as  $I_{\cap}^{\delta}$  is a minimum of differentiable functions;
- (**LB**) follows from (**M<sub>0</sub>**) and (**SR**) (see Prop. 1);
- (**EI**) is satisfied due to invariance under equivalence classes [109];
- (**SE**) follows from (**M<sub>0</sub>**).

### Mages and Rohner’s Decision Reachability Redundancy ( $I_{\cap}^{\text{RDR}}$ )

- (**SR**), (**S<sub>0</sub>**), (**M<sub>0</sub>**), (**GP**), and (**LP<sub>0</sub>**) are satisfied, whereas (**ID**) does not hold by definition [40];
- (**IID**) is violated based on direct checks on the independent TBC;
- (**TM**) is not satisfied, since  $I_{\cap}^{\text{min}}$  fails it and  $I_{\cap}^{\text{RDR}}$  equals  $I_{\cap}^{\text{min}}$  under pointwise Blackwell ordering;
- (**TC**) does not hold as a consequence of (**GP**) combined with the violation of (**TM**) (see Prop. 5);
- (**S<sub>1</sub>**) is violated, as neither  $I_{\cap}^{\text{min}}$  nor  $I_{\cap}^{\text{BROJA}}$  satisfy it, and  $I_{\cap}^{\text{RDR}}$  is equal to them in specific scenarios;
- (**M<sub>1</sub>**) follows from (**M<sub>0</sub>**) and (**TE**) (see Theo. 6);

**(LP<sub>1</sub>)** and **(BP)** are satisfied by definition [40];  
**(TE)** holds for  $n = 2$  as a consequence of **(BP)** (see Theo. 7), or equivalently as  $I_{\cap}^{\text{RDR}}(X_1, Y; Y) = I(X_1; Y)$ , since  $I_{\partial}(X_1; Y) = 0$ ;  
**(\*)** is satisfied, as  $I_{\cap}^{\text{RDR}}$  depends only on  $p(X_1, Y)$  and  $p(X_2, Y)$  (see Eq. (102));  
**(AD)** does not hold, since pointwise information does not factorise for independent subsystems (see denominator of Eq. (102));  
**(CO)** is satisfied, as  $I_{\cap}^{\text{RDR}}$  is a composition of continuous functions;  
**(DI)** fails globally but holds almost everywhere, as  $I_{\cap}^{\text{RDR}}$  is a combination of differentiable functions joint by Blackwell union;  
**(LB)** follows from **(M<sub>0</sub>)** and **(SR)** (see Prop. 1);  
**(EI)** holds due to the invariance under equivalence classes [109];  
**(SE)** follows from **(M<sub>0</sub>)**.  
 We remark that **(M<sub>1</sub>)** and **(TE)** do not hold for  $n > 2$  (counterexamples can be found empirically).

### Lyu *et al.*'s do-calculus Redundancy ( $I_{\cap}^{\text{do}}$ )

**(SR)** holds by construction (see Remark 2);  
**(S<sub>0</sub>)** is satisfied by definition [41];  
**(M<sub>0</sub>)** does not hold, as **(SE)** fails: see Prop. 31;  
**(GP)** is satisfied by definition [41];  
**(LP<sub>0</sub>)** is violated, as it only holds when  $H(Y|X_1, X_2) = 0$  (Remark 1 in Ref. [41]);  
**(ID)** holds: see Prop. 33;  
**(IID)** is satisfied, since **(ID)** holds;  
**(TM)** is violated as observed from empirical tests;  
**(TC)** fails as a consequence of **(GP)** and violation of **(TM)** (see Prop. 5);  
**(S<sub>1</sub>)** is violated due to the clear asymmetry between sources and target;  
**(M<sub>1</sub>)** does not hold, since **(M<sub>0</sub>)** fails;  
**(LP<sub>1</sub>)** does not hold, as  $I_{\cap}^{\text{do}}$  decomposition is defined only for  $n = 2$  sources;  
**(BP)** is violated, since  $X_1 = X_2$  gives non-zero  $I_{\partial}(X_1; Y)$ , or equivalently as **(SE)** fails (see Theo. 7);  
**(TE)** is satisfied, since  $I_{\cap}^{\text{do}}(X_1, Y; Y) = I_{\cap}^{\text{do}}(X'_1; Y) = I_{\cap}^{\text{do}}(X_1; Y)$  (using  $H(X'_1|Y) = H(X_1|Y)$  from Lemma 1 in Ref. [41]);  
**(\*)** is satisfied, as  $I_{\cap}^{\text{do}}(X_1, X_2; Y) = I(X'; Y)$  depends only on pairwise marginals  $p(X_1, Y)$  and  $p(X_2, Y)$ ;  
**(AD)**, **(CO)**, and **(DI)** are satisfied by definition [41];  
**(LB)** does not hold: see Prop. 32;  
**(EI)** holds due to the invariance under equivalence classes [109];  
**(SE)** is violated, since  $I_{\cap}^{\text{do}}(X_1, f(X_1); Y) \neq I_{\cap}^{\text{do}}(X_1; Y)$  (see Prop. 31).

### Redundancy via Auxiliary Variable ( $I_{\cap}^{\text{RAV}}$ )

**(SR)** is satisfied: see Prop. 34;  
**(S<sub>0</sub>)** follows directly from the definition of coinformation;  
**(M<sub>0</sub>)** follows from Prop. 35;  
**(GP)** holds, since  $I_{\cap}^{\text{RAV}}(X_1, \dots, X_n; Y) = \max_f I(X_1; \dots; X_n; Y; f(X_1, \dots, X_n)) \geq I(X_1; \dots; X_n; Y; C) = 0$  by choosing  $f(X_1, \dots, X_n) = C$  to be a constant;  
**(LP<sub>0</sub>)** is satisfied: see Prop. 38;  
**(ID)** holds: see Prop. 36;  
**(IID)** is satisfied as a consequence of **(ID)**;  
**(TM)** is violated as observed from empirical checks;  
**(TC)** fails as a consequence of **(GP)** and violation of **(TM)** (see Prop. 5);  
**(S<sub>1</sub>)** is violated due to the clear asymmetry between sources and targets in the argument of  $f$ ;  
**(M<sub>1</sub>)** follows from **(M<sub>0</sub>)** and **(TE)** (see Theo. 6);  
**(LP<sub>1</sub>)** does not hold as observed from empirical tests;

(**BP**) is violated as observed from empirical checks;

(**TE**) holds (see Prop. 37);

(\*) is violated, since  $I_{\cap}^{\text{RAV}}(X_1, \dots, X_n; Y) = \max_f I(X_1; \dots; X_n; Y; f(X_1, \dots, X_n))$  also depends on  $p(X_1, \dots, X_n)$ ;

(**AD**) does not hold, due to the presence of the maximum and to the function  $f$ , which might correlate independent sources;

(**CO**) is violated, as the  $\arg \max_f$  can change abruptly for small perturbations;

(**DI**) does not hold as (**CO**) fails;

(**LB**) follows from Prop. 1;

(**EI**) is satisfied due to invariance of entropies under equivalence classes [109];

(**SE**) follows from (**M<sub>0</sub>**).

We remark that (**M<sub>0</sub>**), (**M<sub>1</sub>**), (**TE**), and (**LB**) do not hold for  $n > 2$  (empirical counterexamples exist).

## E.2 Proofs

**Proposition 7.** *Consider a system  $\mathcal{S}$  with sources  $X_1, \dots, X_n$  and target  $Y$ . If for such a system the union information satisfies  $I_{\cup}(X_1, \dots, X_n; Y) > I(X_1, \dots, X_n; Y)$ , then any PID performed on  $\mathcal{S}$  violates (**LP**).*

*Proof.* By definition of union information and PID lattice (Eq. (18)), we have that:

$$\begin{aligned} I_{\cup}(X_1, \dots, X_n; Y) &= \sum_{J \subseteq \{1, \dots, n\}} (-1)^{|J|-1} I_{\cap}(X_{J_1}, \dots, X_{J_{|J|}}; Y) \\ &= \sum_{J \subseteq \{1, \dots, n\}} (-1)^{|J|-1} \sum_{\alpha \preceq \{X_{J_1}, \dots, X_{J_{|J|}}\}} I_{\partial}(\alpha; Y) \\ &= \sum_{\substack{\alpha \preceq \{X_{J_i}\} \\ i=1, \dots, n}} I_{\partial}(\alpha; Y), \end{aligned} \tag{109}$$

where the last line follows from the structure of the PID lattice, as the alternating signs reduce all multiplicities of the PID atoms to 1. This can be seen explicitly by writing out the different combinations of the atoms, but can be more easily understood via the set-theoretic relationship between intersection and union information [10, 124]. Hence we obtain

$$\begin{aligned} I(X_1, \dots, X_n; Y) &= \sum_{\alpha \preceq \{X_1, \dots, X_n\}} I_{\partial}(\alpha; Y) \\ &= I_{\cup}(X_1, \dots, X_n; Y) + \sum_{\substack{\alpha \succ \{X_{J_i}\} \\ i=1, \dots, n}} I_{\partial}(\alpha; Y). \end{aligned} \tag{110}$$

Thus, if  $I_{\cup}(X_1, \dots, X_n; Y) > I(X_1, \dots, X_n; Y)$ , the last term must be negative, hence violating (**LP**).  $\blacksquare$

**Lemma 6.** *The maximum entropy distribution of the system  $(X_1, X_2, Y)$  with constraints  $p(X_i, Y) = q(X_i, Y)$   $i = 1, 2$  is*

$$q(x_1, x_2, y) = p(x_1|y)p(x_2|y)p(y) = \frac{p(x_1, y)p(x_2, y)}{p(y)} \tag{111}$$

*Proof.* We prove this explicitly via Lagrange multipliers. Consider the Lagrangian

$$\mathcal{L} = \sum q(x_1, x_2, y) \log q(x_1, x_2, y) - \lambda \left( \sum q(x_1, x_2, y) - 1 \right) - \eta \left( \sum_{x_1} q(x_2, y) - p(x_2, y) \right) - \delta \left( \sum_{x_2} q(x_1, y) - p(x_1, y) \right), \tag{112}$$

differentiating we get

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial q_i} &= \log q_i + 1 - \lambda - \eta_{x_2 y} - \delta_{x_1 y} = 0 \\ q_i &\propto \exp(\eta_{x_2 y}) \exp(\delta_{x_1 y}) \equiv A(x_2, y) B(x_1, y). \end{aligned} \tag{113}$$

Thus, the maximising joint distribution must have the form

$$q(x_1, x_2, y) = C A(x_2, y) B(x_1, y), \quad (114)$$

for some normalisation constant  $C$ . We now impose the constraints to determine  $A(x_2, y)$  and  $B(x_1, y)$ . Summing over  $x_1$  gives

$$\begin{aligned} \sum_{x_1} q(x_1, x_2, y) &= C A(x_2, y) \sum_{x_1} B(x_1, y) \\ q(x_2, y) &= p(x_2, y), \end{aligned} \quad (115)$$

and hence

$$C A(x_2, y) = \frac{p(x_2, y)}{\sum_{x_1} B(x_1, y)}. \quad (116)$$

Similarly, summing over  $x_2$  gives

$$\begin{aligned} \sum_{x_2} q(x_1, x_2, y) &= C B(x_1, y) \sum_{x_2} A(x_2, y) \\ &= q(x_1, y) = p(x_1, y), \end{aligned} \quad (117)$$

and hence

$$C B(x_1, y) = \frac{p(x_1, y)}{\sum_{x_2} A(x_2, y)}. \quad (118)$$

Multiplying (116) and (118) we obtain

$$C^2 A(x_2, y) B(x_1, y) = \frac{p(x_1, y) p(x_2, y)}{\left( \sum_{x_1} B(x_1, y) \right) \left( \sum_{x_2} A(x_2, y) \right)}. \quad (119)$$

But  $q(x_1, x_2, y) = C A(x_2, y) B(x_1, y)$ , so

$$q(x_1, x_2, y) = \frac{p(x_1, y) p(x_2, y)}{Z(y)}, \quad (120)$$

where  $Z(y)$  is the normalisation factor

$$Z(y) = C \left( \sum_{x_1} B(x_1, y) \right) \left( \sum_{x_2} A(x_2, y) \right). \quad (121)$$

Finally, summing over  $x_1, x_2$  we must have  $\sum_{x_1 x_2} q(x_1, x_2, y) = p(y)$ , hence  $Z(y) = p(y)$  and therefore

$$q(x_1, x_2, y) = \frac{p(x_1, y) p(x_2, y)}{p(y)} = p(x_1|y) p(x_2|y) p(y). \quad (122)$$

■

### E.2.1 Proofs regarding $I_{\cap}^{\text{MMI}}$

**Proposition 8.**  $I_{\cap}^{\text{MMI}}$  satisfies  $(\text{LP}_1)$ .

*Proof.* We propose a proof similar to that of  $I_{\cap}^{\text{min}}$  (Theo.4 in Ref. [10]). Using Theo.3 and Lemma 2 of the same

reference, we have:

$$\begin{aligned}
I_{\partial}(\alpha; Y) &= I_{\cap}^{\text{MMI}}(\alpha; Y) - \sum_{k=1}^{|\alpha^-|} (-1)^{k-1} \sum_{\substack{\mathcal{B} \subseteq \alpha^- \\ |\mathcal{B}|=k}} I_{\cap}^{\text{MMI}}(\bigwedge \mathcal{B}; Y) \\
&= I_{\cap}^{\text{MMI}}(\alpha; Y) - \sum_{k=1}^{|\alpha^-|} (-1)^{k-1} \sum_{\substack{\mathcal{B} \subseteq \alpha^- \\ |\mathcal{B}|=k}} \min_{B \in \bigwedge \mathcal{B}} I(B; Y) \\
&= I_{\cap}^{\text{MMI}}(\alpha; Y) - \sum_{k=1}^{|\alpha^-|} (-1)^{k-1} \sum_{\substack{\mathcal{B} \subseteq \alpha^- \\ |\mathcal{B}|=k}} \min_{\beta \in \mathcal{B}} \min_{B \in \beta} I(B; Y) \\
&= I_{\cap}^{\text{MMI}}(\alpha; Y) - \max_{\beta \in \alpha^-} \min_{B \in \beta} I(B; Y) \\
&= \min_{A \in \alpha} I(A; Y) - \max_{\beta \in \alpha^-} \min_{B \in \beta} I(B; Y) \\
&\geq 0.
\end{aligned} \tag{123}$$

The last line follows from Theo. 2 of Ref. [10], i.e.  $I(B; Y) \leq I(A; Y)$  for  $B \subseteq A$ , and the ordering of the antichain lattice  $\beta \preceq \alpha \iff \forall A \in \alpha \exists B \in \beta \mid B \subseteq A$ . In fact, considering any  $\beta \in \alpha^-$ , we have that by definition of  $\alpha^-$   $\beta \preceq \alpha$ , thus defining  $A^* \equiv \arg \min_{A \in \alpha} I(A; Y)$ , there exists  $B^* \in \beta \mid B^* \subseteq A^*$ . Hence, we obtain:

$$I(A^*; Y) = \min_{A \in \alpha} I(A; Y) \geq I(B^*; Y) \geq \min_{B \in \beta} I(B; Y). \tag{124}$$

Since this holds  $\forall \beta \in \alpha^-$ , the last line of Eq.(123) follows. ■

## E.2.2 Proofs regarding $I_{\cap}^{\text{RR}}$

**Proposition 9.**  $I_{\cap}^{\text{RR}}$  satisfies **(GP)**.

*Proof.* Without loss of generality, we assume  $I(X_1; Y) \leq I(X_2; Y)$ . Then we have

$$0 \leq R_{\min} = \max(0, I(X_1; X_2; Y)) = \max(0, I(X_1; Y) - I(X_1; Y|X_2)) \leq \max(0, I(X_1; Y)). \tag{125}$$

Hence  $R_{\text{MMI}} - R_{\min} \geq 0$ , and since  $I_s(X_1, X_2) \geq 0$ , it follows

$$I_{\cap}^{\text{RR}}(X_1, X_2; Y) = R_{\min} + I_s(X_1, X_2)(R_{\text{MMI}} - R_{\min}) \geq 0. \tag{126}$$

**Proposition 10.**  $I_{\cap}^{\text{RR}}$  satisfies **(LP<sub>0</sub>)**.

*Proof.* Since we have **(GP)**, we just need to check unique information and synergy terms. For the former, we have

$$\begin{aligned}
I_{\partial}^{\text{RR}}(X_i; Y) &= I(X_i; Y) - I_{\cap}^{\text{RR}}(X_1, X_2; Y) \\
&\geq R_{\text{MMI}} - I_{\cap}^{\text{RR}}(X_1, X_2; Y) \\
&= R_{\text{MMI}} - R_{\min} - I_s(X_1, X_2)(R_{\text{MMI}} - R_{\min}) \\
&= (R_{\text{MMI}} - R_{\min})(1 - I_s(X_1, X_2)) \\
&\geq 0.
\end{aligned} \tag{127}$$

For synergy, we have:

$$\begin{aligned}
I_{\partial}^{\text{RR}}(X_1 X_2; Y) &= I(X_1, X_2; Y) - I(X_1; Y) - I(X_2; Y) + I_{\partial}^{\text{RR}}(X_1, X_2; Y) \\
&= -I(X_1; X_2; Y) + I_{\partial}^{\text{RR}}(X_1, X_2; Y) \\
&\geq -I(X_1; X_2; Y) + R_{\min} \\
&= \begin{cases} -I(X_1; X_2; Y) + I(X_1; X_2; Y) & \text{if } I(X_1; X_2; X_2) \geq 0 \\ -I(X_1; X_2; Y) & \text{if } I(X_1; X_2; X_2) \leq 0 \end{cases} \\
&\geq 0,
\end{aligned} \tag{128}$$

where we used that  $I_s(X_1, X_2)(R_{\text{MMI}} - R_{\min}) \geq 0$  (see Prop. 9). ■

**Proposition 11.**  $I_{\partial}^{\text{RR}}$  does not satisfy **(BP)**.

*Proof.* We prove this by violating the implication ( $\Leftarrow$ ) from the definition of Eq. (40). Assuming  $I(X_1; X_2; Y) \leq 0$ , we have that  $X_1 \preceq_Y X_2$  implies  $I_{\partial}^{\text{RR}}(X_1, X_2; Y) = I_s(X_1, X_2) I(X_1; Y) \leq I(X_1; Y)$  and hence  $I_{\partial}^{\text{RR}}(X_1; Y)$  can be non-zero.

Alternatively, just note that for  $I(X_1; X_2; Y) \leq 0$  we have  $I_{\partial}^{\text{RR}} \propto I_{\partial}^{\text{MMI}}$ , and  $I_{\partial}^{\text{MMI}}$  does not satisfy **(BP)**. ■

**Proposition 12.**  $I_{\partial}^{\text{RR}}$  satisfies **(TE)**.

*Proof.* Consider the PID with  $X_2 = Y$ , then  $R_{\min}(X_1, Y, Y) = R_{\text{MMI}}(X_1, Y, Y) = I(X_1; Y)$ . It follows that

$$I_{\partial}^{\text{RR}}(X_1, Y; Y) = R_{\min}(X_1, Y, Y) = I(X_1; Y). \tag{129}$$

**Proposition 13.**  $I_{\partial}^{\text{RR}}$  does not allow for mechanistic redundancy.

*Proof.* We just need to prove that if  $X_1 \perp\!\!\!\perp X_2$  then  $I_{\partial}^{\text{RR}}(X_1, X_2; Y) = 0 \forall Y$ . This directly follows from the definition, since  $I_s(X_1, X_2) \propto I(X_1; X_2) = 0$  and  $R_{\min} = 0$ . Thus  $I_{\partial}^{\text{RR}}(X_1, X_2; Y) = 0$ . ■

### E.2.3 Proofs regarding $I_{\partial}^{\text{DEP}}$

**Proposition 14.**  $I_{\partial}^{\text{DEP}}$  satisfies **(TE)**.

*Proof.* Consider the PID  $(X_1, X_2; Y)$  with  $X_2 = Y$ . For **(TE)** to hold, we need  $I_{\partial}^{\text{DEP}}(X_1, X_2; Y) = I_{\partial}^{\text{DEP}}(X_1; Y) = I(X_1; Y)$ , i.e.  $I_{\partial}^{\text{DEP}}(X_1; Y) = 0$ . Using the definition of unique information, we have:

$$\begin{aligned}
I_{\partial}^{\text{DEP}}(X_1; Y) &= \min\left(I_{X_1 Y : X_2 Y}(X_1; Y | X_2), I_{X_1 X_2 : X_1 Y : X_2 Y}(X_1; Y | X_2)\right) \\
&\geq I_{X_1 Y : X_2 Y}(X_1; Y | X_2) \\
&= 0,
\end{aligned} \tag{130}$$

where in the last step we used that the  $X_1 \perp\!\!\!\perp Y | X_2$  on the maximum entropy distribution with constraints  $p(x_i, y) = q(x_i, y)$ ,  $i = 1, 2$  given by Lemma 6. In fact, we have:

$$\begin{aligned}
q(x_1, y | x_2) &= \frac{q(x_1, x_2, y)}{q(x_2)} \\
&= \frac{p(x_1, y) p(x_2, y)}{p(x_2) p(y)} \\
&= \frac{p(x_1, y) \delta(x_2, y)}{p(y)} \\
&= p(x_1 | y) \delta(x_2, y),
\end{aligned} \tag{131}$$

and thus

$$\begin{aligned}
I_{X_1 Y : X_2 Y}(X_1; Y | X_2) &= \sum_{x_1, x_2, y} q(x_1, x_2, y) \log \left( \frac{q(x_1, y | x_2)}{q(x_1 | x_2) q(y | x_2)} \right) \\
&= \sum_{x_1, x_2, y} q(x_1, x_2, y) \log \left( \frac{p(x_1 | y)}{p(x_1 | y)} \delta(x_2, y) \right) \\
&= 0.
\end{aligned} \tag{132}$$

The result then follows from **(LP<sub>0</sub>)**. ■

#### E.2.4 Proofs regarding $I_{\cap}^{\text{MES}}$

**Proposition 15.**  $I_{\cap}^{\text{MES}}(X_1, X_2; Y) = I_Q(X_1; X_2; Y) = I_Q(X_1; X_2)$

*Proof.* From the definition of  $I_{\cup}^{\text{MES}}(X_1, X_2; Y)$  and the inclusion-exclusion principle we have

$$\begin{aligned}
I_{\cap}^{\text{MES}}(X_1, X_2; Y) &= I(X_1; X_2; Y) + I_{\delta}^{\text{MES}}(X_1 X_2; Y) \\
&= I(X_1; Y) + I(X_2; Y) - I(X_1, X_2; Y) + I(X_1, X_2; Y) - I_Q(X_1, X_2; Y) \\
&= I(X_1; Y) + I(X_2; Y) - I_Q(X_1, X_2; Y) \\
&= I_Q(X_1; Y) + I_Q(X_2; Y) - I_Q(X_1, X_2; Y) \\
&= I_Q(X_1; X_2; Y) \\
&= I_Q(X_1; X_2) - I_Q(X_1; X_2 | Y) \\
&= I_Q(X_1; X_2),
\end{aligned} \tag{133}$$

where in the last step we noticed that  $I_Q(X_1; X_2 | Y) = 0$  due to  $Q$  being the constrained maximum entropy distribution (see Lemma 6). ■

**Proposition 16.**  $I_{\delta}^{\text{MES}}(X_i; Y) \geq 0$

*Proof.* From Prop. 15 we have:

$$\begin{aligned}
I_{\delta}^{\text{MES}}(X_i; Y) &= I(X_i; Y) - I_{\cap}^{\text{MES}}(X_1, X_2; Y) \\
&= I(X_i; Y) - I_Q(X_1; X_2),
\end{aligned} \tag{134}$$

but we can also notice that

$$I_Q(X_1, X_2) \leq I_Q(X_1; X_2, Y) = I_Q(X_1; X_2 | Y) + I_Q(X_1; Y) = I_Q(X_1; Y) = I(X_1; Y) \tag{135}$$

where we used  $I_Q(X_1; X_2 | Y) = 0$  and the constraints on  $Q$ . Therefore, Eq. (135) ensures that the RHS in Eq.(135) is non-negative. Analogous derivation holds for  $X_2$ . ■

**Lemma 7.** Consider  $Q$  being the maximum entropy distribution between  $X_1, X_2, Y$  with the constraints given by  $\Delta_P$ . Then  $I_Q(X_1; Y | X_2) = 0$  if and only if  $X_1 \perp\!\!\!\perp Y$  on  $P$ .

*Proof.*  $I_Q(X_1; Y | X_2) = 0$  means that  $q(x_1, y | x_2) = q(x_1 | x_2) q(y | x_2)$ . On the other hand, writing the joint conditional probability explicitly, we have:

$$\begin{aligned}
q(x_1, y | x_2) &= \frac{q(x_1, x_2, y)}{q(x_2)} \\
&= \frac{q(x_1, y) q(x_2, y)}{q(y) q(x_2)} \\
&= q(x_1 | y) q(y | x_2).
\end{aligned} \tag{136}$$

Hence, for the partitioning to hold, we have the condition

$$q(x_1 | x_2) = q(x_1 | y), \tag{137}$$

since the RHS does not depend on  $x_2$  and the LHS does not depend on  $y$ , it implies that  $X_1 \perp\!\!\!\perp (X_2, Y)$  on  $Q$ . Hence, we can write

$$\begin{aligned} q(x_1, x_2, y) &= q(x_1)q(x_2, y) \\ \frac{p(x_1, y)p(x_2, y)}{p(y)} &= p(x_1)p(x_2, y) \\ p(x_1, y) &= p(x_1)p(y) \end{aligned} \tag{138}$$

which shows that  $X_1 \perp\!\!\!\perp Y$  on  $P$ . ■

**Proposition 17.**  $I_{\cap}^{\text{MES}}$  does not satisfy **(LB)** and **(SE)**.

*Proof.* For both cases, we use the result of Lemma 7. First, consider the PID with the two sources to be identical  $X_1 = X_2$  and not orthogonal to an otherwise generic target  $Y$ . Using Prop. 15, we have that

$$\begin{aligned} I_{\cap}^{\text{MES}}(X_1, X_2; Y) &= I_Q(X_1; X_2; Y) \\ &= I_Q(X_1; Y) - I_Q(X_1; Y|X_2) \\ &= I(X_1; Y) - I_Q(X_1; Y|X_2) \\ &< I(X_1; Y), \end{aligned} \tag{139}$$

where the strict inequality follows from Lemma 7 for any  $Y \not\perp\!\!\!\perp X_1$ , since  $I_Q(X_1; Y|X_2) > 0$ . Hence  $I_{\cap}^{\text{MES}}$  does not satisfy **(SE)**.

Similarly, if we consider the PID with first source  $X_1$  and second source the joint state  $(X_1 X_2)$  and a generic target  $Y$ , then the same reasoning above applies, and we obtain  $I_{\cap}^{\text{MES}}(X_1, X_1; Y) < I(X_1; Y)$  for any  $Y \not\perp\!\!\!\perp X_1$ . Hence  $I_{\cap}^{\text{MES}}$  does not satisfy **(SE)**. ■

**Lemma 8.** *The maximum entropy distribution of the TBC given pairwise constraints is the TBC itself.*

*Proof.* The proof directly follows from the closed-form solution for the maximum entropy distribution given pairwise constraints (see Prop. 15):

$$q(x_1, x_2, y) = q(x_1, x_2) = \frac{p(x_1, y)p(x_2, y)}{p(y)} = p(x_1, x_2) = p(x_1, x_2, y). \tag{140}$$

**Proposition 18.**  $I_{\cap}^{\text{MES}}$  satisfies **(CO)**.

*Proof.* For ease of readability, here we denote with  $X, Y$  the sources and  $Z$  the target. Consider two probability distributions  $P, P'$  with full support  $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ , and let

$$Q = \arg \max_{\tilde{Q} \in \Delta_P} H_{\tilde{Q}}(X, Y, Z) \quad Q' = \arg \max_{\tilde{Q} \in \Delta_{P'}} H_{\tilde{Q}}(X, Y, Z) \tag{141}$$

If  $\|Q - Q'\|_1 < \delta_Q$ , by the Fannes–Audenaert continuity bound for mutual information, we have that

$$\begin{aligned} \|I_{\cap, P}^{\text{MES}}(X, Y; Z) - I_{\cap, P'}^{\text{MES}}(X, Y; Z)\|_1 &= \|I_Q(X; Y) - I_{Q'}(X; Y)\|_1 \\ &\leq h(\delta_Q/2) + (\delta_Q/2) \log \min(|\mathcal{X}|, |\mathcal{Y}|) \\ &\equiv g(\delta_Q), \end{aligned} \tag{142}$$

where  $h(x) = -(x \log x + (1-x) \log(1-x))$  is the binary entropy function. The proof is then completed by showing that  $\|Q - Q'\|_1 \leq C\|P - P'\|_1$ :

$$\begin{aligned}
\|Q - Q'\|_1 &= \left\| \frac{p(x, z)p(y, z)}{p(z)} - \frac{p'(x, z)p'(y, z)}{p'(z)} \right\|_1 \\
&= \left\| \frac{p(x, z)p(y, z)p'(z) - p'(x, z)p'(y, z)p(z)}{p(z)p'(z)} \right\|_1 \\
&\leq \frac{1}{\epsilon} \|p(x, z)p(y, z)p'(z) - p'(x, z)p'(y, z)p(z)\|_1 \\
&\leq \frac{1}{\epsilon} \sum_{xyz} (|p'(z)| |p(x, z)p(y, z) - p'(x, z)p'(y, z)| + |p'(x, z)p'(y, z)| |p'(z) - p(z)|) \\
&\leq \frac{1}{\epsilon} \left( |\mathcal{Z}| \sum_{xyz} |p(x, z)p(y, z) - p'(x, z)p'(y, z)| + |\mathcal{Z}|^2 \sum_z |p'(z) - p(z)| \right) \\
&\leq \frac{1}{\epsilon} \left( |\mathcal{Z}| \sum_{xyz} (|p(x, z)| |p(y, z) - p'(y, z)| + |p'(y, z)| |p(x, z) - p'(x, z)|) + |\mathcal{Z}|^2 \|P - P'\|_1 \right) \\
&\leq \frac{1}{\epsilon} \left( |\mathcal{Z}|^2 \sum_{yz} |p(y, z) - p'(y, z)| + |\mathcal{Z}|^2 \sum_{xz} |p(x, z) - p'(x, z)| + |\mathcal{Z}|^2 \|P - P'\|_1 \right) \\
&\leq \frac{3|\mathcal{Z}|^2}{\epsilon} \|P - P'\|_1,
\end{aligned} \tag{143}$$

where we introduced  $\epsilon \equiv \min_{z \in \mathcal{Z}} (p(z), p'(z))$ , used that  $|ab - a'b'| \leq |a||b - b'| + |b'||a - a'|$  which follows from the triangle inequality, and used the fact that

$$\begin{aligned}
\|P - P'\|_1 &= \sum_{xyz} |p(x, y, z) - p'(x, y, z)| \\
&\geq \sum_{xz} \left| \sum_y (p(x, y, z) - p'(x, y, z)) \right| \\
&= \sum_{xz} |p(x, z) - p'(x, z)| \\
&= \|p(x, z) - p'(x, z)\|_1,
\end{aligned} \tag{144}$$

and similarly for the other marginals. Finally, to make the constant of Eq.(143) uniform, we define  $m \equiv \min_{z \in \mathcal{Z}} p(z)$  and consider a small neighbourhood of  $P$  given by  $\delta_P = m$ . This way for  $\|P - P'\|_1 < \delta_P$  we have that for every  $z$   $|p'(z) - p(z)| \leq \|P - P'\|_1 < m/2$ , and hence  $p'(z) \geq p(z) - m/2 \geq m/2$ . Thus  $\epsilon \geq m/2$ , and consequently, the constant in the above bound can be taken as

$$C_0 \equiv \frac{3|\mathcal{Z}|^2}{m/2} = \frac{6|\mathcal{Z}|^2}{m}, \tag{145}$$

so that for all  $P'$  with  $\|P - P'\|_1 < \delta_P$  we have the uniform inequality

$$\|Q - Q'\|_1 \leq C_0 \|P - P'\|_1. \tag{146}$$

Finally, taking  $\delta \equiv \min(\delta_Q, \delta_P)$ , we have that for  $\|P - P'\|_1 < \delta$ :

$$|I_{\square, P}^{\text{MES}}(X_1, X_2; Y) - I_{\square, P'}^{\text{MES}}(X_1, X_2; Y)| \leq g(\|Q - Q'\|_1) \leq g(C_0 \|P - P'\|_1) < \epsilon. \tag{147}$$

which is the definition of continuity. ■

**Proposition 19.**  $I_{\square}^{\text{MES}}$  does not satisfy (BP).

*Proof.* We prove this by showing that the requirements for Blackwell sufficiency and having zero unique information are different. Consider the Blackwell ordering  $X_1 \preceq_Y X_2$ , then  $X_1$  is a garbling of  $X_2$ :

$$p(x_1|y) = \sum_{x_2} k(x_1|x_2)p(x_2|y). \quad (148)$$

On the other hand, having vanishing unique information  $I_{\theta}^{\text{MES}}(X_1; Y)$  means

$$I(X_1; Y) = I_Q(X_1; X_2), \quad (149)$$

which implies

$$p(x_1|y) = q(x_1|x_2). \quad (150)$$

The two conditions of Eqs.(148)-(150) are clearly different, hence **(BP)** cannot hold. An easy way to see it is to consider a probability distribution  $P$  where  $X_1 \perp\!\!\!\perp X_2|Y$ , hence  $P = Q$  and Eq.(150) implies that  $p(x_1|\cdot)$  should be a constant not depending on the conditioning. Clearly, this is a stronger requirement than Eq. 148, hence Blackwell sufficiency cannot imply vanishing unique information.  $\blacksquare$

**Proposition 20.**  $I_{\cap}^{\text{MES}}$  satisfies **(AD)**.

*Proof.* Consider the PID with sources  $S_1 \equiv (X_1, X'_1)$  and  $S_2 \equiv (X_2, X'_2)$ , and target  $T \equiv (Y, Y')$ , where  $(X_1, X_2, Y_1) \perp\!\!\!\perp (X'_1, X'_2, Y')$ . It is easy to see that the solution to the maximum entropy projection with constraints  $q(S_1, T) = p(S_1, T)$  and  $q(S_2, T) = p(S_2, T)$  factorises as

$$\begin{aligned} q(s_1, s_2, t) &= \frac{p(s_1, t)p(s_2, t)}{p(t)} \\ &= \frac{p(x_1, x'_1, y, y')p(x_2, x'_2, y, y')}{p(y, y')} \\ &= \frac{p(x_1, y)p(x_2, y)}{p(y)} \frac{p(x'_1, y')p(x'_2, y')}{p(y')} \\ &= q(x_1, x_2, y)q(x'_1, x'_2, y'). \end{aligned} \quad (151)$$

Thus we have  $(X_1, X_2, Y_1) \perp\!\!\!\perp (X'_1, X'_2, Y')$  on  $Q$ , from which it follows that

$$\begin{aligned} I_{\cap}^{\text{MES}}((X_1, X'_1), (X_2, X'_2); YY') &= I_Q(X_1, X'_1; X_2, X'_2; YY') \\ &= I_Q(X_1; X_2; Y) + I_Q(X'_1; X'_2; Y') \\ &= I_{\cap}^{\text{MES}}(X_1, X_2; Y) + I_{\cap}^{\text{MES}}(X'_1, X'_2; Y'). \end{aligned} \quad (152)$$

### E.2.5 Proofs regarding $I_{\cap}^{\text{IG}}$

**Proposition 21.**  $I_{\cap}^{\text{IG}}$  satisfies **(S<sub>0</sub>)**.

*Proof.* This follows easily from the definition of unique information:

$$\begin{aligned} I_{\cap}^{\text{IG}}(X_1, X_2; Y) &= D(\bar{P}_1 || P_{X_1 X_2} P_Y) - D(P^* || \bar{P}_2) \\ &= D(\bar{P}_2 || P_{X_1 X_2} P_Y) - D(P^* || \bar{P}_1) \\ &= I_{\cap}^{\text{IG}}(X_2, X_1; Y) \end{aligned} \quad (153)$$

**Proposition 22.**  $I_{\cap}^{\text{IG}}$  satisfies **(M<sub>0</sub>)**.

*Proof.* The inequality condition follows from **(SR)** and the non-negativity of unique information, as it is a KL divergence:

$$I_{\cap}^{\text{IG}}(X_1, X_2; Y) \leq I_{\cap}^{\text{IG}}(X_i; Y) = I(X_i; Y) \leq I_{\cap}^{\text{IG}}(X_1 X_2; Y) = I(X_1 X_2; Y) \quad i = 1, 2, \quad (154)$$

so it only remains to check the equality condition, which, in a bivariate case, reduces to

$$X_1 = f(X_2) \implies I_{\cap}^{\text{IG}}(X_1, X_2; Y) = I_{\cap}^{\text{IG}}(X_1; Y) = I(X_1; Y). \quad (155)$$

We have that

$$\bar{P}_1 = \frac{P(X_1, Y)P(X_2)\delta(f(X_2), X_1)}{P(X_1)} \quad (156)$$

$$\bar{P}_2 = \frac{P(X_2, Y)P(X_2)\delta(f(X_2), X_1)}{P(X_2)} = P(X_2, Y)\delta(f(X_2), X_1) = P(X_1, X_2, Y). \quad (157)$$

Thus the  $P^*$  that minimises  $D(P||P^*)$  is exactly  $\bar{P}_2$ . This implies that:

$$I_{\partial}^{\text{IG}}(X_1\{X_1 X_2\}; Y) = D(P||P^*) = D(P||P) = 0 \quad (158)$$

$$I_{\partial}^{\text{IG}}(X_1; Y) = D(P^*||\bar{P}_2) = D(P||P) = 0 \quad (159)$$

$$I_{\partial}^{\text{IG}}(\{X_1 X_2\}; Y) = D(P^*||\bar{P}_1) = D(P||\bar{P}_1) = I(X_1 X_2; Y|X_1). \quad (160)$$

Hence  $I_{\cap}^{\text{IG}}(X_1, X_1 X_2; Y) = I(X_1; Y)$ , concluding the proof.  $\blacksquare$

**Proposition 23.**  $I_{\cap}^{\text{IG}}$  satisfies **(TE)**.

*Proof.* Considering the PID  $(X_1, X_2; Y)$  with  $X_2 = Y$ , we have

$$P(X_1, X_2, Y) = P(X_1, Y)\delta(X_2, Y) \quad (161)$$

$$\bar{P}_1 = \frac{P(X_1, Y)P(X_1, X_2)\delta(X_2, Y)}{P(X_1)} \quad (162)$$

$$\bar{P}_2 = \frac{P(X_2, Y)P(X_2, X_1)\delta(X_2, Y)}{P(Y)} = P(X_1, Y)\delta(X_2, Y). \quad (163)$$

Thus the  $P^*$  that minimises  $D(P||P^*)$  is exactly  $\bar{P}_2$ . Similarly to Prop. 22, this quickly implies that:

$$I_{\partial}^{\text{IG}}(X_1; Y) = D(P^*||\bar{P}_2) = D(P||P) = 0, \quad (164)$$

and hence  $I_{\cap}^{\text{IG}}(X_1, X_2; Y) = I(X_1; Y) = I_{\cap}^{\text{IG}}(X_1; Y)$ .  $\blacksquare$

**Proposition 24.**  $I_{\cap}^{\text{IG}}$  does not satisfy **(\*)**.

*Proof.* Recall the definition of redundancy (Eq. (89)):

$$I_{\cap}^{\text{IG}}(X_1, X_2; Y) = D(\bar{P}_1||P_{X_1 X_2} P_Y) - D(P^*||\bar{P}_2). \quad (165)$$

The first term is the mutual information  $I(X_1; Y)$ , and hence only depends on the marginal distribution  $P(X_1, Y)$ . On the other hand, writing the second term explicitly, we obtain:

$$\begin{aligned} D(P^*||\bar{P}_2) &= \sum_{x_1, x_2, y} \mathcal{N} \bar{p}_1^t \bar{p}_2^{1-t} \log \frac{\mathcal{N} \bar{p}_1^t \bar{p}_2^{1-t} p(x_1)}{p(x_1, x_2) p(x_1, y)} = \\ &= \sum_{x_1, x_2, y} \mathcal{N} p(x_1, x_2) p(y|x_1)^t p(y|x_2)^{1-t} \log \frac{\mathcal{N} p(x_1, x_2) p(y|x_1)^t p(y|x_2)^{1-t} p(x_1)}{p(x_1, x_2) p(x_1, y)} \\ &= \sum_{x_1, x_2, y} \mathcal{N} p(x_1, x_2) p(y|x_1)^t p(y|x_2)^{1-t} \log (\mathcal{N} p(y|x_1)^{t-1} p(y|x_2)^{1-t}), \end{aligned} \quad (166)$$

with  $\mathcal{N} = (p(x_1, x_2) p(y|x_1)^t p(y|x_2)^{1-t})^{-1}$ . Hence, although there is no explicit term including  $p(x_1, x_2)$  inside the logarithm, the redundancy evidently still depends on such a term in the summation and inside the normalisation constant  $\mathcal{N}$ . Thus  $I_{\cap}^{\text{IG}}$  does not satisfy **(\*)**.  $\blacksquare$

### E.2.6 Proofs regarding $I_{\cap}^{\text{CT}}$

**Proposition 25.**  $I_{\cap}^{\text{CT}}$  does not allow for mechanistic redundancy.

*Proof.* We prove that if  $X_1 \perp\!\!\!\perp X_2$  then  $I_{\cap}^{\text{CT}}(X_1, X_2; Y) = 0 \forall Y$ :

$$\begin{aligned} I_{\cap}^{\text{CT}}(X_1, X_2; Y) &= \sum_{ijk} p^{ijk} \log \left( \frac{\sum_l A_i^l B_l^k}{p^k} \right) \\ &= \sum_{ijk} p^{ijk} \log \left( \frac{\sum_l p^l B_l^k}{p^k} \right) \\ &= \sum_{ijk} p^{ijk} \log \left( \frac{p^k}{p^k} \right) = 0, \end{aligned} \tag{167}$$

where in the first step we assumed without loss of generality that the maximum inside the redundancy is obtained by  $I_{P_{X_1-Y}}\{X_1\}\{X_2\}\{Y\}$ , and in the second step we made use of the fact that  $X_1 \perp\!\!\!\perp X_2$  and hence  $A_i^j = p^j$ .  $\blacksquare$

**Proposition 26.**  $I_{\cap}^{\text{CT}}$  does not satisfy **(BP)**.

*Proof.* We prove the statement by showing that **(BP)** does not necessarily imply vanishing unique information. Consider the Blackwell ordering  $X_2 \preceq_Y X_1$ , this implies that  $I_{\cap}^{\text{CT}}(X_1, X_2; Y) \leq I(X_2; Y)$ , in fact to have equality we would need

$$\sum_m A_j^{\dagger m} C_m^k = B_j^k, \tag{168}$$

but the LHS can be written as

$$\sum_m A_j^{\dagger m} C_m^k = \sum_m p(m|x_2)p(y|m) = \sum_m \frac{p(x_2|m)p(m, y)}{p(x_2)} \tag{169}$$

where we used the following identity on the Blackwell order:

$$\begin{aligned} p(x_2|y) &= \sum_x k(x_2|x_1)p(x_1|y) \\ p(x_2, y) &= \sum_x k(x_2|x_1)p(x_1, y) \\ p(y|x_2) &= \sum_x \frac{k(x_2|x_1)p(x_1, y)}{p(x_2)}. \end{aligned} \tag{170}$$

On the other hand, the RHS is

$$B_j^k = p(y|x_2) = \sum_m \frac{k(x_2|m)p(m, y)}{p(x_2)}. \tag{171}$$

However, Eqs.(169)-(171) need not be the same, as the stochastic channel  $k$  can differ from the conditional probability distribution  $p$ . Since redundancy is smaller than the marginal mutual information, unique information cannot be zero.  $\blacksquare$

Although Prop. 26 denies **(BP)** for  $I_{\cap}^{\text{CT}}$ , one side of the implication still holds.

**Remark 1.** *Vanishing unique information in  $I_{\cap}^{\text{CT}}$  implies **(BP)**.*

*Proof.* We prove this by showing that if  $I_{\cap}^{\text{CT}}(X_1, X_2; Y) = I(X_2; Y)$ , and hence  $I_{\cap}^{\text{CT}}(X_2; Y) = 0$ , then  $X_1 \preceq_Y X_2$ . The general proof for an arbitrary number of sources is a direct extension of the following. First, we notice that the Blackwell ordering imposes that the cascade  $X_2 \rightarrow X_1 \rightarrow Y$  is equal to  $X_2 \rightarrow Y$ :

$$\sum_m A_j^{\dagger m} C_m^k = \sum_m p(m|x_2)p(y|m) = \sum_m \frac{p(x_2|m)p(m, y)}{p(x_2)} = p(y|x_2) = B_j^k, \tag{172}$$

where in the last step we used Eq.(170) and took the stochastic channel  $k$  to be the conditional probability  $p$ . With this at hand, we can write:

$$\begin{aligned}
I_{\cap}^{CT}(X_1, X_2; Y) &= \min(I_{P_{X_1-Y}}\{X_1\}\{X_2\}\{Y\}, I_{P_{X_2-Y}}\{X_2\}\{X_1\}\{Y\}) \\
&= I_{P_{X_2-Y}}\{X_2\}\{X_1\}\{Y\} \\
&= \sum_{ijk} p^{ijk} \log \left( \frac{\sum_m A_j^{\dagger m} C_m^k}{p^k} \right) \\
&= \sum_{ijk} p^{ijk} \log \left( \frac{B_j^k}{p^k} \right) \\
&= I(X_1; Y),
\end{aligned} \tag{173}$$

where in the first step we used that the cascade  $X_2 \rightarrow Y$  is directed ( $\sum_l A_i^l B_l^k \neq C_i^k$ ), so it is not a valid path for redundancy. ■

### E.2.7 Proofs regarding $I_{\cap}^{\prec}$

**Proposition 27.**  $I_{\cap}^{\prec}$  does not satisfy (TM)<sup>4</sup>.

*Proof.* The proof follows by a direct counterexample. Consider a TBC system with full support, i.e.  $Y = (X_1, X_2)$  and  $p(x_1, x_2) > 0 \forall x_1, x_2$ . In Ref. [38] it was shown that  $I_{\cap}^{\prec}(X_1, X_2; X_1, X_2)$  is equal to the Gács-Körner common information, which vanishes for distributions with full support. Hence  $I_{\cap}^{\prec}(X_1, X_2; X_1, X_2) = 0$ .

On the other hand, by (BP) we have that  $I_{\cap}^{\prec}(X_1, X_2; X_2) = I(X_1; X_2) > 0$ , where the strict positivity comes from the full-support requirement. Thus  $0 = I_{\cap}^{\prec}(X_1, X_2; X_1, X_2) < I_{\cap}^{\prec}(X_1, X_2; X_2) = I(X_1; X_2)$  and (TM) does not hold. ■

### E.2.8 Proofs regarding $I_{\cap}^{\delta}$

**Proposition 28.**  $I_{\cap}^{\delta}$  satisfies (M<sub>0</sub>).

*Proof.* The inequality condition follows from (LP<sub>0</sub>). For the equality condition, it is enough to note that  $X_2 = f(X_1)$  is a stronger condition than the Blackwell ordering  $X_2 \preceq_Y X_1$ , and hence from (BP) we have  $I_{\cap}^{\delta}(X_2; Y) = 0$ , and thus  $I_{\cap}^{\delta}(X_1, X_2; Y) = I(X_2; Y) = I_{\cap}^{\delta}(X_2; Y)$ .

*Alternative proof for equality condition.* For the equality condition, we need to prove that  $I_{\cap}^{\delta}(X_1, f(X_1); Y) = I_{\cap}^{\delta}(f(X_1); Y)$ . Consider  $X_2 = f(X_1)$ , then we have

$$\begin{aligned}
I_{\cap}^{\delta}(X_1, f(X_1); Y) &= \min\left(I(X_1; Y) - \delta(Y; X_1 \setminus f(X_1)), I(f(X_1); Y) - \delta(Y; f(X_1) \setminus X_1)\right) \\
&= \min\left(I(X_1; Y) - \delta(Y; X_1 \setminus f(X_1)), I(f(X_1); Y)\right) \\
&= \min\left(I(X_1; Y) - I(X_1; Y|f(X_1)), I(f(X_1); Y)\right) \\
&= I(f(X_1); Y) \\
&= I(X_2; Y),
\end{aligned} \tag{174}$$

---

<sup>4</sup>We thank Artemy Kolchinsky for the suggestion of the reported counterexample.

where we used that:

$$\begin{aligned}
\delta(Y; X_1 \setminus f(X_1)) &= \inf_k \mathbb{E}_{P_Y} [D(P_{X_1|Y} \| K_{X_1|f(X_1)} \circ P_{f(X_1)|Y})] \\
&= \inf_k \mathbb{E}_{P_Y} \left[ \int dx_1 p(x_1|y) \log \frac{p(x_1|y)}{\int df(x_1)' k(x_1|f(x_1)') p(f(x_1)'|y)} \right] \\
&= \mathbb{E}_{P_Y} \left[ \int dx_1 p(x_1|y) \log \frac{p(x_1|y)}{\int df(x_1) \frac{p(x_1)p(f(x_1)'|x_1)}{p(f(x_1)')} p(f(x_1)'|y)} \right] \\
&= \mathbb{E}_{P_Y} \left[ \int dx_1 p(x_1|y) \log \frac{p(x_1|y)p(f(x_1))}{p(x_1)p(f(x_1)|y)} \right] \\
&= \mathbb{E}_{P_Y} \left[ \int dx_1 p(x_1|y) \log \frac{p(x_1|f(x_1), y)}{p(x_1|f(x_1))} \right] \\
&= I(X_1; Y|f(X_1)),
\end{aligned} \tag{175}$$

where we used that the infimum is attained by the true channel  $p(x_1|f(x_1))$ , and that  $p(x_1, f(x_1)) = p(x_1)$ .  $\blacksquare$

**Proposition 29.**  $I_{\cap}^{\delta}$  satisfies **(TE)**.

*Proof.* The proof follows trivially from the Blackwell ordering relation. Considering any two stochastic variables  $(X_1, Y)$ , we have that  $X_1 \preceq_Y Y$ :

$$\int dy' k(x|y') p(y'|y) = \int dy' k(x|y') \delta(y', y) = k(x|y) = p(x|y). \tag{176}$$

Hence it follows that  $\delta(Y; X_1 \setminus Y) = 0$  and

$$I_{\cap}^{\delta}(X_1, Y; Y) = I(X_1; Y) = I_{\cap}^{\delta}(X_1; Y). \tag{177}$$

**Proposition 30.**  $I_{\cap}^{\delta}$  satisfies **(ID)**.

*Proof.* Consider the TBC system where the target is  $Y = (X_1, X_2)$ . From the definition, we have that the deficiency of  $X_1$  w.r.t.  $X_2$  about  $Y$  is

$$\begin{aligned}
\delta(Y; X_1 \setminus X_2) &= \inf_k \mathbb{E}_{P_Y} [D(p(X_1|y) \| k(X_1|X_2) \circ p(X_2|y))] \\
&= \inf_k \mathbb{E}_{P_Y} \left[ \int dx'_1 p(x'_1|y) \log \left( \frac{p(x'_1|y)}{\int k(x'_1|x'_2) p(x'_2|y) dx'_2} \right) \right] \\
&= \inf_k \int dx_1 dx_2 p(x_1, x_2) \left[ \int dx'_1 \delta(x'_1 - x_1) \log \left( \frac{\delta(x'_1 - x_1)}{\int k(x_1|x'_2) \delta(x'_2 - x_2) dx'_2} \right) \right] \\
&= \inf_k \int dx_1 dx_2 p(x_1, x_2) [-\log k(x_1|x_2)] \\
&= \inf_k \int dx_2 p(x_2) \int dx_1 p(x_1|x_2) [-\log k(x_1|x_2)] \\
&= \int dx_2 p(x_2) H(X_1|X_2 = x_2) \\
&= H(X_1|X_2),
\end{aligned} \tag{178}$$

where the infimum over the stochastic channel  $k$  follows from Gibb's inequality [8]. An analogous derivation holds for  $\delta(Y; X_2 \setminus X_1)$ . Hence, we have that

$$\begin{aligned}
I_{\cap}^{\delta}(X_1, X_2; Y) &= \min(I(X_1; Y) - \delta(Y; X_1 \setminus X_2), I(X_2; Y) - \delta(Y; X_2 \setminus X_1)) \\
&= \min(H(X_1) - H(X_1|X_2), H(X_2) - H(X_2|X_1)) \\
&= \min(I(X_1; X_2), I(X_2; X_1)) \\
&= I(X_1; X_2),
\end{aligned} \tag{179}$$

which concludes the proof. ■

### E.2.9 Proofs regarding $I_{\cap}^{\text{do}}$

Although in the original publication the authors of  $I_{\cap}^{\text{do}}$  claimed that the measure satisfies monotonicity and self-redundancy (Axiom 3 in Ref. [41]), we point out that their proof does not effectively show self-redundancy and does not check the subset equality condition ((**SE**)) in the monotonicity definition, i.e. that  $I_{\cap}(X_1, X_2; Y) = I_{\cap}(X_1; Y)$  if  $X_1 = f(X_2)$ . In fact, it turns out that  $I_{\cap}^{\text{do}}$  does not satisfy (**SE**). Below, we discuss what implications this has on (**SR**) and (**M<sub>0</sub>**).

As for (**SR**), we underline that a PID measure based on the definition of unique information can be built so that it satisfies the (**SR**) axiom constructively [27], since there is otherwise no way to prove directly that  $I_{\cap}(X_1; Y) = I(X_1; Y)$ , as one cannot use the definition of unique information for a single source. Hence, we make the following remark:

**Remark 2.**  $I_{\cap}^{\text{do}}$  satisfies (**SR**) constructively, i.e. we define  $I_{\cap}^{\text{do}}(X_1; Y) \equiv I(X_1; Y)$ .

We can use this constructive definition to rewrite  $I_{\cap}^{\text{do}}$  in a more convenient form.

**Lemma 9.** *The redundancy function can be written as a coinformation:  $I_{\cap}^{\text{do}}(X_1, X_2; Y) = I(X'_1; X_2; Y) = I(X_1; X'_2; Y)$*

*Proof.* The proof readily follows from the definition of unique information and (**SR**):

$$\begin{aligned} I_{\cap}^{\text{do}}(X_1, X_2; Y) &= I_{\cap}^{\text{do}}(X_1; Y) - I_{\cap}^{\text{do}}(X'_1; Y|X_2) \\ &= I(X_1; Y) - I_{\cap}^{\text{do}}(X'_1; Y|X_2) \\ &= I(X'_1; Y) - I_{\cap}^{\text{do}}(X'_1; Y|X_2) \\ &= I(X_1; X'_2; Y), \end{aligned} \tag{180}$$

where we used that  $I(X'_1; Y) = I(X_1; Y)$  since  $H(X) = H(X')$  and  $H(X|Y) = H(X'|Y)$ . An analogous procedure would give  $I_{\cap}^{\text{do}}(X_1, X_2; Y) = I(X_1; X'_2; Y)$ . ■

Using this Lemma we can also obtain Lemma 2 of Ref. [41] by noting that  $X'_1$  and  $X_2$  are conditionally independent given  $Y$ , in fact:

$$\begin{aligned} p(X'_1 = x_1, X_2 = x_2|Y = y) &= \frac{p(X_2 = x_2)}{p(Y = y)} p(X'_1 = x_1, Y = y|X_2 = x_2) \\ &= \frac{p(X_2 = x_2)}{p(Y = y)} p(X_1 = x_1|Y = y) p(Y = y|X_2 = x_2) \\ &= p(X_1 = x_1|Y = y) p(X_2 = x_2|Y = y). \end{aligned} \tag{181}$$

Hence we have

$$I_{\cap}^{\text{do}}(X_1, X_2; Y) = I(X'_1; X_2; Z) = I(X'_1; X_2) - I(X'_1; X_2|Y) = I(X'_1; X_2). \tag{182}$$

On the other hand, the lack of the equality condition in the monotonicity definition has an important consequence, which cannot be so easily alleviated:  $I_{\cap}^{\text{do}}$  assigns non-zero unique information even when the sources are the same. Before proving this result, we provide a useful Lemma.

**Lemma 10.** *Consider the system  $(X_1, X_2, Y)$  and the random variable  $X'_1$  as defined in Eq.(105). Then  $I_Q(X'_1; Y|X_2) = 0$  if and only if  $X_1 \perp\!\!\!\perp Y$ .*

*Proof.*  $I_Q(X'_1; Y|X_2) = 0$  means that  $p(x'_1, y|x_2) = p(x'_1|x_2)p(y|x_2)$ . This happens when

$$p(x'_1, y|x_2) = p(x_1|y)p(y|x_2) \tag{183}$$

$$p(x'_1|x_2) = \sum_{\tilde{y}} p(x'_1, \tilde{y}|x_2) = \sum_{\tilde{y}} p(x_1|\tilde{y})p(\tilde{y}|x_2). \tag{184}$$

Hence, for the partitioning to hold, we have the condition

$$p(x_1|y) = \sum_{\tilde{y}} p(x_1|\tilde{y})p(\tilde{y}|x_2), \quad (185)$$

since the RHS does not depend on  $y$ , it implies that  $X_1 \perp\!\!\!\perp Y$ . ■

Using this Lemma, we have:

**Proposition 31.**  $I_{\cap}^{\text{do}}$  does not satisfy **(SE)**.

*Proof.* Consider the PID with identical first and second source  $X_1 = X_2$  and generic target  $Y$ . From Lemma 9 we have that

$$\begin{aligned} I_{\cap}^{\text{do}}(X_1, X_1; Y) &= I(X'_1; Y) - I(X'_1; Y|X_2) \\ &= I(X_1; Y) - I(X'_1; Y|X_2) \\ &< I(X_1; Y), \end{aligned} \quad (186)$$

where the strict inequality follows from Lemma 10 when  $X_1 \not\perp\!\!\!\perp Y$ , since  $I(X'_1; Y|X_2) > 0$ . Hence  $I_{\cap}^{\text{do}}$  does not satisfy **(SE)**. ■

A similar reasoning can be used for the following property.

**Proposition 32.**  $I_{\cap}^{\text{do}}$  does not satisfy **(LB)**.

*Proof.* We prove it by counterexample. Consider the PID with first source  $X_1$  and second source the joint state  $X_1X_2$  and a generic target  $Y$ , then using Lemma 9 we have

$$\begin{aligned} I_{\cap}^{\text{do}}(X_1, X_1X_2; Y) &= I(X'_1; Y) - I(X'_1; Y|X_1X_2) \\ &= I(X_1; Y) - I(X'_1; Y|X_1X_2) \\ &< I(X_1; Y), \end{aligned} \quad (187)$$

where the strict inequality follows from Lemma 10 when  $X_1 \not\perp\!\!\!\perp Y$ . On the other hand, **(LB)** would require  $I_{\cap}^{\text{do}}(X_1, X_1X_2; Y) \geq I(X_1; Y)$  since  $X_1 = f(X_1X_2)$ . Hence  $I_{\cap}^{\text{do}}$  does not satisfy **(LB)**. ■

**Proposition 33.**  $I_{\cap}^{\text{do}}$  satisfies **(ID)**.

*Proof.* Considering the TBC system, and denoting as  $y = (\tilde{x}_1, \tilde{x}_2)$  the outcome of the target, we have

$$\begin{aligned} p(X'_1 = x_1, Y = y|X_2 = x_2) &= p(X_1 = x_1|Y = y)p(Y = y|X_2 = x_2) \\ &= p(X_1 = x_1|X_1 = \tilde{x}_1, X_2 = \tilde{x}_2)p(X_1 = \tilde{x}_1, X_2 = \tilde{x}_2|X_2 = x_2) \\ &= p(X_1 = x_1|X_1 = x_1, X_2 = x_2)p(X_1 = x_1, X_2 = x_2|X_2 = x_2)\delta(\tilde{x}_1, x_1)\delta(\tilde{x}_2, x_2) \\ &= p(X_1 = x_1, X_2 = x_2|X_2 = x_2), \end{aligned} \quad (188)$$

which means that  $X'_1 = X_1$ . Hence **(ID)** follows from Eq.(107). ■

### E.2.10 Proofs regarding $I_{\cap}^{\text{RAV}}$

The following proofs only hold for bivariate PIDs ( $n = 2$  sources).

**Proposition 34.**  $I_{\cap}^{\text{RAV}}$  satisfies **(SR)**.

*Proof.*

$$\begin{aligned} I_{\cap}^{\text{RAV}}(X_1; Y) &= \max_f I(X_1; Y; f(X_1)) \\ &= \max_f (I(f(X_1); Y) - I(f(X_1); Y|X_1)) \\ &= \max_f I(f(X_1); Y) \\ &= I(X_1; Y). \end{aligned} \quad (189)$$

■

**Proposition 35.**  $I_{\cap}^{\text{RAV}}$  satisfies  $(\mathbf{M}_0)$  for  $n = 2$  sources.

*Proof.* For  $n = 2$  sources, for the inequality condition we only need to prove  $I_{\cap}^{\text{RAV}}(X_i; Y) - I_{\cap}^{\text{RAV}}(X_1, X_2; Y) \geq 0$ . Considering e.g.  $i = 1$ :

$$\begin{aligned} I_{\cap}^{\text{RAV}}(X_i; Y) - I_{\cap}^{\text{RAV}}(X_1, X_2; Y) &= I(X_1; Y) - \max_f I(X_1; X_2; Y; f(X_1, X_2)) \\ &= I(X_1; Y) - I(X_1; X_2; Y) + \min_f I(X_1; X_2; Y|f(X_1, X_2)) \\ &= I(X_1; Y|X_2) + \min_f I(X_1; X_2; Y|f(X_1, X_2)) \geq 0. \end{aligned} \quad (190)$$

For the equality condition, if we have  $X_2 = g(X_1)$  then

$$I_{\cap}^{\text{RAV}}(X_1, X_2; Y) = \max_f I(X_1; g(X_1); Y; f(X_1, X_2)) = \max_f I(X_1; Y; f(X_1)) = I_{\cap}^{\text{RAV}}(X_1; Y), \quad (191)$$

which concludes the proof. ■

**Proposition 36.**  $I_{\cap}^{\text{RAV}}$  satisfies  $(\mathbf{ID})$ .

*Proof.* Consider the joint PID target  $Y = (X_1, X_2)$ . It follows that

$$\begin{aligned} I_{\cap}^{\text{RAV}}(X_1, X_2; Y) &= \max_f I(X_1; X_2; Y; f(X_1, X_2)) \\ &= \max_f (I(X_1; X_2; Y; f(X_1, X_2)) - I(X_1; X_2; f(X_1, X_2)|X_1 X_2)) \\ &= \max_f (I(X_1; X_2) - I(X_1, X_2|f(X_1, X_2))) \\ &= I(X_1; X_2), \end{aligned} \quad (192)$$

where in the last step it is enough to take  $f(X_1, X_2) = X_i$ . ■

**Proposition 37.**  $I_{\cap}^{\text{RAV}}$  satisfies  $(\mathbf{TE})$  for  $n = 2$  sources.

*Proof.* For  $n = 2$  we just need to prove that  $I_{\cap}^{\text{RAV}}(X_1, Y; Y) = I_{\cap}^{\text{RAV}}(X_1; Y)$ .

$$\begin{aligned} I_{\cap}^{\text{RAV}}(X_1, Y; Y) &= \max_f I(X_1; Y; Y; f(X_1, Y)) \\ &= \max_f I(X_1; Y; f(X_1, Y)) \\ &= I(X_1; Y) - \min_f I(X_1; Y; Y|f(X_1, Y)) \\ &= I(X_1; Y), \end{aligned} \quad (193)$$

where we took  $f(X_1, Y) = \{X_1, Y\}$ . ■

**Proposition 38.**  $I_{\cap}^{\text{RAV}}$  satisfies  $(\mathbf{LP}_0)$ .

*Proof.* Redundancy is non-negative from  $(\mathbf{GP})$ , and unique information is non-negative from  $(\mathbf{M}_0)$  and  $(\mathbf{GP})$ . It remains to prove non-negativity for synergy:

$$\begin{aligned} I_{\partial}(X_1 X_2; Y) &= I_{\cap}^{\text{RAV}}(X_1, X_2; Y) - I(X_1; X_2; Y) \\ &= I(X_1; X_2; Y) - \min_f I(X_1; X_2; Y|f(X_1, X_2)) - I(X_1; X_2; Y) \\ &= \max_f (-I(X_1; X_2; Y|f(X_1, X_2))) \\ &= \max_f (I(X_1; Y|X_2, f(X_1, X_2)) - I(X_1; Y|f(X_1, X_2))) \\ &\geq I(X_1; Y|X_2, X_2) - I(X_1; Y|X_2) \\ &= 0, \end{aligned} \quad (194)$$

where in the second last step we took  $f(X_1, X_2) = X_2$ . ■

## References

- [1] C. E. Shannon, “A mathematical theory of communication,” *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.
- [2] H. Nyquist, “Certain factors affecting telegraph speed 1,” *Bell System Technical Journal*, vol. 3, no. 2, pp. 324–346, 1924.
- [3] R. V. Hartley, “Transmission of information 1,” *Bell System technical journal*, vol. 7, no. 3, pp. 535–563, 1928.
- [4] R. Clausius, *Ueber verschiedene für die Anwendung bequeme Formen der Hauptgleichungen der mechanischen Wärmetheorie: vorgetragen in der naturforsch. Gesellschaft den 24. April 1865*. Verlag nicht ermittelbar, 1865.
- [5] L. Boltzmann, “Weitere studien über das wärmeleichgewicht unter gasmolekülen,” *Sitzungsberichte der Kaiserlichen Akademie der Wissenschaften in Wien*, vol. 66, no. 2, pp. 275–370, 1872.
- [6] L. Boltzmann, “Weitere studien über das wärmeleichgewicht unter gasmolekülen,” in *Kinetische Theorie II: Irreversible Prozesse Einführung und Originaltexte*, pp. 115–225, Springer, 1970.
- [7] E. T. Jaynes, “Information theory and statistical mechanics,” *Physical review*, vol. 106, no. 4, p. 620, 1957.
- [8] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, 1999.
- [9] R. G. James and J. P. Crutchfield, “Multivariate dependence beyond shannon information,” *Entropy*, vol. 19, no. 10, p. 531, 2017.
- [10] P. L. Williams and R. D. Beer, “Nonnegative decomposition of multivariate information,” *arXiv preprint arXiv:1004.2515*, 2010.
- [11] F. E. Rosas, P. A. Mediano, A. I. Luppi, T. F. Varley, J. T. Lizier, S. Stramaglia, H. J. Jensen, and D. Marinazzo, “Disentangling high-order mechanisms and high-order behaviours in complex systems,” *Nature Physics*, vol. 18, no. 5, pp. 476–477, 2022.
- [12] A. I. Luppi, P. A. Mediano, F. E. Rosas, N. Holland, T. D. Fryer, J. T. O’Brien, J. B. Rowe, D. K. Menon, D. Bor, and E. A. Stamatakis, “A synergistic core for human brain evolution and cognition,” *BioRxiv*, 2020.
- [13] K. J. Down, J. Huntley, P. A. Mediano, and D. Bor, “Synergistic and redundant information dynamics are modulated by alzheimer’s disease and cognitive impairment,” *bioRxiv*, pp. 2026–02, 2026.
- [14] M. Neri, A. Brovelli, S. Castro, F. Fraioli, M. Gatica, R. Herzog, P. A. Mediano, I. Mindlin, G. Petri, D. Bor, *et al.*, “A taxonomy of neuroscientific strategies based on interaction orders,” *European Journal of Neuroscience*, vol. 61, no. 3, p. e16676, 2025.
- [15] E. Combrisson, R. Basanisi, M. Neri, G. Auzias, G. Petri, D. Marinazzo, S. Panzeri, and A. Brovelli, “Higher-order and distributed synergistic functional interactions encode information gain in goal-directed learning,” *Nature Communications*, vol. 16, no. 1, p. 7179, 2025.
- [16] F. Battiston, G. Cencetti, I. Iacopini, V. Latora, M. Lucas, A. Patania, J.-G. Young, and G. Petri, “Networks beyond pairwise interactions: Structure and dynamics,” *Physics reports*, vol. 874, pp. 1–92, 2020.
- [17] T. Robiglio, M. Neri, D. Coppes, C. Agostinelli, F. Battiston, M. Lucas, and G. Petri, “Synergistic signatures of group mechanisms in higher-order systems,” *Physical review letters*, vol. 134, no. 13, p. 137401, 2025.
- [18] M. Harder, C. Salge, and D. Polani, “Bivariate measure of redundant information,” *Physical Review E*, vol. 87, no. 1, p. 012130, 2013.

- [19] N. Bertschinger, J. Rauh, E. Olbrich, and J. Jost, “Shared information—new insights and problems in decomposing information in complex systems,” in *Proceedings of the European conference on complex systems 2012*, pp. 251–269, Springer, 2013.
- [20] V. Griffith, E. K. Chong, R. G. James, C. J. Ellison, and J. P. Crutchfield, “Intersection information based on common randomness,” *Entropy*, vol. 16, no. 4, pp. 1985–2000, 2014.
- [21] N. Bertschinger, J. Rauh, E. Olbrich, J. Jost, and N. Ay, “Quantifying unique information,” *Entropy*, vol. 16, no. 4, pp. 2161–2183, 2014.
- [22] V. Griffith and C. Koch, “Quantifying synergistic mutual information,” in *Guided self-organization: inception*, pp. 159–190, Springer, 2014.
- [23] V. Griffith and T. Ho, “Quantifying redundant information in predicting a target random variable,” *Entropy*, vol. 17, no. 7, pp. 4644–4653, 2015.
- [24] A. B. Barrett, “Exploration of synergistic and redundant information sharing in static and dynamical gaussian systems,” *Physical Review E*, vol. 91, no. 5, p. 052802, 2015.
- [25] A. E. Goodwell and P. Kumar, “Temporal information partitioning: Characterizing synergy, uniqueness, and redundancy in interacting environmental variables,” *Water Resources Research*, vol. 53, no. 7, pp. 5920–5942, 2017.
- [26] R. A. Ince, “Measuring multivariate redundant information with pointwise common change in surprisal,” *Entropy*, vol. 19, no. 7, p. 318, 2017.
- [27] R. G. James, J. Emenheiser, and J. P. Crutchfield, “Unique information via dependency constraints,” *Journal of Physics A: Mathematical and Theoretical*, vol. 52, no. 1, p. 014002, 2018.
- [28] J. W. Kay and R. A. Ince, “Exact partial information decompositions for gaussian systems based on dependency constraints,” *Entropy*, vol. 20, no. 4, p. 240, 2018.
- [29] R. G. James, J. Emenheiser, and J. P. Crutchfield, “Unique information and secret key agreement,” *Entropy*, vol. 21, no. 1, p. 12, 2018.
- [30] C. Finn and J. T. Lizier, “Pointwise partial information decomposition using the specificity and ambiguity lattices,” *Entropy*, vol. 20, no. 4, p. 297, 2018.
- [31] X. Niu and C. J. Quinn, “A measure of synergy, redundancy, and unique information using information geometry,” in *2019 IEEE International Symposium on Information Theory (ISIT)*, pp. 3127–3131, IEEE, 2019.
- [32] J. W. Kay, “A partial information decomposition for multivariate gaussian systems based on information geometry,” *Entropy*, vol. 26, no. 7, p. 542, 2024.
- [33] D. Sigtermans, “A partial information decomposition based on causal tensors,” *arXiv preprint arXiv:2001.10481*, 2020.
- [34] D. Sigtermans, “Towards a framework for observational causality from time series: when shannon meets turing,” *Entropy*, vol. 22, no. 4, p. 426, 2020.
- [35] A. Makkeh, A. J. Gutknecht, and M. Wibral, “Introducing a differentiable measure of pointwise shared information,” *Physical Review E*, vol. 103, no. 3, p. 032149, 2021.
- [36] K. Schick-Poland, A. Makkeh, A. J. Gutknecht, P. Wollstadt, A. Sturm, and M. Wibral, “A partial information decomposition for discrete and continuous variables,” *arXiv preprint arXiv:2106.12393*, 2021.

- [37] D. A. Ehrlich, K. Schick-Poland, A. Makkeh, F. Lanfermann, P. Wollstadt, and M. Wibral, “Partial information decomposition for continuous variables based on shared exclusions: Analytical formulation and estimation,” *Physical Review E*, vol. 110, no. 1, p. 014115, 2024.
- [38] A. Kolchinsky, “A novel approach to the partial information decomposition,” *Entropy*, vol. 24, no. 3, p. 403, 2022.
- [39] P. Venkatesh and G. Schamberg, “Partial information decomposition via deficiency for multivariate gaussians,” in *2022 IEEE International Symposium on Information Theory (ISIT)*, pp. 2892–2897, IEEE, 2022.
- [40] T. Mages and C. Rohner, “Decomposing and tracing mutual information by quantifying reachable decision regions,” *Entropy*, vol. 25, no. 7, p. 1014, 2023.
- [41] A. Lyu, A. Clark, and N. Raviv, “Explicit formula for partial information decomposition,” in *2024 IEEE International Symposium on Information Theory (ISIT)*, pp. 2329–2334, IEEE, 2024.
- [42] J. Rauh, N. Bertschinger, E. Olbrich, and J. Jost, “Reconsidering unique information: Towards a multivariate information decomposition,” in *2014 IEEE International Symposium on Information Theory*, pp. 2232–2236, IEEE, 2014.
- [43] L. De Moura and N. Bjørner, “Z3: An efficient smt solver,” in *International conference on Tools and Algorithms for the Construction and Analysis of Systems*, pp. 337–340, Springer, 2008.
- [44] G.-C. Rota, “On the foundations of combinatorial theory: I. theory of möbius functions,” in *Classic Papers in Combinatorics*, pp. 332–360, Springer, 1964.
- [45] A. Jansma, P. A. Mediano, and F. E. Rosas, “Fast möbius transform: An algebraic approach to information decomposition,” *Physical Review Research*, vol. 7, no. 3, p. 033049, 2025.
- [46] A. J. Gutknecht, M. Wibral, and A. Makkeh, “Bits and pieces: Understanding information decomposition from part-whole relationships and formal logic,” *Proceedings of the Royal Society A*, vol. 477, no. 2251, p. 20210110, 2021.
- [47] R. D. Beer and P. L. Williams, “Information processing and dynamics in minimally cognitive agents,” *Cognitive science*, vol. 39, no. 1, pp. 1–38, 2015.
- [48] M. Wibral, V. Priesemann, J. W. Kay, J. T. Lizier, and W. A. Phillips, “Partial information decomposition as a unified approach to the specification of neural goal functions,” *Brain and cognition*, vol. 112, pp. 25–38, 2017.
- [49] T. M. Tax, P. A. Mediano, and M. Shanahan, “The partial information decomposition of generative neural network models,” *Entropy*, vol. 19, no. 9, p. 474, 2017.
- [50] A. M. Proca, F. E. Rosas, A. I. Luppi, D. Bor, M. Crosby, and P. A. Mediano, “Synergistic information supports modality integration and flexible learning in neural networks solving multiple tasks,” *PLOS Computational Biology*, vol. 20, no. 6, p. e1012178, 2024.
- [51] T. F. Varley, M. Pope, M. Grazia, Joshua, and O. Sporns, “Partial entropy decomposition reveals higher-order information structures in human brain activity,” *Proceedings of the National Academy of Sciences*, vol. 120, no. 30, p. e2300888120, 2023.
- [52] T. F. Varley, M. Pope, J. Faskowitz, and O. Sporns, “Multivariate information theory uncovers synergistic subsystems of the human cerebral cortex,” *Communications biology*, vol. 6, no. 1, p. 451, 2023.
- [53] A. I. Luppi, F. E. Rosas, P. A. Mediano, D. K. Menon, and E. A. Stamatakis, “Information decomposition and the informational architecture of the brain,” *Trends in Cognitive Sciences*, 2024.

- [54] M. Gatica, R. Cofré, P. A. Mediano, F. E. Rosas, P. Orio, I. Diez, S. P. Swinnen, and J. M. Cortes, “High-order interdependencies in the aging brain,” *Brain connectivity*, vol. 11, no. 9, pp. 734–744, 2021.
- [55] M. Gatica, F. E. Rosas, P. AM Mediano, I. Diez, S. P. Swinnen, P. Orio, R. Cofré, and J. M. Cortes, “High-order functional redundancy in ageing explained via alterations in the connectome in a whole-brain model,” *PLoS Computational Biology*, vol. 18, no. 9, p. e1010431, 2022.
- [56] C. Coronel-Oliveros, M. Gatica, R. Herzog, and M. Neri, “An integrated computational approach for diversity-sensitive personalized medicine,” *Neuroscience*, 2025.
- [57] A. Santoro, M. Neri, S. Poetto, D. Orsenigo, M. Diano, M. Gatica, and G. Petri, “Beyond pairwise interactions: Charting higher-order models of brain function,” *bioRxiv*, pp. 2025–06, 2025.
- [58] T. E. Chan, M. P. Stumpf, and A. C. Babbie, “Gene regulatory network inference from single-cell data using multivariate information measures,” *Cell systems*, vol. 5, no. 3, pp. 251–267, 2017.
- [59] S. Chen and J. C. Mar, “Evaluating methods of inferring gene regulatory networks highlights their lack of performance for single cell gene expression data,” *BMC bioinformatics*, vol. 19, no. 1, pp. 1–21, 2018.
- [60] C. Finn and J. T. Lizier, “Quantifying information modification in cellular automata using pointwise partial information decomposition,” in *Artificial Life Conference Proceedings*, pp. 386–387, MIT Press One Rogers Street, Cambridge, MA 02142-1209, USA journals-info . . . , 2018.
- [61] F. Rosas, P. A. Mediano, M. Ugarte, and H. J. Jensen, “An information-theoretic approach to self-organisation: Emergence of complex interdependencies in coupled dynamical systems,” *Entropy*, vol. 20, no. 10, p. 793, 2018.
- [62] H. Rajpal and O. Guerrero, “Synergistic small worlds that drive technological sophistication,” *PNAS nexus*, vol. 4, no. 4, p. pgaf102, 2025.
- [63] J. Pearl, *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Elsevier, 2014.
- [64] D. Chicharro, G. Pica, and S. Panzeri, “The identity of information: How deterministic dependencies constrain information synergy and redundancy,” *Entropy*, vol. 20, no. 3, p. 169, 2018.
- [65] P. H. Matthias, A. Makkeh, M. Wibral, and A. J. Gutknecht, “Novel inconsistency results for partial information decomposition,” *arXiv preprint arXiv:2512.16662*, 2025.
- [66] P. K. Banerjee and V. Griffith, “Synergy, redundancy and common information,” *arXiv preprint arXiv:1509.03706*, 2015.
- [67] J. Rauh, P. K. Banerjee, E. Olbrich, J. Jost, and N. Bertschinger, “On extractable shared information,” *Entropy*, vol. 19, no. 7, p. 328, 2017.
- [68] J. Rauh, P. K. Banerjee, E. Olbrich, G. Montúfar, and J. Jost, “Continuity and additivity properties of information decompositions,” *International Journal of Approximate Reasoning*, vol. 161, p. 108979, 2023.
- [69] D. Blackwell, “Equivalent comparisons of experiments,” *The annals of mathematical statistics*, pp. 265–272, 1953.
- [70] J. Rauh, “Secret sharing and shared information,” *Entropy*, vol. 19, no. 11, p. 601, 2017.
- [71] A. J. Gutknecht, A. Makkeh, and M. Wibral, “From babel to boole: the logical organization of information decompositions,” *Proceedings of the Royal Society A*, vol. 481, no. 2310, p. 20240174, 2025.
- [72] C. Finn and J. T. Lizier, “Probability mass exclusions and the directed components of mutual information,” *Entropy*, vol. 20, no. 11, p. 826, 2018.
- [73] M. Zwick, “An overview of reconstructability analysis,” *Kybernetes*, vol. 33, no. 5/6, pp. 877–905, 2004.

- [74] K. Krippendorff, “Ross ashby’s information theory: a bit of history, some solutions to problems, and what we face today,” *International journal of general systems*, vol. 38, no. 2, pp. 189–212, 2009.
- [75] S.-I. Amari, “Information geometry on hierarchy of probability distributions,” *IEEE transactions on information theory*, vol. 47, no. 5, pp. 1701–1711, 2001.
- [76] J. Pearl, “Causal diagrams for empirical research,” *Biometrika*, vol. 82, no. 4, pp. 669–688, 1995.
- [77] J. Pearl, “Causal inference in statistics: An overview,” *Statistics Surveys*, vol. 3, pp. 96–146, 2009.
- [78] J. Pearl, “The do-calculus revisited,” in *Proceedings of the Twenty-Eighth Conference on Uncertainty in Artificial Intelligence*, pp. 3–11, 2012.
- [79] R. G. James, C. J. Ellison, and J. P. Crutchfield, ““dit“: A Python package for discrete information theory,” *Journal of Open Source Software*, vol. 3, no. 25, p. 738, 2018.
- [80] L. Faes, G. Mijatovic, L. Sparacino, and A. Porta, “Predictive information decomposition as a tool to quantify emergent dynamical behaviors in physiological networks,” *IEEE Transactions on Biomedical Engineering*, 2025.
- [81] A. Liardi, F. E. Rosas, R. L. Carhart-Harris, G. Blackburne, D. Bor, and P. A. Mediano, “Null models for comparing information decomposition across complex systems,” *PLoS Computational Biology*, vol. 21, no. 11, p. e1013629, 2025.
- [82] P. K. Banerjee, E. Olbrich, J. Jost, and J. Rauh, “Unique informations and deficiencies,” in *2018 56th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 32–38, IEEE, 2018.
- [83] J. Walpole, J. A. Papin, and S. M. Peirce, “Multiscale computational models of complex biological systems,” *Annual review of biomedical engineering*, vol. 15, no. 1, pp. 137–154, 2013.
- [84] J. O. Dada and P. Mendes, “Multi-scale modelling and simulation in systems biology,” *Integrative Biology*, vol. 3, no. 2, pp. 86–96, 2011.
- [85] A. Liardi, M. I. Sas, G. Blackburne, W. J. Knottenbelt, P. A. Mediano, and H. J. Jensen, “Simple physical systems as a reference for multivariate information dynamics,” *arXiv preprint arXiv:2504.10372*, 2025.
- [86] N. Ay, D. Polani, and N. Virgo, “Information decomposition based on cooperative game theory,” *Kybernetika*, vol. 56, no. 5, pp. 979–1014, 2021.
- [87] F. E. Rosas, P. A. Mediano, B. Rassouli, and A. B. Barrett, “An operational information decomposition via synergistic disclosure,” *Journal of Physics A: Mathematical and Theoretical*, vol. 53, no. 48, p. 485001, 2020.
- [88] T. F. Varley, “A scalable synergy-first backbone decomposition of higher-order structures in complex systems,” *npj Complexity*, vol. 1, no. 1, p. 9, 2024.
- [89] D. Chicharro and S. Panzeri, “Synergy and redundancy in dual decompositions of mutual information gain and information loss,” *Entropy*, vol. 19, no. 2, p. 71, 2017.
- [90] G. Pica, E. Piasini, D. Chicharro, and S. Panzeri, “Invariant components of synergy, redundancy, and unique information among three variables,” *Entropy*, vol. 19, no. 9, p. 451, 2017.
- [91] P. Perrone and N. Ay, “Hierarchical quantification of synergy in channels,” *Frontiers in Robotics and AI*, vol. 2, p. 35, 2016.
- [92] A. Lyu, A. Clark, and N. Raviv, “The whole is less than the sum of parts: Subsystem inconsistency in partial information decomposition,” *arXiv preprint arXiv:2510.14864*, 2025.

- [93] P. A. Mediano, F. E. Rosas, A. I. Luppi, R. L. Carhart-Harris, D. Bor, A. K. Seth, and A. B. Barrett, “Toward a unified taxonomy of information dynamics via integrated information decomposition,” *Proceedings of the National Academy of Sciences*, vol. 122, no. 39, p. e2423297122, 2025.
- [94] R. A. Ince, “The partial entropy decomposition: Decomposing multivariate entropy and mutual information via pointwise common surprisal,” *arXiv preprint arXiv:1702.01591*, 2017.
- [95] T. F. Varley, “Generalized decomposition of multivariate information,” *Plos one*, vol. 19, no. 2, p. e0297128, 2024.
- [96] R. Quax, O. Har-Shemesh, and P. M. Sloom, “Quantifying synergistic information using intermediate stochastic variables,” *Entropy*, vol. 19, no. 2, p. 85, 2017.
- [97] E. Olbrich, N. Bertschinger, and J. Rauh, “Information decomposition and synergy,” *Entropy*, vol. 17, no. 5, pp. 3501–3517, 2015.
- [98] F. E. Rosas, P. A. Mediano, M. Gastpar, and H. J. Jensen, “Quantifying high-order interdependencies via multivariate extensions of the mutual information,” *Physical Review E*, vol. 100, no. 3, p. 032305, 2019.
- [99] G. Chechik, A. Globerson, M. Anderson, E. Young, I. Nelken, and N. Tishby, “Group redundancy measures reveal redundancy reduction in the auditory pathway,” *Advances in neural information processing systems*, vol. 14, 2001.
- [100] A. J. Bell, “The co-information lattice,” in *Proceedings of the fifth international workshop on independent component analysis and blind signal separation: ICA*, vol. 2003, 2003.
- [101] K. Reing, G. Ver Steeg, and A. Galstyan, “Discovering higher-order interactions through neural information decomposition,” *Entropy*, vol. 23, no. 1, p. 79, 2021.
- [102] E. Schneidman, S. Still, M. J. Berry, and W. Bialek, “Network information and connected correlations,” *Physical review letters*, vol. 91, no. 23, p. 238701, 2003.
- [103] E. Schneidman, W. Bialek, and M. J. Berry, “Synergy, redundancy, and independence in population codes,” *Journal of Neuroscience*, vol. 23, no. 37, pp. 11539–11553, 2003.
- [104] E. Schneidman, M. J. Berry, R. Segev, and W. Bialek, “Weak pairwise correlations imply strongly correlated network states in a neural population,” *Nature*, vol. 440, no. 7087, pp. 1007–1012, 2006.
- [105] F. Rosas, V. Ntranos, C. J. Ellison, S. Pollin, and M. Verhelst, “Understanding interdependency through complex information sharing,” *Entropy*, vol. 18, no. 2, p. 38, 2016.
- [106] A. J. Gutknecht, F. E. Rosas, D. A. Ehrlich, A. Makkeh, P. A. M. Mediano, and M. Wibral, “Shannon invariants: A scalable approach to information decomposition,” *arXiv preprint arXiv:2504.15779*, 2025.
- [107] K. J. A. Down and P. A. M. Mediano, “A logarithmic decomposition and a signed measure space for entropy,” *IEEE Transactions on Information Theory*, vol. 72, no. 1, pp. 17–41, 2026.
- [108] K. J. Down and P. A. Mediano, “Algebraic representations of entropy and fixed-sign information quantities,” *Entropy*, vol. 27, no. 2, p. 151, 2025.
- [109] H. Li and E. K. Chong, “On a connection between information and group lattices,” *Entropy*, vol. 13, no. 3, pp. 683–708, 2011.
- [110] C. Shannon, “The lattice theory of information,” *Transactions of the IRE professional Group on Information Theory*, vol. 1, no. 1, pp. 105–107, 1953.
- [111] P. Gács, J. Körner, *et al.*, “Common information is far less than mutual information.,” *Problems of Control and Information Theory*, vol. 2, pp. 149–162, 1973.

- [112] R. J. Aumann, “Agreeing to disagree,” *Readings in Formal Epistemology*, pp. 859–862, 2016.
- [113] E. Torgersen, *Comparison of statistical experiments*, vol. 36. Cambridge University Press, 1991.
- [114] J. Rauh, P. K. Banerjee, E. Olbrich, J. Jost, N. Bertschinger, and D. Wolpert, “Coarse-graining and the blackwell order,” *Entropy*, vol. 19, no. 10, p. 527, 2017.
- [115] W. McGill, “Multivariate information transmission,” *Transactions of the IRE Professional Group on Information Theory*, vol. 4, no. 4, pp. 93–111, 1954.
- [116] R. W. Yeung, “A new outlook on shannon’s information measures,” *IEEE transactions on information theory*, vol. 37, no. 3, pp. 466–474, 1991.
- [117] R. P. Dilworth, “A decomposition theorem for partially ordered sets,” in *The Dilworth Theorems: Selected Papers of Robert P. Dilworth*, pp. 7–12, Springer, 1990.
- [118] A. B. Barrett, L. Barnett, and A. K. Seth, “Multivariate granger causality and generalized variance,” *Physical Review E*, vol. 81, no. 4, p. 041907, 2010.
- [119] N. Bertschinger and J. Rauh, “The blackwell relation defines no lattice,” in *2014 IEEE International Symposium on Information Theory*, pp. 2479–2483, IEEE, 2014.
- [120] S. Wolf and J. Wulschleger, “Zero-error information and applications in cryptography,” in *Information Theory Workshop*, pp. 1–6, IEEE, 2004.
- [121] D. Eppstein, “Zonohedra and zonotopes,” 1995.
- [122] J. Pearl, *Causality*. Cambridge University Press, 2009.
- [123] J. Rauh, M. Schünemann, and J. Jost, “Properties of unique information,” *Kybernetika*, vol. 57, no. 3, pp. 383–403, 2021.
- [124] P. L. Williams, *Information dynamics: Its theory and application to embodied cognitive systems*. PhD thesis, Indiana University, 2011.