

Article

Orders between Channels and Implications for Partial Information Decomposition

André F. C. Gomes *  and Mário A. T. Figueiredo 

Instituto de Telecomunicações and LUMIS (Lisbon ELLIS Unit), Instituto Superior Técnico, Universidade de Lisboa, 1049-001 Lisboa, Portugal; mario.figueiredo@tecnico.ulisboa.pt

* Correspondence: andrefcgomes@tecnico.ulisboa.pt

Abstract: The partial information decomposition (PID) framework is concerned with decomposing the information that a set of random variables has with respect to a target variable into three types of components: redundant, synergistic, and unique. Classical information theory alone does not provide a unique way to decompose information in this manner, and additional assumptions have to be made. Recently, Kolchinsky proposed a new general axiomatic approach to obtain measures of redundant information based on choosing an order relation between information sources (equivalently, order between communication channels). In this paper, we exploit this approach to introduce three new measures of redundant information (and the resulting decompositions) based on well-known preorders between channels, contributing to the enrichment of the PID landscape. We relate the new decompositions to existing ones, study several of their properties, and provide examples illustrating their novelty. As a side result, we prove that any preorder that satisfies Kolchinsky's axioms yields a decomposition that meets the axioms originally introduced by Williams and Beer when they first proposed PID.

Keywords: information theory; partial information decomposition; channel preorders; intersection information; shared information; redundancy



Citation: Gomes, A.F.C.; Figueiredo, M.A.T. Orders between Channels and Implications for Partial Information Decomposition. *Entropy* **2023**, *25*, 975. <https://doi.org/10.3390/e25070975>

Academic Editors: Maria Longobardi and Francesco Buono

Received: 5 May 2023

Revised: 21 June 2023

Accepted: 22 June 2023

Published: 25 June 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Williams and Beer [1] proposed the partial information decomposition (PID) framework as a way to characterize or analyze the information that a set of random variables (often called *sources*) has about another variable (referred to as the *target*). PID is a useful tool for gathering insights and analyzing the way information is stored, modified, and transmitted within complex systems [2,3]. It has found applications in areas such as cryptography [4] and neuroscience [5,6], with many other potential use cases, such as in understanding how information flows function in gene regulatory networks [7], neural coding [8], financial markets [9], and network design [10].

Consider the simplest case, that of a three-variable joint distribution $p(y_1, y_2, t)$ describing three random variables: two sources Y_1 and Y_2 and a target T . Notice that despite the names *sources* and *target*, there is no directionality assumption, either causal or otherwise. The goal of PID is to *decompose* the information that $Y = (Y_1, Y_2)$ has about T into the sum of four non-negative quantities: the information that is present in both Y_1 and Y_2 , known as *redundant* information R ; the information that only Y_1 (respectively, Y_2) has about T , known as *unique* information U_1 (respectively, U_2); and the *synergistic* information S that is present in the pair (Y_1, Y_2) , and is not present in either Y_1 or Y_2 alone. That is, in this case with two variables, the goal is to write

$$I(T; Y) = R + U_1 + U_2 + S, \quad (1)$$

where $I(T; Y)$ is the mutual information between T and Y [11]. In this paper, mutual information is always assumed to refer to Shannon's mutual information, which for two discrete variables $X \in \mathcal{X}$ and $Z \in \mathcal{Z}$ is provided by

$$I(X; Z) = \sum_{x \in \mathcal{X}} \sum_{z \in \mathcal{Z}} p(x, z) \log \frac{p(x, z)}{p(x)p(z)},$$

and satisfies the following well-known fundamental properties: $I(X; Z) \geq 0$ and $I(X; Z) = 0 \Leftrightarrow X \perp Z$ (X and Z are independent) [11].

Because unique information and redundancy satisfy the relationship $U_i = I(T; Y_i) - R$ (for $i \in \{1, 2\}$), it turns out that defining how to compute one of these quantities (R , U_i , or S) is enough to fully determine the others [1]. As the number of variables grows, the number of terms appearing in the PID of $I(T; Y)$ grows super-exponentially [12]. Williams and Beer [1] suggested a set of axioms that a measure of redundancy should satisfy and proposed a measure of their own. These axioms have become known as the Williams–Beer axioms, although the measure they proposed has subsequently been criticized for not capturing informational content, only information size [13].

Spawned by their initial work, other measures and axioms for information decomposition have been introduced; see, for example, the work by Bertschinger et al. [14], Griffith and Koch [15], and James et al. [16]. There is no consensus about what axioms any measure should satisfy or whether a given measure is *capturing the information* that it should capture other than the Williams–Beer axioms. Today, debate continues about which axioms a measure of redundant information ought to satisfy, and there is no general agreement on what makes for an appropriate PID [16–20].

Recently, Kolchinsky [21] suggested a new general approach to defining measures of redundant information, known as *intersection information* (II), the designation that we adopt hereinafter. The core of this approach is the choice of an order relation between information sources (random variables), which allows two sources to be compared in terms of how informative they are with respect to the target variable.

In this work, we use previously studied preorders between communication channels, which correspond to preorders between the corresponding output variables in terms of information content with respect to the input. Following Kolchinsky's approach, we show that these orders lead to the definition of new II measures. The rest of the paper is organized as follows. In Sections 2 and 3, we review Kolchinsky's definition of an II measure and the *degradation* order. In Section 4, we describe a number of preorders between channels then, based on the work by Korner and Marton [22] and Américo et al. [23], we derive the resulting II measures and study of their properties. Section 5 presents comments on the optimization problems involved in computation of the proposed measures. In Section 6, we explore the relationships between the new II measures and previous PID approaches, then apply the proposed II measures to several famous PID problems. Section 7 concludes the paper by pointing out suggestions for future work.

2. Kolchinsky's Axioms and Intersection Information

Consider a set of n discrete random variables $Y_1 \in \mathcal{Y}_1, \dots, Y_n \in \mathcal{Y}_n$, called the source variables, and let $T \in \mathcal{T}$ be the target variable (also discrete), with a joint distribution (probability mass function) $p(y_1, \dots, y_n, t)$. Let \preceq denote some preorder between random variables that satisfies the following axioms, herein referred to as *Kolchinsky's axioms* [21]:

- (i) Monotonicity of *mutual information* w.r.t. T : $Y_i \preceq Y_j \Rightarrow I(Y_i; T) \leq I(Y_j; T)$.
- (ii) Reflexivity: $Y_i \preceq Y_i$ for all Y_i .
- (iii) For any $Y_i, C \preceq Y_i \preceq (Y_1, \dots, Y_n)$, where $C \in \mathcal{C}$ is any variable taking a constant value with probability one, i.e., with a distribution that is a delta function or such that \mathcal{C} is a singleton.

Kolchinsky [21] showed that such an order can be used to define an II measure via

$$I_{\cap}(Y_1, \dots, Y_n \rightarrow T) := \sup_{Q: Q \preceq Y_i, i \in \{1, \dots, n\}} I(Q; T), \tag{2}$$

and we now show that this implies that the II measure in (2) satisfies the Williams–Beer axioms [1,2], establishing a strong connection between these formulations. Before stating and proving this result, we first recall the Williams–Beer axioms [2], where the definition of a source A_i is that of a set of random variables, e.g., $A_1 = \{X_1, X_2\}$.

Definition 1. Let A_1, \dots, A_r be an arbitrary number of $r \geq 2$ sources. An intersection information measure I_{\cap} is said to satisfy the Williams–Beer axioms if it satisfies the following:

1. *Symmetry:* I_{\cap} is symmetric in the A_i s.
2. *Self-redundancy:* $I_{\cap}(A_i) = I(A_i; T)$.
3. *Monotonicity:* $I_{\cap}(A_1, \dots, A_{r-1}, A_r) \leq I_{\cap}(A_1, \dots, A_{r-1})$.
4. *Equality for Monotonicity:* If $A_{r-1} \subseteq A_r$, then $I_{\cap}(A_1, \dots, A_{r-1}, A_r) = I_{\cap}(A_1, \dots, A_{r-1})$.

Theorem 1. Let \preceq be some preorder that satisfies Kolchinsky’s axioms, and define its corresponding II measure as in (2). Then, the corresponding II measure satisfies the Williams–Beer axioms.

Proof. Symmetry and monotonicity follow trivially given the form of (2) (the definition of the supremum and restriction set). Self-redundancy follows from the reflexivity of the preorder and monotonicity of mutual information. Now, suppose $A_{r-1} \subseteq A_r$, and let Q be a solution of $I_{\cap}(A_1, \dots, A_{r-1})$, implying that $Q \preceq A_{r-1}$. Now, because $A_{r-1} \subseteq A_r$, the third Kolchinsky axiom and transitivity of the preorder \preceq guarantee that $Q \preceq A_{r-1} \preceq A_r$, meaning that Q is an admissible point of $I_{\cap}(A_1, \dots, A_r)$. Therefore, $I_{\cap}(A_1, \dots, A_{r-1}, A_r) \geq I_{\cap}(A_1, \dots, A_{r-1})$ and monotonicity guarantees that $I_{\cap}(A_1, \dots, A_{r-1}, A_r) = I_{\cap}(A_1, \dots, A_{r-1})$. \square

In conclusion, every preorder relation that satisfies the set of axioms introduced by Kolchinsky [21] yields a valid II measure, in the sense that the measure satisfies the Williams–Beer axioms. Having a *more informative* relation \preceq allows us to draw conclusions about information flowing from different sources, and allows for the construction of PID measures that are well-defined for more than two sources. In the following, we omit “ $\rightarrow T$ ” from the notation unless we need to explicitly refer to it, with the understanding that the target variable is always some arbitrary discrete random variable T .

3. Channels and the Degradation/Blackwell Order

In an information-theoretical perspective, given two discrete random variables $X \in \mathcal{X}$ and $Z \in \mathcal{Z}$, the corresponding conditional distribution $p(z|x)$ corresponds to a discrete memoryless channel with a channel matrix K such that $K[x, z] = p(z|x)$ [11]. This matrix is row-stochastic, i.e., $K[x, z] \geq 0$ for any $x \in \mathcal{X}$ and $z \in \mathcal{Z}$, and $\sum_{z \in \mathcal{Z}} K[x, z] = 1$.

The comparison of different channels (equivalently, different stochastic matrices) is an object of study with many applications in different fields [24]. Such investigations address order relations between channels and their properties. One such order, named the *degradation order* (or *Blackwell order*) and defined next, was used by Kolchinsky to obtain a particular II measure [21].

Consider the distribution $p(y_1, \dots, y_n, t)$ and the channels $K^{(i)}$ between T and each Y_i , that is, $K^{(i)}$ is a $|\mathcal{T}| \times |\mathcal{Y}_i|$ row-stochastic matrix with the conditional distribution $p(y_i|t)$.

Definition 2. We say that channel $K^{(i)}$ is a degradation of channel $K^{(j)}$, and write $K^{(i)} \preceq_d K^{(j)}$ or $Y_i \preceq_d Y_j$, if there exists a channel K^U from Y_j to Y_i , i.e., a $|\mathcal{Y}_j| \times |\mathcal{Y}_i|$ row-stochastic matrix, such that $K^{(i)} = K^{(j)}K^U$.

Intuitively, consider two agents, one with access to Y_i and the other with access to Y_j . The agent with access to Y_j has at least as much information about T as the one with access

to Y_i , as it has access to channel K^U , which permits sampling from Y_i conditionally on Y_j [19]. Blackwell [25] showed that this is equivalent to saying that, for whatever decision game where the goal is to predict T and for whatever utility function, the agent with access to Y_i cannot do better on average than the agent with access to Y_j .

Based on the degradation/Blackwell order, Kolchinsky [21] introduced the *degradation II measure* by plugging the “ \preceq_d ” order into (2):

$$I_{\cap}^d(Y_1, \dots, Y_n) := \sup_{Q: Q \preceq_d Y_i, i \in \{1, \dots, n\}} I(Q; T). \tag{3}$$

As noted by Kolchinsky [21], this II measure has the following operational interpretation. Supposing that $n = 2$ and considering two agents, 1 and 2, with access to variables Y_1 and Y_2 , respectively, $I_{\cap}^d(Y_1, Y_2)$ is the maximum information that agent 1 (respectively 2) can have with respect to T without being able to do better than agent 2 (respectively 1) on any decision problem that involves guessing T . That is, the degradation II measure quantifies the existence of a dominating strategy for any guessing game.

4. Other Orders and Corresponding II Measures

4.1. The “Less Noisy” Order

Korner and Marton [22] introduced and studied preorders between channels with the same input. We follow most of their definitions, and change others when appropriate. We interchangeably write $Y_1 \preceq Y_2$ to mean $K^{(1)} \preceq K^{(2)}$, where $K^{(1)}$ and $K^{(2)}$ are the channel matrices defined above.

Before introducing the next channel order, we need to review the notion of Markov chains [11]. We can say that three random variables X_1, X_2 , and X_3 form a Markov chain, for which we write $X_1 \rightarrow X_2 \rightarrow X_3$, if the following equality holds: $p(x_1, x_3|x_2) = p(x_1|x_2) p(x_3|x_2)$, i.e., if X_1 and X_3 are conditionally independent given X_2 . Of course, $X_1 \rightarrow X_2 \rightarrow X_3$ if and only if $X_3 \rightarrow X_2 \rightarrow X_1$.

Definition 3. We say that channel $K^{(2)}$ is less noisy than channel $K^{(1)}$, and write $K^{(1)} \preceq_{ln} K^{(2)}$, if for any discrete random variable U with finite support (such that both $U \rightarrow T \rightarrow Y_1$ and $U \rightarrow T \rightarrow Y_2$ hold) we have $I(U; Y_1) \leq I(U; Y_2)$.

The *less noisy* order has been primarily used in network information theory to study the problems of the capacity regions of broadcast channels [26] and the secrecy capacity of wiretap and eavesdrop channels [27]. The secrecy capacity (C_S) is the maximum rate at which information can be transmitted over a communication channel while keeping the communication secure from eavesdroppers, that is, having zero information leakage [28,29]. It has been shown that $C_S > 0$ unless $K^{(2)} \preceq_{ln} K^{(1)}$, where C_S is the secrecy capacity of the Wyner wiretap channel, with $K^{(2)}$ as the main channel and $K^{(1)}$ as the eavesdropper channel ([27], Corollary 17.11).

Plugging the *less noisy* order \preceq_{ln} into (2) yields a new II measure

$$I_{\cap}^{ln}(Y_1, \dots, Y_n) := \sup_{Q: Q \preceq_{ln} Y_i, i \in \{1, \dots, n\}} I(Q; T). \tag{4}$$

Intuitively, $I_{\cap}^{ln}(Y_1, \dots, Y_n)$ is the most information that a channel K^Q can have about T such that it is *less noisy* than any other channel $K^{(i)}$, $i = 1, \dots, n$, that is, a channel that leads to a positive secrecy capacity, as compared to any other channel $K^{(i)}$.

4.2. The “More Capable” Order

The next order we consider, termed “*more capable*”, has been used in calculating the capacity region of broadcast channels [30] and to help determine whether one system is more secure than another [31]; see the book by Cohen et al. [24] for more applications of the *degradation*, *less noisy*, and *more capable* orders.

Definition 4. We say that channel $K^{(2)}$ is more capable than $K^{(1)}$, and write $K^{(1)} \preceq_{mc} K^{(2)}$, if for any distribution $p(t)$ we have $I(T; Y_1) \leq I(T; Y_2)$.

Inserting the “more capable” order into (2) leads to

$$I_{\cap}^{mc}(Y_1, \dots, Y_n) := \sup_{Q: Q \preceq_{mc} Y_i, i \in \{1, \dots, n\}} I(Q; T), \tag{5}$$

that is, $I_{\cap}^{mc}(Y_1, \dots, Y_n)$ is the information that the ‘largest’ (in the more capable sense), though no larger than any Y_i , that the random variable Q has with respect to T . Whereas under the degradation order it is guaranteed that agent 2 will make better decisions if $Y_1 \preceq_d Y_2$ for whatever decision game, on average, under the “more capable” order, such a guarantee is not available. However, we do have a guarantee that, if $Y_1 \preceq_{mc} Y_2$, then for a given distribution $p(t)$ we know that agent 2 always has more information about T than agent 1. This has an interventional approach meaning; if we intervene on variable T by changing its distribution $p(t)$ in whichever way we see fit, we have $I(Y_1; T) \leq I(Y_2; T)$ (assuming that the distribution $p(Y_1, \dots, Y_n, T)$ can be modeled as a set of channels from T to each Y_i); that is to say, $I_{\cap}^{mc}(Y_1, \dots, Y_n)$ is the highest information that a channel K^Q can have about T such that for any change in $p(t)$, K^Q knows less about T than any $Y_i, i = 1, \dots, n$. Because PID is concerned with decomposing a distribution that has fixed $p(t)$, the “more capable” measure is concerned with the mechanism by which T generates Y_1, \dots, Y_n for any $p(t)$, and is not concerned with the specific distribution $p(t)$ yielded by $p(Y_1, \dots, Y_n, T)$.

For the sake of completeness, we could additionally study the Π measure that would result from the capacity order. Recall that the capacity of the channel from a variable X to another variable Z , which is only a function of the conditional distribution $p(z|x)$, is defined as [11]

$$C = \max_{p(x)} I(X; Z). \tag{6}$$

Definition 5. We can write $W \preceq_c V$ if the capacity of V is at least as large as the capacity of W .

Even though it is clear that $W \preceq_{mc} V \Rightarrow W \preceq_c V$, the \preceq_c order does not comply with the first of Kolchinsky’s axioms, as the definition of capacity involves the choice of a particular marginal that achieves the maximum in (6), which may not coincide with the marginal corresponding to $p(y_1, \dots, y_n, t)$. For this reason, we do not define an Π measure based on it.

4.3. The “Degradation/Supermodularity” Order

In order to introduce the last Π measure, we follow the work and notation of Américo et al. [23]. Given two real vectors r and s with dimension n , let $r \vee s := (\max(r_1, s_1), \dots, \max(r_n, s_n))$ and $r \wedge s := (\min(r_1, s_1), \dots, \min(r_n, s_n))$. Consider an arbitrary channel K , and let K_i be its i th column. From K , we may define a new channel which we construct column by column using the *JoinMeet* operator $\diamond_{i,j}$. Column l of the new channel is defined for $i \neq j$ as

$$(\diamond_{i,j}K)_l = \begin{cases} K_i \vee K_j, & \text{if } l = i \\ K_i \wedge K_j, & \text{if } l = j \\ K_l, & \text{otherwise} \end{cases} .$$

Américo et al. [23] used this operator to define the two new orders described below. Intuitively, the operator $\diamond_{i,j}$ makes the rows of the channel matrix more similar to each other by putting all the maxima in column i and the minima in column j between every pair of elements in columns i and j of every row. In the following definitions, the s stands for supermodularity, a concept we need not introduce in this work.

Definition 6. We can write $W \preceq_s V$ if there exists a finite collection of tuples (i_k, j_k) such that $W = \diamond_{i_1, j_1}(\diamond_{i_2, j_2}(\dots(\diamond_{i_m, j_m} V))$.

Definition 7. We write $W \preceq_{ds} V$ if there are m channels $U^{(1)}, \dots, U^{(m)}$ such that $W \preceq_0 U^{(1)} \preceq_1 U^{(2)} \preceq_2 \dots \preceq_{m-1} U^{(m)} \preceq_m V$, where each \preceq_i stands for \preceq_d or \preceq_s . We call this the degradation/supermodularity order.

Using the “degradation/supermodularity” (ds) order, we can define the ds I measure as follows:

$$I_{\cap}^{ds}(Y_1, \dots, Y_n) := \sup_{Q: Q \preceq_{ds} Y_i, i \in \{1, \dots, n\}} I(Q; T). \tag{7}$$

The *ds* order was recently introduced in the context of *core-concave entropies* [23]. Given a core-concave entropy H , the *leakage* about T through Y_1 is defined as $I_H(T; Y_1) = H(T) - H(T|Y_1)$. In this work, we are mainly concerned with the Shannon entropy H ; however, as we elaborate in the future work section at the end of this paper, PID may be applied to other core-concave entropies. Although the operational interpretation of the *ds* order is not yet clear, it has found applications in privacy/security contexts, as well as in finding the most secure deterministic channel (under certain constraints) [23].

4.4. Relations between Orders

Korner and Marton [22] proved that $W \preceq_d V \Rightarrow W \preceq_{ln} V \Rightarrow W \preceq_{mc} V$, and provided examples to show that the reverse implications do not hold in general. As Américo et al. [23] note, the degradation (\preceq_d), supermodularity (\preceq_s), and degradation/supermodularity (\preceq_{ds}) orders are *structural orders*, in the sense that they only depend on the conditional probabilities that are defined by each channel. On the other hand, the *less noisy* and *more capable* orders are concerned with information measures resulting from different distributions. It is trivial to see (directly from the definition) that the degradation order implies the degradation/supermodular order. In turn, Américo et al. [23] showed that the degradation/supermodular order implies the more capable order. This set of implications is schematically depicted in Figure 1.

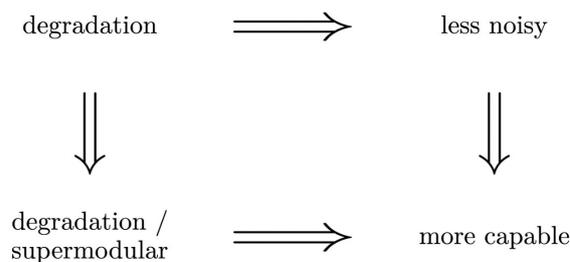


Figure 1. Implications satisfied by the orders. The reverse implications do not hold in general.

For any set of variables Y_1, \dots, Y_n, T , these relations between the orders imply, via the corresponding definitions, that

$$I_{\cap}^d(Y_1, \dots, Y_n) \leq I_{\cap}^{ln}(Y_1, \dots, Y_n) \leq I_{\cap}^{mc}(Y_1, \dots, Y_n) \tag{8}$$

and

$$I_{\cap}^d(Y_1, \dots, Y_n) \leq I_{\cap}^{ds}(Y_1, \dots, Y_n) \leq I_{\cap}^{mc}(Y_1, \dots, Y_n). \tag{9}$$

These, in turn, imply the following result.

Theorem 2. The preorders \preceq_{ln} , \preceq_{mc} , and \preceq_{ds} , satisfy Kolchinsky’s axioms.

Proof. Let $i \in \{1, \dots, n\}$. Because any of the introduced orders implies the *more capable* order, it follows that they all satisfy the axiom of monotonicity of mutual information. Axiom 2 is trivially true, as reflexivity is guaranteed by the definition of preorder. For axiom 3, the rows of a channel corresponding to a variable C taking a constant value must all be the same (and yield zero mutual information with any target variable T), from which it is clear that any Y_i satisfies $C \preceq Y_i$ for any of the introduced orders per the definition of each order. To see that $Y_i \preceq Y = (Y_1, \dots, Y_n)$ for the *less noisy* and the *more capable*, recall that for any U such that $U \rightarrow T \rightarrow Y_i$ and $U \rightarrow T \rightarrow Y$ it is trivial that $I(U; Y_i) \leq I(U; Y)$; hence, $Y_i \preceq_{ln} Y$. A similar argument can be used to show that $Y_i \preceq_{mc} Y$, as $I(T; Y_i) \leq I(T; Y)$. Finally, to see that $Y_i \preceq_{ds} (Y_1, \dots, Y_n)$, note that $Y_i \preceq_d (Y_1, \dots, Y_n)$ [21]; hence, $Y_i \preceq_{ds} (Y_1, \dots, Y_n)$. \square

5. Optimization Problems

We now focus on certain observations around optimization problems involving the introduced Π measures. All of these problems seek to maximize $I(Q; T)$ (under different constraints) as a function of the conditional distribution $p(q|t)$, and equivalently with respect to the channel from T to Q , which we denote as $K^Q := K^{Q|T}$. For fixed $p(t)$, as is the case in PID, $I(Q; T)$ is a convex function of K^Q ([11], Theorem 2.7.4). As we will see, the admissible region of all problems is a compact set, and because $I(Q; T)$ is a continuous function of the parameters of K^Q , the supremum is achieved; thus, we replace sup here with max.

As noted by Kolchinsky [21], the computation of (3) can be rewritten as an optimization problem using auxiliary variables such that it involves only linear constraints, and because the objective function is convex, its maximum is attained at one of the vertices of the admissible region. The computation of the other measures, however, is not as simple, as shown in the following subsections.

5.1. The “Less Noisy” Order

To solve (4), we can use one of the necessary and sufficient conditions presented by (Makur and Polyanskiy [26], Theorem 1). For instance, let V and W be two channels with input T , and let Δ^{T-1} be the probability simplex of the target T . Then, $V \preceq_{ln} W$, if and only if the inequality

$$\chi^2(p(t)W||q(t)W) \geq \chi^2(p(t)V||q(t)V) \tag{10}$$

holds for any pair of distributions $p(t), q(t) \in \Delta^{T-1}$, where χ^2 in the above equation denotes the χ^2 -distance between two vectors. The χ^2 distance between two vectors u and v of dimension n is given by $\chi^2(u||v) = \sum_{i=1}^n (u_i - v_i)^2 / v_i$. Notice that $p(t)W$ is the distribution of the output of channel W for input distribution $p(t)$; thus, intuitively, the condition in (10) means that the two output distributions of the *less noisy* channel are more different from each other than those of the other channel. Hence, computing $I_{\cap}^{ln}(Y_1, \dots, Y_n)$ can be formulated as solving the problem

$$\begin{aligned} &\max_{K^Q} I(Q; T) \\ &\text{s.t. } K^Q \text{ is a stochastic matrix,} \\ &\quad \forall p(t), q(t) \in \Delta^{T-1}, \forall i \in \{1, \dots, n\}, \chi^2(p(t)K^{(i)}||q(t)K^{(i)}) \geq \chi^2(p(t)K^Q||q(t)K^Q). \end{aligned}$$

Although the restriction set is convex, as the χ^2 -divergence is an f -divergence with convex f [27], the problem is intractable because we have an infinite (uncountable) number of restrictions. It is possible to construct a set \mathcal{S} by taking an arbitrary number of samples S of $p(t) \in \Delta^{T-1}$ to define the problem

$$\begin{aligned} &\max_{K^Q} I(Q; T), \\ &\text{s.t. } K^Q \text{ is a stochastic matrix,} \\ &\quad \forall p(t), q(t) \in \mathcal{S}, \forall i \in \{1, \dots, n\}, \chi^2(p(t)K^{(i)}||q(t)K^{(i)}) \geq \chi^2(p(t)K^Q||q(t)K^Q). \end{aligned} \tag{11}$$

The above problem yields an upper bound on $I_{\cap}^{ln}(Y_1, \dots, Y_n)$.

5.2. The “More Capable” Order

To compute $I_{\cap}^{mc}(Y_1, \dots, Y_n)$, we can define the problem

$$\begin{aligned} \max_{K^Q} \quad & I(Q; T) \\ \text{s.t.} \quad & K^Q \text{ is a stochastic matrix,} \\ & \forall p(t) \in \Delta^{T-1}, \forall i \in \{1, \dots, n\}, I(Y_i; T) \geq I(Q; T), \end{aligned} \tag{12}$$

which again leads to a convex restriction set, as $I(Q; T)$ is a convex function of K^Q . We can discretize the problem in the same manner as above to obtain a tractable version

$$\begin{aligned} \max_{K^Q} \quad & I(Q; T) \\ \text{s.t.} \quad & K^Q \text{ is a stochastic matrix,} \\ & \forall p(t) \in \mathcal{S}, \forall i \in \{1, \dots, n\}, I(Y_i; T) \geq I(Q; T), \end{aligned} \tag{13}$$

which again yields an upper bound on $I_{\cap}^{mc}(Y_1, \dots, Y_n)$.

5.3. The “Degradation/Supermodularity” Order

The final introduced measure, $I_{\cap}^{ds}(Y_1, \dots, Y_n)$, is provided by

$$\begin{aligned} \max_{K^Q} \quad & I(Q; T) \\ \text{s.t.} \quad & K^Q \text{ is a stochastic matrix,} \\ & \forall i, K^Q \preceq_{ds} K^{(i)}. \end{aligned} \tag{14}$$

To the best of our knowledge, there is currently no known condition to check whether $K^Q \preceq_{ds} K^{(i)}$.

6. Relation to Existing PID Measures

Griffith et al. [32] introduced a measure of Π as

$$I_{\cap}^s(Y_1, \dots, Y_n) := \max_Q I(Q; T), \text{ such that } \forall i \ Q \triangleleft Y_i, \tag{15}$$

with the order relation \triangleleft defined by $A \triangleleft B$ if $A = f(B)$ for some deterministic function f , that is, I_{\cap}^s is used to quantify the redundancy as the presence of deterministic relations between the input and target. If Q is a solution of (15), then there exist functions f_1, \dots, f_n such that $Q = f_i(Y_i), i = 1, \dots, n$, which implies that for all i it is the case that $T \rightarrow Y_i \rightarrow Q$ is a Markov chain. Therefore, Q is an admissible point of the optimization problem that defines $I_{\cap}^d(Y_1, \dots, Y_n)$, and we have $I_{\cap}^s(Y_1, \dots, Y_n) \leq I_{\cap}^d(Y_1, \dots, Y_n)$.

Barrett [33] introduced the so-called *minimum mutual information* (MMI) measure of bivariate redundancy as

$$I_{\cap}^{MMI}(Y_1, Y_2) := \min\{I(T; Y_1), I(T; Y_2)\}.$$

It turns out that if (Y_1, Y_2) is jointly Gaussian and T is univariate, then most of the introduced PIDs in the literature are equivalent to this measure [33]. Furthermore, as noted by Kolchinsky [21], it may be generalized to more than two sources:

$$I_{\cap}^{MMI}(Y_1, \dots, Y_n) := \sup_Q I(Q; T), \text{ such that } \forall i \ I(Q; T) \leq I(Y_i; T),$$

which allows us to trivially conclude that for any set of variables Y_1, \dots, Y_n, T ,

$$I_{\cap}^s(Y_1, \dots, Y_n) \leq I_{\cap}^d(Y_1, \dots, Y_n) \leq I_{\cap}^{mc}(Y_1, \dots, Y_n) \leq I_{\cap}^{\text{MMI}}(Y_1, \dots, Y_n).$$

One of the appeals of measures of II as defined by Kolchinsky [21] is that the underlying preorder determines what is intersection (or redundant) information. For example, taking the degradation II measure in the $n = 2$ case, its solution Q satisfies $T \perp Q | Y_1$ and $T \perp Q | Y_2$; that is, if either Y_1 or Y_2 are known, then Q has no additional information about T . The same is not necessarily the case for the *less noisy* or the *more capable* II measures, where the solution Q may have additional information about T even when a source is known. However, the three proposed measures satisfy the property that any solution Q of the optimization problem satisfies

$$\forall i \in \{1, \dots, n\}, \forall t \in S_T, I(Y_i; T = t) \geq I(Q; T = t),$$

where S_T is the support of T and $I(T = t; Y_i)$ refers to the so-called specific information [1,34]. This means that, independent of the outcome of T , Q has less specific information about $T = t$ than any source variable Y_i . This can be seen by noting that any of the introduced orders imply the more capable order. This is not the case, for example, for I_{\cap}^{MMI} , which is arguably one of the reasons why it has been criticized for depending only on the amount of information and not on its content [21]. As mentioned, there is not much consensus as to which properties a measure of II should satisfy. The three proposed measures for partial information decomposition do not satisfy the so-called *Blackwell property* [14,35]:

Definition 8. An intersection information measure $I_{\cap}(Y_1, Y_2)$ is said to satisfy the Blackwell property if the equivalence $Y_1 \preceq_d Y_2 \Leftrightarrow I_{\cap}(Y_1, Y_2) = I(T; Y_1)$ holds.

This definition is equivalent to demanding that $Y_1 \preceq_d Y_2$ if and only if Y_1 has no unique information about T . Although the (\Rightarrow) implication holds for the three proposed measures, the reverse implication does not, as shown by specific examples presented by Korner and Marton [22], which we mention below. If we define the “more capable” property by replacing the *degradation* order with the *more capable* order in the original definition of the Blackwell property, then it is clear that measure k satisfies the k property, with k referring to any of the three introduced intersection information measures.

In PID, the *identity property* (IP) has been frequently studied [13]. For this property, let the target T be a copy of the source variables, that is, let $T = (Y_1, Y_2)$. An II measure I_{\cap} is said to satisfy the IP if

$$I_{\cap}(Y_1, Y_2) = I(Y_1; Y_2).$$

Criticism has been levied against this proposal for being too restrictive [16,36]. A less strict property was introduced by [20] under the name *independent identity property* (IIP). If the target T is a copy of the input, an II measure is said to satisfy the IIP if

$$I(Y_1; Y_2) = 0 \Rightarrow I_{\cap}(Y_1, Y_2) = 0.$$

Note that the IIP is implied by the IP, while the reverse does not hold. It turns out that all the introduced measures, as is the case for the degradation II measure, satisfy the IIP and not the IP, as we show later. This can be seen from (8) and (9), as well as from the fact that $I_{\cap}^{mc}(Y_1, Y_2 \rightarrow (Y_1, Y_2))$ equals 0 if $I(Y_1; Y_2) = 0$, as we argue now. Consider the distribution where T is a copy of (Y_1, Y_2) , as presented in Table 1.

Table 1. Copy distribution.

T	Y ₁	Y ₂	p(t, y ₁ , y ₂)
(0, 0)	0	0	p(T = (0, 0))
(0, 1)	0	1	p(T = (0, 1))
(1, 0)	1	0	p(T = (1, 0))
(1, 1)	1	1	p(T = (1, 1))

We assume that each of the four events has non-zero probability. In this case, channels $K^{(1)}$ and $K^{(2)}$ are provided by

$$K^{(1)} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}, \quad K^{(2)} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Note that for any distribution $p(t) = [p(0, 0), p(0, 1), p(1, 0), p(1, 1)]$, if $p(1, 0) = p(1, 1) = 0$, then $I(T; Y_1) = 0$, which implies that for any such distributions the solution Q of (12) must satisfy $I(Q; T) = 0$. Thus, the first and second rows of K^Q must be the same. The same is the case for any distribution $p(t)$ with $p(0, 0) = p(0, 1) = 0$; on the other hand, if $p(0, 0) = p(1, 0) = 0$ or $p(1, 1) = p(0, 1) = 0$, then $I(T; Y_2) = 0$, implying that $I(Q; T) = 0$ for such distributions. Hence, K^Q must be an arbitrary channel, that is, a channel that satisfies $Q \perp T$, yielding $I_{\cap}^{mc}(Y_1, Y_2) = 0$.

Now, recall the Gács–Korner *common information* [37], defined as

$$C(Y_1 \wedge Y_2) := \sup_Q H(Q) \tag{16}$$

s.t. $Q \triangleleft Y_1$
 $Q \triangleleft Y_2$

We use a similar argument, while slightly changing the notation, to show the following result.

Theorem 3. *Let $T = (X, Y)$ be a copy of the source variables; then, $I_{\cap}^{ln}(X, Y) = I_{\cap}^{ds}(X, Y) = I_{\cap}^{mc}(X, Y) = C(X \wedge Y)$.*

Proof. As shown by Kolchinsky [21], $I_{\cap}^d(X, Y) = C(X \wedge Y)$. Thus, (8) implies that $I_{\cap}^{mc}(X, Y) \geq C(X \wedge Y)$. The proof is completed by showing that $I_{\cap}^{mc}(X, Y) \leq C(X \wedge Y)$. Construct the bipartite graph with vertex set $\mathcal{X} \cup \mathcal{Y}$ and edges (x, y) if $p(x, y) > 0$. Consider the set of *maximally connected components* $MCC = \{CC_1, \dots, CC_l\}$ for some $l \geq 1$, where each CC_i refers to a maximal set of connected edges. Let $CC_i, i \leq l$ be an arbitrary set in MCC . Suppose that the edges (x_1, y_1) and (x_1, y_2) (with $y_1 \neq y_2$) are in CC_i . This means that the channels $K^X := K^{X|T}$ and $K^Y := K^{Y|T}$ have rows corresponding to the outcomes $T = (x_1, y_1)$ and $T = (x_1, y_2)$ of the form

$$K^X = \begin{bmatrix} \vdots \\ 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ \vdots \end{bmatrix}, \quad K^Y = \begin{bmatrix} \vdots \\ 0 & \dots & 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots \end{bmatrix}.$$

Choosing $p(t) = [0, \dots, 0, a, 1 - a, 0, \dots, 0]$, that is, $p(T = (x_1, y_1)) = a$ and $p(T = (x_1, y_2)) = 1 - a$, we have $\forall a \in [0, 1], I(X; T) = 0$, which implies that the solution Q must be such that $\forall a \in [0, 1], I(Q; T) = 0$ (from the definition of the *more capable* order),

which in turn implies that the rows of K^Q corresponding to these outcomes must be the same to ensure that they yield $I(Q; T) = 0$ under this set of distributions. We may choose the values of those rows to be the same as those rows from K^X , that is, a row that is composed of zeros except for one of the positions whenever $T = (x_1, y_1)$ or $T = (x_1, y_2)$. On the other hand, if the edges (x_1, y_1) and (x_2, y_1) (with $x_1 \neq x_2$) are in CC_i , the same argument leads to the conclusion that the rows of K^Q corresponding to the outcomes $T = (x_1, y_1)$, $T = (x_1, y_2)$ and $T = (x_2, y_1)$ must be the same. Applying this argument to every edge in CC_i , we can conclude that the rows of K^Q corresponding to outcomes $(x, y) \in CC_i$ must all be the same. Using this argument for every set CC_1, \dots, CC_i implies that if two edges are in the same CC , the corresponding rows of K^Q must be the same. These corresponding rows of K^Q may vary between different CC s; however, for the same CC they must be the same.

We are left with the choice of appropriate rows of K^Q for each corresponding CC_i . Because $I(Q; T)$ is maximized by a deterministic relation between Q and T and, as suggested before, we choose a row that is composed of zeros except for one of the positions for each CC_i such that Q is a deterministic function of T , this admissible point Q implies that $Q = f_1(X)$ and $Q = f_2(Y)$, as X and Y are also functions of T under the channel perspective. For this choice of rows, we have

$$I_{\cap}^{mc}(X, Y) = \sup_Q I(Q; T) \leq \sup_Q H(Q) = C(X \wedge Y)$$

$$\text{s.t. } Q \preceq_{mc} X \quad \text{s.t. } Q = f_1(X)$$

$$Q \preceq_{mc} Y \quad \quad \quad Q = f_2(Y)$$

where we have used the fact that $I(Q; T) \leq \min\{H(Q), H(T)\}$ to conclude that $I_{\cap}^{mc}(X, Y) \leq C(X \wedge Y)$. Hence $I_{\cap}^{ln}(X, Y) = I_{\cap}^{ds}(X, Y) = I_{\cap}^{mc}(X, Y) = C(X \wedge Y)$ if T is a copy of the input. \square

Bertschinger et al. [14] suggested what later became known as the (*) assumption, which states that in the bivariate source case any sensible measure of unique information should only depend on $K^{(1)}, K^{(2)}$, and $p(t)$. It is not clear that this assumption should hold for every PID. It is trivial to see that all the introduced Π measures satisfy the (*) assumption.

We conclude with several applications of the proposed measures to famous (bivariate) PID problems; the results are shown in Table 2. Due to the channel design in these problems, computation of the proposed measures is fairly trivial. We assume that the input variables are binary (taking values in $\{0, 1\}$), independent, and equiprobable.

Table 2. Application of the proposed measures to famous PID problems.

Target	I_{\cap}^{\leq}	I_{\cap}^d	I_{\cap}^{ln}	I_{\cap}^{ds}	I_{\cap}^{mc}	I_{\cap}^{MMI}
$T = Y_1 \text{ AND } Y_2$	0	0.311	0.311	0.311	0.311	0.311
$T = Y_1 + Y_2$	0	0.5	0.5	0.5	0.5	0.5
$T = Y_1$	0	0	0	0	0	0
$T = (Y_1, Y_2)$	0	0	0	0	0	1

We note that in these fairly simple toy distributions all of the introduced measures yield the same value. This is not surprising when the distribution $p(t, y_1, y_2)$ yields $K^{(1)} = K^{(2)}$, which implies that $I(T; Y_1) = I(T; Y_2) = I_{\cap}^k(Y_1, Y_2)$, where k refers to any of the introduced preorders, as is the case in the $T = Y_1 \text{ AND } Y_2$ and $T = Y_1 + Y_2$ examples. Less trivial examples lead to different values over the introduced measures. We present distributions showing that our three introduced measures lead to novel information decompositions by comparing them to the following existing measures: I_{\cap}^{\leq} from Griffith et al. [32], I_{\cap}^{MMI} from Barrett [33], I_{\cap}^{WB} from Williams and Beer [1], I_{\cap}^{GH} from Griffith and Ho [38], I_{\cap}^{Ince} from Ince [20], I_{\cap}^{FL} from Finn and Lizier [39], I_{\cap}^{BROJA} from Bertschinger et al. [14], I_{\cap}^{Harder} from Harder et al. [13], and I_{\cap}^{dep} from [16]. We used the dit package [40] to compute

them, along with the code provided in [21]. Consider counterexample 1 from [22] with $p = 0.25, \epsilon = 0.2, \delta = 0.1$, provided by

$$K^{(1)} = \begin{bmatrix} 0.25 & 0.75 \\ 0.35 & 0.65 \end{bmatrix}, \quad K^{(2)} = \begin{bmatrix} 0.675 & 0.325 \\ 0.745 & 0.255 \end{bmatrix}.$$

These channels satisfy $K^{(2)} \preceq_{ln} K^{(1)}$ and $K^{(2)} \not\preceq_d K^{(1)}$ from Korner and Marton [22]. This is an example that satisfies $I_{\cap}^{ln}(Y_1, Y_2) = I(T; Y_2)$ for a given distribution $p(t)$. It is noteworthy to see that even though there is no degradation order between the two channels, we nonetheless have $I_{\cap}^d(Y_1, Y_2) > 0$, as there is some non-trivial channel K^Q that satisfies $K^Q \preceq_d K^{(1)}$ and $K^Q \preceq_d K^{(2)}$. In Table 3, we present various PIDs under different measures after choosing $p(t) = [0.4, 0.6]$ (which yields $I(T; Y_2) \approx 0.004$) and assuming $p(t, y_1, y_2) = p(t)p(y_1|t)p(y_2|t)$.

Table 3. Different decompositions of $p(t, y_1, y_2)$.

I_{\cap}^s	I_{\cap}^d	I_{\cap}^{ln}	I_{\cap}^{ds}	I_{\cap}^{mc}	I_{\cap}^{MMI}	I_{\cap}^{WB}	I_{\cap}^{GH}	I_{\cap}^{Ince}	I_{\cap}^{FL}	I_{\cap}^{BROJA}	I_{\cap}^{Harder}	I_{\cap}^{dep}
0	0.002	0.004	*	0.004	0.004	0.004	0.002	0.003	0.047	0.003	0.004	0

We write $I_{\cap}^{ds} = *$ here, as we do not yet have a way to find the ‘largest’ Q such that $Q \preceq_{ds} K^{(1)}$ and $Q \preceq_{ds} K^{(2)}$ (see counterexample 2 from [22] for an example of channels $K^{(1)}, K^{(2)}$ that satisfy $K^{(2)} \preceq_{mc} K^{(1)}$ while $K^{(2)} \not\preceq_{ln} K^{(1)}$, leading to different values of the proposed Π measures). An example of $K^{(3)}, K^{(4)}$ that satisfy $K^{(4)} \preceq_{ds} K^{(3)}$ while $K^{(4)} \not\preceq_d K^{(3)}$ is presented by (Américo et al. [23], page 10), provided by

$$K^{(3)} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0.5 & 0.5 \end{bmatrix}, \quad K^{(4)} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0.5 & 0.5 \end{bmatrix}.$$

There is no stochastic matrix K^U such that $K^{(4)} = K^{(3)}K^U$ while $K^{(4)} \preceq_{ds} K^{(3)}$, as $K^{(4)} = \diamond_{1,2}K^{(3)}$. Using (10), it is possible to check whether there is any *less noisy* relation between the two channels. (Compute (10) with $V = K^{(4)}, W = K^{(3)}, p(t) = [0, 0, 1]$, and $q(t) = [0.1, 0.1, 0.8]$ to conclude that $K^{(4)} \not\preceq_{ln} K^{(3)}$, then switch the roles of V and W and set $p(t) = [0, 1, 0]$ and $q(t) = [0.1, 0, 0.9]$ to conclude that $K^{(3)} \not\preceq_{ln} K^{(4)}$). We present the decomposition of $p(t, y_3, y_4) = p(t)p(y_3|t)p(y_4|t)$ for the choice of $p(t) = [0.3, 0.3, 0.4]$ (which yields $I(T; Y_4) \approx 0.322$) in Table 4.

Table 4. Different decompositions of $p(t, y_3, y_4)$.

I_{\cap}^s	I_{\cap}^d	I_{\cap}^{ln}	I_{\cap}^{ds}	I_{\cap}^{mc}	I_{\cap}^{MMI}	I_{\cap}^{WB}	I_{\cap}^{GH}	I_{\cap}^{Ince}	I_{\cap}^{FL}	I_{\cap}^{BROJA}	I_{\cap}^{Harder}	I_{\cap}^{dep}
0	0	0*	0.322	0.322	0.322	0.193	0	0	0.058	0	0	0

We write $I_{\cap}^{ln} = 0*$ because we conjecture, after numerical experiments based on (10), that the ‘largest’ channel that is less noisy than both $K^{(3)}$ and $K^{(4)}$ is a channel that satisfies $I(Q; T) = 0$. (We tested all 3×3 row-stochastic matrices with entries that take values in $\{0, 0.1, 0.2, \dots, 0.9, 1\}$ with all distributions $p(t)$ and $q(t)$ having entries that take values in the same set.)

7. Conclusions and Future Work

In this paper, we have introduced three new measures of *intersection information* for the *partial information decomposition* (PID) framework based on preorders between channels implied by the degradation/Blackwell order. The new measures were obtained from the orders by following the approach recently proposed by Kolchinsky [21]. The main contributions and conclusions of this paper can be summarized as follows:

- We show that a measure of intersection information that satisfies the axioms by Kolchinsky [21] and is based on a preorder, satisfies the Williams–Beer axioms as well [1].
- As a corollary of the previous result, the proposed measures satisfy the Williams–Beer axioms, and can be extended beyond two sources.
- We demonstrate that if there is a degradation ordering between the sources, then the measures coincide in their decomposition. Conversely, if there is no degradation ordering (i.e., only a weaker ordering) between the source variables, the proposed measures lead to novel finer information decompositions that capture different finer information.
- We show that while the proposed measures do not satisfy the *identity property* (IP) [13], they do satisfy the *independent identity property* (IIP) [20].
- We formulate the optimization problems that yield the proposed measures, and derive bounds by relating them to existing measures.

Finally, we believe that this paper opens several avenues for future research; thus, we point to several directions that could be pursued in upcoming work:

- Investigating conditions to verify whether two channels $K^{(1)}$ and $K^{(2)}$ satisfy $K^{(1)} \preceq_{ds} K^{(2)}$.
- Kolchinsky [21] showed that when computing $I_{\cap}^d(Y_1, \dots, Y_n)$, it is sufficient to consider variables Q with a support size of at most $\sum_i |S_{Y_i}| - n + 1$, which is a consequence of the admissible region of $I_{\cap}^d(Y_1, \dots, Y_n)$ being a polytope. The same is not the case with the *less noisy* or the *more capable* measures; hence, it is not clear whether it is sufficient to consider Q with the same support size, which could represent a direction for future research.
- Studying the conditions under which different *intersection information* measures are continuous.
- Implementing the introduced measures by addressing their corresponding optimization problems.
- Considering the usual PID framework, except that instead of decomposing $I(T; Y) = H(Y) - H(Y|T)$, where H denotes the Shannon entropy, other mutual informations induced by different entropy measures could be considered, such as the *guessing entropy* [41] or the *Tsallis entropy* [42] (see the work of Américo et al. [23] for other core-concave entropies that may be decomposed under the introduced preorders, as these entropies are consistent with the introduced orders).
- Another line for future work might be to define measures of union information using the introduced preorders, as suggested by Kolchinsky [21], and to study their properties.
- As a more long-term research direction, it would be interesting to study how the approach taken in this paper can be extended to quantum information; the fact that partial quantum information can be negative might open up new possibilities or create novel difficulties [43].

Author Contributions: Conceptualization, A.F.C.G.; writing—original draft preparation, A.F.C.G.; writing—review and editing, A.F.C.G. and M.A.T.F.; supervision, M.A.T.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially funded by: FCT—*Fundação para a Ciência e a Tecnologia* under the grants SFRH/BD/145472/2019 and UIDB/50008/2020; Instituto de Telecomunicações; Portuguese Recovery and Resilience Plan, project C645008882-00000055 (NextGenAI, CenterforResponsibleAI).

Institutional Review Board Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We thank Artemy Kolchinsky for helpful discussions and suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Williams, P.; Beer, R. Nonnegative decomposition of multivariate information. *arXiv* **2010**, arXiv:1004.2515.
2. Lizier, J.; Flecker, B.; Williams, P. Towards a synergy-based approach to measuring information modification. In Proceedings of the 2013 IEEE Symposium on Artificial Life (ALIFE), Singapore, 16–19 April 2013; pp. 43–51.
3. Wibral, M.; Finn, C.; Wollstadt, P.; Lizier, J.; Priesemann, V. Quantifying information modification in developing neural networks via partial information decomposition. *Entropy* **2017**, *19*, 494. [[CrossRef](#)]
4. Rauh, J. Secret sharing and shared information. *Entropy* **2017**, *19*, 601. [[CrossRef](#)]
5. Vicente, R.; Wibral, M.; Lindner, M.; Pipa, G. Transfer entropy—A model-free measure of effective connectivity for the neurosciences. *J. Comput. Neurosci.* **2011**, *30*, 45–67. [[CrossRef](#)]
6. Ince, R.; Van Rijsbergen, N.; Thut, G.; Rousselet, G.; Gross, J.; Panzeri, S.; Schyns, P. Tracing the flow of perceptual features in an algorithmic brain network. *Sci. Rep.* **2015**, *5*, 17681. [[CrossRef](#)] [[PubMed](#)]
7. Gates, A.; Rocha, L. Control of complex networks requires both structure and dynamics. *Sci. Rep.* **2016**, *6*, 24456. [[CrossRef](#)]
8. Faber, S.; Timme, N.; Beggs, J.; Newman, E. Computation is concentrated in rich clubs of local cortical networks. *Netw. Neurosci.* **2019**, *3*, 384–404. [[CrossRef](#)]
9. James, R.; Ayala, B.; Zakirov, B.; Crutchfield, J. Modes of information flow. *arXiv* **2018**, arXiv:1808.06723.
10. Arellano-Valle, R.; Contreras-Reyes, J.; Genton, M. Shannon Entropy and Mutual Information for Multivariate Skew-Elliptical Distributions. *Scand. J. Stat.* **2013**, *40*, 42–62. [[CrossRef](#)]
11. Cover, T. *Elements of Information Theory*; John Wiley & Sons: Hoboken, NJ, USA, 1999.
12. Gutknecht, A.; Wibral, M.; Makkeh, A. Bits and pieces: Understanding information decomposition from part-whole relationships and formal logic. *Proc. R. Soc. A* **2021**, *477*, 20210110. [[CrossRef](#)]
13. Harder, M.; Salge, C.; Polani, D. Bivariate measure of redundant information. *Phys. Rev. E* **2013**, *87*, 012130. [[CrossRef](#)] [[PubMed](#)]
14. Bertschinger, N.; Rauh, J.; Olbrich, E.; Jost, J.; Ay, N. Quantifying unique information. *Entropy* **2014**, *16*, 2161–2183. [[CrossRef](#)]
15. Griffith, V.; Koch, C. Quantifying synergistic mutual information. In *Guided Self-Organization: Inception*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 159–190.
16. James, R.; Emenheiser, J.; Crutchfield, J. Unique information via dependency constraints. *J. Phys. A Math. Theor.* **2018**, *52*, 014002. [[CrossRef](#)]
17. Chicharro, D.; Panzeri, S. Synergy and redundancy in dual decompositions of mutual information gain and information loss. *Entropy* **2017**, *19*, 71. [[CrossRef](#)]
18. Bertschinger, N.; Rauh, J.; Olbrich, E.; Jost, J. Shared information—New insights and problems in decomposing information in complex systems. In Proceedings of the European Conference on Complex Systems 2012, Brussels, Belgium, 2–7 September 2012; Springer: Berlin/Heidelberg, Germany, 2013; pp. 251–269.
19. Rauh, J.; Banerjee, P.; Olbrich, E.; Jost, J.; Bertschinger, N.; Wolpert, D. Coarse-graining and the Blackwell order. *Entropy* **2017**, *19*, 527. [[CrossRef](#)]
20. Ince, R. Measuring multivariate redundant information with pointwise common change in surprisal. *Entropy* **2017**, *19*, 318. [[CrossRef](#)]
21. Kolchinsky, A. A Novel Approach to the Partial Information Decomposition. *Entropy* **2022**, *24*, 403. [[CrossRef](#)]
22. Korner, J.; Marton, K. Comparison of two noisy channels. In *Topics in Information Theory*; Csiszr, I., Elias, P., Eds.; North-Holland Pub. Co.: Amsterdam, The Netherlands, 1977; pp. 411–423.
23. Américo, A.; Khouzani, A.; Malacaria, P. Channel-Supermodular Entropies: Order Theory and an Application to Query Anonymization. *Entropy* **2021**, *24*, 39. [[CrossRef](#)]
24. Cohen, J.; Kempermann, J.; Zbaganu, G. *Comparisons of Stochastic Matrices with Applications in Information Theory, Statistics, Economics and Population*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 1998.
25. Blackwell, D. Equivalent comparisons of experiments. *Ann. Math. Stat.* **1953**, *24*, 265–272. [[CrossRef](#)]
26. Makur, A.; Polyanskiy, Y. Less noisy domination by symmetric channels. In Proceedings of the 2017 IEEE International Symposium on Information Theory (ISIT), Aachen, Germany, 25–30 June 2017; pp. 2463–2467.
27. Csiszár, I.; Körner, J. *Information Theory: Coding Theorems for Discrete Memoryless Systems*; Cambridge University Press: Cambridge, UK, 2011.
28. Wyner, A. The wire-tap channel. *Bell Syst. Tech. J.* **1975**, *54*, 1355–1387. [[CrossRef](#)]
29. Bassi, G.; Piantanida, P.; Shamai, S. The secret key capacity of a class of noisy channels with correlated sources. *Entropy* **2019**, *21*, 732. [[CrossRef](#)]
30. Gamal, A. The capacity of a class of broadcast channels. *IEEE Trans. Inf. Theory* **1979**, *25*, 166–169. [[CrossRef](#)]
31. Clark, D.; Hunt, S.; Malacaria, P. Quantitative information flow, relations and polymorphic types. *J. Log. Comput.* **2005**, *15*, 181–199. [[CrossRef](#)]
32. Griffith, V.; Chong, E.; James, R.; Ellison, C.; Crutchfield, J. Intersection information based on common randomness. *Entropy* **2014**, *16*, 1985–2000. [[CrossRef](#)]
33. Barrett, A. Exploration of synergistic and redundant information sharing in static and dynamical Gaussian systems. *Phys. Rev. E* **2015**, *91*, 052802. [[CrossRef](#)]
34. DeWeese, M.; Meister, M. How to measure the information gained from one symbol. *Netw. Comput. Neural Syst.* **1999**, *10*, 325. [[CrossRef](#)]
35. Rauh, J.; Banerjee, P.; Olbrich, E.; Jost, J.; Bertschinger, N. On extractable shared information. *Entropy* **2017**, *19*, 328. [[CrossRef](#)]

36. Rauh, J.; Bertschinger, N.; Olbrich, E.; Jost, J. Reconsidering unique information: Towards a multivariate information decomposition. In Proceedings of the 2014 IEEE International Symposium on Information Theory, Honolulu, HI, USA, 29 June–4 July 2014; pp. 2232–2236.
37. Gács, P.; Körner, J. Common information is far less than mutual information. *Probl. Control Inf. Theory* **1973**, *2*, 149–162.
38. Griffith, V.; Ho, T. Quantifying redundant information in predicting a target random variable. *Entropy* **2015**, *17*, 4644–4653. [[CrossRef](#)]
39. Finn, C.; Lizier, J. Pointwise Partial Information Decomposition Using the Specificity and Ambiguity Lattices. *Entropy* **2018**, *20*, 297. [[CrossRef](#)]
40. James, R.; Ellison, C.; Crutchfield, J. “dit”: A Python package for discrete information theory. *J. Open Source Softw.* **2018**, *3*, 738. [[CrossRef](#)]
41. Massey, J. Guessing and entropy. In Proceedings of the 1994 IEEE International Symposium on Information Theory, Trondheim, Norway, 27 June–1 July 1994; p. 204.
42. Tsallis, C. Possible generalization of Boltzmann-Gibbs statistics. *J. Stat. Phys.* **1988**, *52*, 479–487. [[CrossRef](#)]
43. Horodecki, M.; Oppenheim, J.; Winter, A. Partial quantum information. *Nature* **2005**, *436*, 673–676. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.