Intersection Information based on Common Randomness

Virgil Griffith[1], Edwin K. P. Chong[2], Ryan G. James[3], Christopher J. Ellison[4], and James P. Crutchfield[5,6]

[1]Computation and Neural Systems, Caltech, Pasadena, CA 91125

[2]Dept. of Electrical & Computer Engineering
Colorado State University, Fort Collins, CO 80523

[3]Computer Science Department, University of Colorado, Boulder, CO 80309

[4]Center for Complexity and Collective Computation
University of Wisconsin-Madison, Madison WI 53706

[5]Complexity Sciences Center and Physics Dept.
University of California Davis, Davis, CA 95616

[6]Santa Fe Institute, Santa Fe, NM 87501

Abstract

Since the seminal paper [1], there's been a flurry of research [2–5] towards defining an *intersection information* that quantifies how much of "the same information" two or more random variables specify about a target random variable. A palatable measure of intersection information would provide a principled way to quantify slippery concepts such as synergy. Here we introduce an intersection information measure based on the Gács-Körner common random variable which is the first to satisfy the coveted Target Monotonicity property. Our measure is imperfect and we suggest directions for improvement.

1 Introduction

Introduced in [1], *Partial Information Decomposition* (PID) is an immensely useful framework for deepening our understanding of multivariate interactions—particularly our understanding of informational redundancy and synergy. To harness the PID framework, the user brings her own measure of *intersection information*, $\mathrm{I}_\cap(X_1, \ldots, X_n : Y)$, which quantifies the magnitude of information that each of the n predictors $X_1, \ldots, X_n$ conveys about a target random variable Y. An antichain lattice of redundant, unique, and synergistic partial informations is built from the intersection information.

In [1], the authors propose to use the following quantity, $\mathrm{I}_{\min}$, as the intersection information measure:

$$\begin{aligned} \mathrm{I}_{\min}\left(X_1, \ldots, X_n : Y\right) &\equiv \sum_y \Pr(y) \min_i \mathrm{I}(X_i : Y = y) \\ &= \sum_y \Pr(y) \min_i \mathrm{D}_{\mathrm{KL}}\left[\Pr\big(X_i|y\big) \Big\| \Pr\big(X_i\big)\right] , \end{aligned} \tag{1}$$

where D_{KL} is the Kullback-Leibler divergence.

Though $\mathrm{I}_{\min}$ is an intuitive and plausible choice for the intersection information, [2] showed that $\mathrm{I}_{\min}$ has counterintuitive properties. In particular, $\mathrm{I}_{\min}$ calculates one bit of redundant information for example UNQ (Figure 1). It does this because each input shares one bit of information with the output. However, its quite clear that the shared informations are, in fact, different: X_1 provides the low bit, while X_2 provides the high bit. This led to the conclusion that $\mathrm{I}_{\min}$ *over-estimates* the ideal intersection information measure by focusing only on *how much* information the inputs provide to the output. An ideal measure of intersection information must recognize that there are non-equivalent ways of providing information to the output. The search for an improved intersection information measure ensued, continued through [3–5], and today a widely accepted intersection information measure remains undiscovered.

Here we do not definitively solve this problem, but we present a strong candidate intersection information measure for the special case of *zero-error* information. This is useful in of itself because it provides a template for how the yet undiscovered ideal intersection information measure for Shannon mutual information could work. Alternatively, if a Shannon intersection information measure with the same properties does not exist, then we have learned something significant.

In the next section, we introduce some definitions, some notation, and a necessary lemma. We also extend and clarify the desired properties for intersection information. In Section 3 we introduce zero-error information and its intersection information measure. In Section 4 we use the same methodology to produce a novel candidate for the Shannon intersection information. In Section 5 we show the successes and shortcomings of our candidate intersection information measure using example circuits. Finally in Section 7, we summarize our progress towards the ideal intersection information measure and suggest directions for improvement. The Appendix is devoted to technical lemmas and their proofs, to which we refer in the main text.

2 Preliminaries

2.1 Informational Partial Order and Equivalence

We assume an underlying probability space on which we define random variables denoted by capital letters (e.g., X, Y, and Z). In this paper, we consider only random variables taking values on finite spaces.

Given random variables X and Y, we write $X \preceq Y$ to signify that there exists a measurable function f such that $X = f(Y)$ almost surely (i.e., with probability one). In this case, following the terminology in [6], we say that X is *informationally poorer* than Y; this induces a partial order on the set of random variables. Similarly, we write $X \succeq Y$ if $Y \preceq X$, in which case we say X is *informationally richer* than Y.

If X and Y are such that $X \preceq Y$ and $X \succeq Y$, then we write $X \cong Y$. In this case, again following [6], we say that X and Y are *informationally equivalent*. In other words, $X \cong Y$ if and only if one can relabel the values of X to obtain a random value that is equal to Y almost surely, and vice versa.

This "information-equivalence" relation can easily be shown to be an equivalence relation, so that we can partition the set of all random variables into disjoint equivalence classes. The $\preceq$ ordering is invariant within these equivalence classes in the following sense. If $X \preceq Y$ and $Y \cong Z$, then $X \preceq Z$. Similarly, if $X \preceq Y$ and $X \cong Z$, then $Z \preceq Y$. Moreover, within each equivalence class, the entropy is invariant, as stated formally in Lemma 1 below.

2.2 Information Lattice

Next, we follow [6] and consider the *join* and *meet* operators. These operators were defined for *information elements*, which are σ-algebras, or, equivalently, equivalence classes of random variables. We deviate from [6], though, by defining the join and meet operators for random variables, but we do preserve their conceptual properties.

Given random variables X and Y, we define $X \curlyvee Y$ (called the *join* of X and Y) to be an informationally poorest ("smallest" in the sense of the partial order $\preceq$) random variable such that $X \preceq X \curlyvee Y$ and $Y \preceq X \curlyvee Y$. In other words, if Z is such that $X \preceq Z$ and $Y \preceq Z$, then $X \curlyvee Y \preceq Z$. Note that $X \curlyvee Y$ is unique only up to equivalence with respect to $\cong$. In other words, $X \curlyvee Y$ does not define a specific, unique random variable. Nonetheless, standard information-theoretic quantities are invariant over the set of random variables satisfying the condition specified above. For example, the entropy of $X \curlyvee Y$ is invariant over the entire equivalence class of random variables satisfying the condition above (by Lemma 1(a) below). Similarly, the inequality $Z \preceq X \curlyvee Y$ does not depend on the specific random variable chosen, as long as it satisfies the condition above. Note that the pair (X, Y) is an instance of $X \curlyvee Y$.

In a similar vein, given random variables X and Y, we define $X \curlywedge Y$ (called the *meet* of X and Y) to be an informationally richest random variable ("largest" in the sense of $\succeq$) such that $X \curlywedge Y \preceq X$ and $X \curlywedge Y \preceq Y$. In other words, if Z is such that $Z \preceq X$ and $Z \preceq Y$, then $Z \preceq X \curlywedge Y$. Following [7], we also call $X \curlywedge Y$ the *common random variable* of X and Y. Again, considering the entropy of $X \curlywedge Y$ or the inequality $Z \preceq X \curlywedge Y$ does not depend on the specific random variable chosen, as long as it satisfies the condition above.

The $\curlyvee$ and $\curlywedge$ operators satisfy the algebraic properties of a *lattice* [6]. In particular, the following hold:

- commutative laws: $X \curlyvee Y \cong Y \curlyvee X$ and $X \curlywedge Y \cong Y \curlywedge X$
- associative laws: $X \curlyvee (Y \curlyvee Z) \cong (X \curlyvee Y) \curlyvee Z$ and $X \curlywedge (Y \curlywedge Z) \cong (X \curlywedge Y) \curlywedge Z$
- absorption laws: $X \curlyvee (X \curlywedge Y) \cong X$ and $X \curlywedge (X \curlyvee Y) \cong X)$
- idempotent laws: $X \curlyvee X \cong X$ and $X \curlywedge X \cong X$
- generalized absorption laws: if $X \preceq Y$, then $X \curlyvee Y \cong Y$and $X \curlywedge Y \cong X$.

Finally, the partial order $\preceq$ is preserved under $\curlyvee$ and $\curlywedge$, i.e., if $X \preceq Y$, then $X \curlyvee Z \preceq Y \curlyvee Z$ and $X \curlywedge Z \preceq X \curlywedge Z$.

2.3 Invariance and Monotonicity of Entropy

Let $\mathrm{H}(\cdot)$ represent the entropy function, and $\mathrm{H}\big(\cdot|\cdot\big)$ the conditional entropy. To be consistent with the colon in the intersection information, we denote the Shannon mutual information between X and Y by $\mathrm{I}(X\!:\!Y)$ instead of the more common $\mathrm{I}(X;Y)$. Lemma 1 establishes the invariance and monotonicity of the entropy and conditional entropy functions with respect to $\cong$ and $\preceq$.

Lemma 1. *The following hold:*

(a) If $X \cong Y$*, then* $\mathrm{H}(X) = \mathrm{H}(Y)$*,* $\mathrm{H}(X|Z) = \mathrm{H}(Y|Z)$*, and* $\mathrm{H}(Z|X) = \mathrm{H}(Z|Y)$*.*

(b) If $X \preceq Y$*, then* $\mathrm{H}(X) \leq \mathrm{H}(Y)$*,* $\mathrm{H}(X|Z) \leq \mathrm{H}(Y|Z)$*, and* $\mathrm{H}(Z|X) \geq \mathrm{H}(Z|Y)$*.*

(c) $X \preceq Y$ *if and only if* $\mathrm{H}(X|Y) = 0$*.*

Proof. Part (a) follows from [6], Proposition 1. Part (c) follows from [6], Proposition 4. The first two desired inequalities in part (b) follow from [6], Proposition 5. Now we show that if $X \preceq Y$, then $\mathrm{H}\big(Z|X\big) \geq \mathrm{H}\big(Z|Y\big)$. Suppose that $X \preceq Y$. Then, by the generalized absorption law, $X \curlyvee Y \cong Y$. We have

$$\begin{aligned}
\mathrm{I}(Z\!:\!Y) &= \mathrm{H}(Y) - \mathrm{H}(Y|Z) \\
&= \mathrm{H}(X \curlyvee Y) - \mathrm{H}(X \curlyvee Y|Z) \quad \text{by part (a)} \\
&= \mathrm{I}(Z\!:\!X \curlyvee Y) \\
&= \mathrm{I}(Z\!:\!X) + \mathrm{I}(Z\!:\!Y|X) \\
&\geq \mathrm{I}(Z\!:\!X)\,.
\end{aligned}$$

Substituting $\mathrm{I}(Z\!:\!Y) = \mathrm{H}(Z) - \mathrm{H}(Z|Y)$ and $\mathrm{I}(Z\!:\!X) = \mathrm{H}(Z) - \mathrm{H}(Z|X)$, we obtain $\mathrm{H}(Z|X) \geq \mathrm{H}(Z|Y)$ as desired. □

Remark: Because $(X, Y) \cong X \curlyvee Y$ as noted before, we also have $\mathrm{H}(X, Y) = \mathrm{H}(X \curlyvee Y)$ by Lemma 1(a).

2.4 Desired Properties of Intersection Information

There are currently 12 intuitive properties that we wish the ideal intersection information measure $\mathrm{I}_\cap$ to satisfy. Some are new (e.g. $(\mathbf{M_1})$, $(\mathbf{Eq})$, $(\mathbf{LB})$), but most were introduced earlier, in various forms, Refs. [1–5]. They are as follows:

- $(\mathbf{GP})$ Global Positivity: $\mathrm{I}_\cap(X_1, \ldots, X_n : Y) \geq 0$, and $\mathrm{I}_\cap(X_1, \ldots, X_n : Y) = 0$ if Y is a constant.
- $(\mathbf{Eq})$ Equivalence-Class Invariance: $\mathrm{I}_\cap(X_1, \ldots, X_n : Y)$ is invariant under substitution of X_i (for any $i = 1, \ldots, n$) or Y by an informationally equivalent random variable.
- $(\mathbf{TM})$ Target Monotonicity: If $Y \preceq Z$, then $\mathrm{I}_\cap(X_1, \ldots, X_n : Y) \leq \mathrm{I}_\cap(X_1, \ldots, X_n : Z)$.
- $(\mathbf{M_0})$ Weak Monotonicity: $\mathrm{I}_\cap(X_1, \ldots, X_n, W : Y) \leq \mathrm{I}_\cap(X_1, \ldots, X_n : Y)$ with equality if there exists $Z \in \{X_1, \ldots, X_n\}$ such that $Z \preceq W$.
- $(\mathbf{S_0})$ Weak Symmetry: $\mathrm{I}_\cap(X_1, \ldots, X_n : Y)$ is invariant under reordering of $X_1, \ldots, X_n$.

Remark: If $(\mathbf{S_0})$ is satisfied, the first argument of $\mathrm{I}_\cap(X_1, \ldots, X_n : Y)$ can be treated as a *set* of random variables rather than a *list*. In this case, the notation $\mathrm{I}_\cap\big(\{X_1, \ldots, X_n\} : Y\big)$ would also be appropriate.

For the next set of properties, $\mathcal{I}(X : Y)$ is a given normative measure of information between X and Y. For example, $\mathcal{I}(X : Y)$ could denote the Shannon mutual information; i.e., $\mathcal{I}(X : Y) = \mathrm{I}(X : Y)$. Alternatively, as discussed in the next section, we might take $\mathcal{I}(X : Y)$ to be the zero-error information. Yet other possibilities for $\mathcal{I}(X : Y)$ include the Wyner common information [8] or the quantum mutual information [9]. The following are desired properties of intersection information *relative to* the given information measure $\mathcal{I}$.

- $(\mathbf{LB})$ Lowerbound: If $Q \preceq X_i$ for all $i = 1, \ldots, n$, then $\mathrm{I}_\cap(X_1, \ldots, X_n : Y) \geq \mathcal{I}(Q : Y)$. Under a mild assumption,[1] this equates to $\mathrm{I}_\cap(X_1, \ldots, X_n : Y) \geq \mathcal{I}(X_1 \curlywedge \cdots \curlywedge X_n : Y)$.
- $(\mathbf{SR})$ Self-Redundancy: $\mathrm{I}_\cap(X_1 : Y) = \mathcal{I}(X_1 : Y)$. The intersection information a single predictor X_1 conveys about the target Y is equal to the information between the predictor and the target given by the information measure $\mathcal{I}$.
- $(\mathbf{Id})$ Identity: $\mathrm{I}_\cap(X, Y : X \curlyvee Y) = \mathcal{I}(X : Y)$.
- $(\mathbf{LP_0})$ Weak Local Positivity: $\mathrm{I}_\cap(X_1, X_2 : Y) \geq \mathcal{I}(X_1 : Y) + \mathcal{I}(X_2 : Y) - \mathcal{I}(X_1 \curlyvee X_2 : Y)$. In other words, for $n = 2$ predictors, the derived "partial informations" defined in [1] are nonnegative when both $(\mathbf{LP_0})$ and $(\mathbf{GP})$ hold.

Finally, we have the less obvious "strong" properties.

- $(\mathbf{M_1})$ Strong Monotonicity: $\mathrm{I}_\cap(X_1, \ldots, X_n, W : Y) \leq \mathrm{I}_\cap(X_1, \ldots, X_n : Y)$ with equality if there exists $Z \in \{X_1, \ldots, X_n, Y\}$ such that $Z \preceq W$.
- $(\mathbf{S_1})$ Strong Symmetry: $\mathrm{I}_\cap(X_1, \ldots, X_n : Y)$ is invariant under reordering of $X_1, \ldots, X_n, Y$.
- $(\mathbf{LP_1})$ Strong Local Positivity: For all n, the derived "partial informations" defined in [1] are nonnegative.

Properties $(\mathbf{Eq})$, $(\mathbf{LB})$, and $(\mathbf{M_1})$ are novel and are introduced for the first time here. Given $\mathrm{I}_\cap$, $X_1, \ldots, X_n$, Y, and Z, we define the conditional $\mathrm{I}_\cap$ as:

$$\mathrm{I}_\cap(X_1, \ldots, X_n : Z | Y) \equiv \mathrm{I}_\cap(X_1, \ldots, X_n : Y \curlyvee Z) - \mathrm{I}_\cap(X_1, \ldots, X_n : Y) \ .$$

This definition of $\mathrm{I}_\cap(X_1, \ldots, X_n : Z | Y)$ gives rise to the familiar "chain rule":

$$\mathrm{I}_\cap(X_1, \ldots, X_n : Y \curlyvee Z) = \mathrm{I}_\cap(X_1, \ldots, X_n : Y) + \mathrm{I}_\cap(X_1, \ldots, X_n : Z | Y) \ .$$

Some provable[2] properties are:

[1] See Lemmas 2 and 3 in Appendix C.1.
[2] See Lemma 4 in Appendix C.1.

- $\mathrm{I}_\cap(X_1, \ldots, X_n : Z|Y) \geq 0$.
- $\mathrm{I}_\cap(X_1, \ldots, X_n : Z|Y) = \mathrm{I}_\cap(X_1, \ldots, X_n : Z)$ if Y is a constant.

3 Candidate Intersection Information for Zero-Error Information

3.1 Zero-Error Information

Introduced in [10], the *zero-error information*, or *Gács-Körner common information*, is a stricter variant of Shannon mutual information. Whereas the mutual information $\mathrm{I}(A : B)$ quantifies the magnitude of information A conveys about B with an arbitrarily small error $\epsilon > 0$, the zero-error information, denoted $\mathrm{I}^0(A : B)$, quantifies the magnitude of information A conveys about B with *exactly zero* error, i.e., $\epsilon = 0$. The zero-error information between A and B equals the entropy of the *common random variable* $A \curlywedge B$,

$$\mathrm{I}^0(A : B) \equiv \mathrm{H}(A \curlywedge B) \ .$$

An algorithm for computing an instance of the common random variable between two random variables is provided in [10], and straightforwardly generalizes to n random variables.[3]

Zero-error information has several notable properties, but the most salient is that it is nonnegative and bounded by the mutual information,

$$0 \leq \mathrm{I}^0(A : B) \leq \mathrm{I}(A : B) \,.$$

This generalizes to arbitrary n:

$$0 \leq \mathrm{I}^0(X_1 : \cdots : X_n) \leq \min_{i,j} \mathrm{I}\big(X_i : X_j\big) \,.$$

3.2 Intersection Information for Zero-Error Information

It is pleasingly straightforward to define a palatable intersection information for zero-error information (i.e., setting $\mathcal{I} = \mathrm{I}^0$ as the normative measure of information). We propose the zero-error intersection information, $\mathrm{I}^0_\curlywedge(X_1, \ldots, X_n : Y)$, as the maximum zero-error information $\mathrm{I}^0(Q : Y)$ that some random variable Q conveys about Y, subject to Q being a function of each predictor $X_1, \ldots, X_n$:

$$\begin{aligned} \mathrm{I}^0_\curlywedge(X_1, \ldots, X_n : Y) \equiv & \max_{\Pr(Q|Y)} \mathrm{I}^0(Q : Y) \\ & \text{subject to } Q \preceq X_i \ \forall i \in \{1, \ldots, n\} \ . \end{aligned} \tag{2}$$

Basic algebra[4] shows that a maximizing Q is the common random variable across all predictors. This substantially simplifies eq. (2) to:

$$\begin{aligned} \mathrm{I}^0_\curlywedge(X_1, \ldots, X_n : Y) &= \mathrm{I}^0(X_1 \curlywedge \cdots \curlywedge X_n : Y) \\ &= \mathrm{H}\big[(X_1 \curlywedge \cdots \curlywedge X_n) \curlywedge Y\big] \\ &= \mathrm{H}(X_1 \curlywedge \cdots \curlywedge X_n \curlywedge Y) \ . \end{aligned} \tag{3}$$

Importantly, the zero-error information, $\mathrm{I}^0_\curlywedge(X_1, \ldots, X_n : Y)$ satisfies 10 of the 12 desired properties from Section 2.4, leaving only ($\mathbf{LP_0}$) and ($\mathbf{LP_1}$) unsatisfied.[5]

4 Candidate Intersection Information for Shannon Information

In the last section, we defined an intersection information for zero-error information which satisfies the vast majority of desired properties. This is a solid start, but an intersection information for Shannon mutual information remains the goal. Towards this end, we use the

[3] See Appendix A.
[4] See Lemma 11 in Appendix D.
[5] See Lemmas 5, 6, 7 in Appendix C.2.

same method as in eq. (2), leading to $\mathrm{I}_\curlywedge$, our candidate intersection information measure for Shannon mutual information,

$$\begin{aligned}\mathrm{I}_\curlywedge(X_1,\ldots,X_n:Y) \equiv \max_{\Pr(Q|Y)} &\ \mathrm{I}(Q:Y) \\ &\text{subject to } Q \preceq X_i\ \forall i \in \{1,\ldots,n\} \ .\end{aligned} \tag{4}$$

With some algebra[6] this similarly simplifies to,

$$\mathrm{I}_\curlywedge(X_1,\ldots,X_n:Y) = \mathrm{I}(X_1 \curlywedge \cdots \curlywedge X_n : Y) \ . \tag{5}$$

Unfortunately $\mathrm{I}_\curlywedge$ does not satisfy as many of the desired properties as $\mathrm{I}^0_\curlywedge$. However, our candidate $\mathrm{I}_\curlywedge$ still satisfies 7 of the 12 properties—most importantly the enviable (**TM**),[7] which has, until now, not been satisfied by any proposed measure. Table 1 lists the desired properties satisfied by $\mathrm{I}_{\min}$, $\mathrm{I}_\curlywedge$, and $\mathrm{I}^0_\curlywedge$. For reference, we also include $\mathrm{I}_{\mathrm{red}}$, the proposed measure from [3].

Comparing the three subject intersection information measures,[8] we have:

$$0 \le \mathrm{I}^0_\curlywedge(X_1,\ldots,X_n:Y) \le \mathrm{I}_\curlywedge(X_1,\ldots,X_n:Y) \le \mathrm{I}_{\min}(X_1,\ldots,X_n:Y) \ . \tag{6}$$

	Property	$\mathrm{I}_{\min}$	$\mathrm{I}_{\mathrm{red}}$	$\mathrm{I}_\curlywedge$	$\mathrm{I}^0_\curlywedge$
(**GP**)	Global Positivity	✓	✓	✓	✓
(**Eq**)	Equivalence-Class Invariance	✓	✓	✓	✓
(**TM**)	Target Monotonicity			✓	✓
($\mathbf{M_0}$)	Weak Monotonicity	✓		✓	✓
($\mathbf{S_0}$)	Weak Symmetry	✓	✓	✓	✓
(**LB**)	Lowerbound	✓	✓	✓	✓
(**SR**)	Self-Redundancy	✓	✓	✓	✓
(**Id**)	Identity		✓		✓
($\mathbf{LP_0}$)	Weak Local Positivity	✓	✓		
($\mathbf{M_1}$)	Strong Monotonicity				✓
($\mathbf{S_1}$)	Strong Symmetry				✓
($\mathbf{LP_1}$)	Strong Local Positivity	✓			

Table 1: The $\mathrm{I}_\cap$ desired properties each measure satisfies.

Despite not satisfying ($\mathbf{LP_0}$), $\mathrm{I}_\curlywedge$ remains an important stepping-stone towards the ideal Shannon $\mathrm{I}_\cap$. First, $\mathrm{I}_\curlywedge$ captures what is inarguably redundant information (the common random variable)—this makes $\mathrm{I}_\curlywedge$ necessarily a lower bound on any reasonable redundancy measure. Second, it is the first proposal to satisfy target monotonicity and the associated chain rule. Lastly, $\mathrm{I}_\curlywedge$ is the first measure to reach intuitive answers in many canonical situations, while also being generalizable to an arbitrary number of inputs.

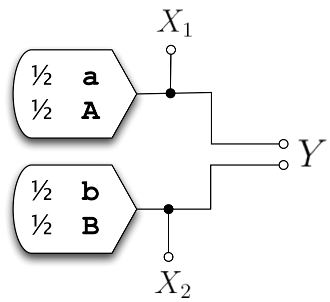
5 Three Examples Comparing $\mathrm{I}_{\min}$ and $\mathrm{I}_\curlywedge$

Examples Unq and RdnXor illustrate $\mathrm{I}_\curlywedge$'s successes and example ImperfectRdn illustrates $\mathrm{I}_\curlywedge$'s paramount deficiency. For each example we show the joint distribution $\Pr(x_1, x_2, y)$, a

[6] See Lemma 12 in Appendix D.
[7] See Lemmas 8, 9, 10 in Appendix C.3.
[8] See Lemma 13 in Appendix D.

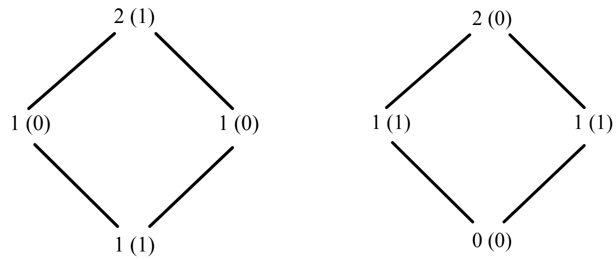
X_1 X_2	Y	
a b	ab	1/4
a B	aB	1/4
A b	Ab	1/4
A B	AB	1/4

(a) $\Pr(x_1, x_2, y)$



(b) circuit diagram

$$\begin{aligned} \mathrm{I}(X_1 \curlyvee X_2 : Y) &= 2 \\ \mathrm{I}(X_1 : Y) &= 1 \\ \mathrm{I}(X_2 : Y) &= 1 \\ \mathrm{I}_{\min}(X_1, X_2 : Y) &= 1 \\ \mathrm{I}_{\curlywedge}(X_1, X_2 : Y) &= 0 \end{aligned}$$



(c) $\mathrm{I}_{\min}$ (d) $\mathrm{I}_{\curlywedge}$ and $\mathrm{I}^0_{\curlywedge}$

Figure 1: Example Unq. This is the canonical example of unique information. X_1 and X_2 each uniquely specify a single bit of Y. This is the simplest example where $\mathrm{I}_{\min}$ calculates an undesirable decomposition (c) of one bit of redundancy and one bit of synergy. $\mathrm{I}_{\curlywedge}$ and $\mathrm{I}^0_{\curlywedge}$ each calculate the desired decomposition (d).

diagram, and the decomposition derived from setting $\mathrm{I}_{\min} / \mathrm{I}_{\curlywedge}$ as the $\mathrm{I}_{\cap}$ measure. At each lattice junction, the left number is the $\mathrm{I}_{\cap}$ value of that node, and the number in parentheses is the I_{∂} value.[9] Readers unfamiliar with the $n = 2$ partial information lattice should consult [1], but in short, I_{∂} measures the amount of "new" information at this node in the lattice compared to nodes lower in the lattice. Except for ImperfectRdn, measures $\mathrm{I}_{\curlywedge}$ and $\mathrm{I}^0_{\curlywedge}$ reach the same decomposition for all presented examples. Per [1], the four partial informations are calculated as follows:

$$\begin{aligned} \mathrm{I}_{\partial}(X_1, X_2 : Y) &= \mathrm{I}_{\cap}(X_1, X_2 : Y) \\ \mathrm{I}_{\partial}(X_1 : Y) &= \mathrm{I}(X_1 : Y) - \mathrm{I}_{\cap}(X_1, X_2 : Y) \\ \mathrm{I}_{\partial}(X_2 : Y) &= \mathrm{I}(X_2 : Y) - \mathrm{I}_{\cap}(X_1, X_2 : Y) \\ \mathrm{I}_{\partial}(X_1 \curlyvee X_2 : Y) &= \mathrm{I}(X_1 \curlyvee X_2 : Y) - \mathrm{I}(X_1 : Y) - \mathrm{I}(X_2 : Y) + \mathrm{I}_{\cap}(X_1, X_2 : Y) \\ &= \mathrm{I}(X_1 \curlyvee X_2 : Y) - \mathrm{I}_{\partial}(X_1 : Y) - \mathrm{I}_{\partial}(X_2 : Y) - \mathrm{I}_{\partial}(X_1, X_2 : Y) \ . \end{aligned} \tag{7}$$

Example Unq (Figure 1). The desired decomposition for this example is two bits of unique information; X_1 uniquely specifies one bit of Y and X_2 uniquely specifies the other bit of Y. The chief criticism of $\mathrm{I}_{\min}$ in [2] was that $\mathrm{I}_{\min}$ calculated one bit of redundancy and one bit of synergy for Unq (Figure 1c). We see that unlike $\mathrm{I}_{\min}$, $\mathrm{I}_{\curlywedge}$ satisfyingly arrives at two bits of unique information. This is easily seen by the inequality,

$$0 \leq \mathrm{I}_{\curlywedge}(X_1, X_2 : Y) \leq \mathrm{H}(X_1 \curlywedge X_2) \leq \mathrm{I}(X_1 : X_2) = 0 \text{ bits} \ . \tag{8}$$

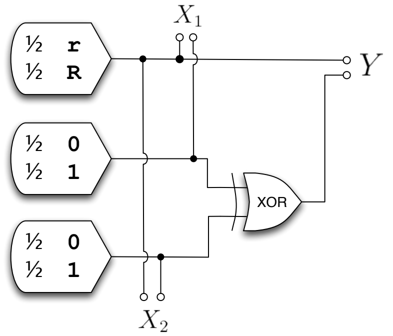
Therefore, as $\mathrm{I}(X_1 : X_2) = 0$, we have $\mathrm{I}_{\curlywedge}(X_1, X_2 : Y) = 0$ bits leading to $\mathrm{I}_{\partial}(X_1 : Y) = 1$ bit and $\mathrm{I}_{\partial}(X_2 : Y) = 1$ bit (Figure 1d).

Example RdnXor (Figure 2). In [2], RdnXor was an example where $\mathrm{I}_{\min}$ shined by reaching the desired decomposition of one bit of redundancy and one bit of synergy. We see

[9] This is the same notation used in [4].

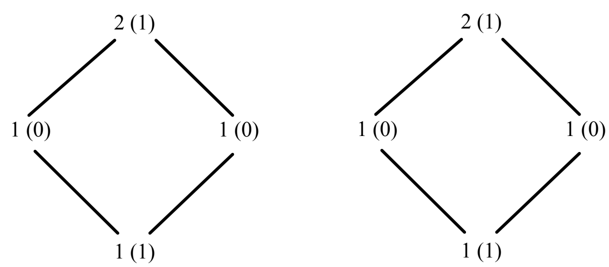
X_1	X_2	Y	
r0	r0	r0	1/8
r0	r1	r1	1/8
r1	r0	r1	1/8
r1	r1	r0	1/8
R0	R0	R0	1/8
R0	R1	R1	1/8
R1	R0	R1	1/8
R1	R1	R0	1/8

(a) $\Pr(x_1, x_2, y)$



(b) circuit diagram

$$\mathrm{I}(X_1 \curlyvee X_2 : Y) = 2$$
$$\mathrm{I}(X_1 : Y) = 1$$
$$\mathrm{I}(X_2 : Y) = 1$$
$$\mathrm{I}_{\min}(X_1, X_2 : Y) = 1$$
$$\mathrm{I}_{\curlywedge}(X_1, X_2 : Y) = 1$$



(c) $\mathrm{I}_{\min}$ (d) $\mathrm{I}_{\curlywedge}$ and $\mathrm{I}^0_{\curlywedge}$

Figure 2: Example RdnXor. This is the canonical example of redundancy and synergy coexisting. $\mathrm{I}_{\min}$ and $\mathrm{I}_{\curlywedge}$ each reach the desired decomposition of one bit of redundancy and one bit of synergy. This is the simplest example demonstrating $\mathrm{I}_{\curlywedge}$ and $\mathrm{I}^0_{\curlywedge}$ correctly extracting the embedded redundant bit within X_1 and X_2.

that $\mathrm{I}_{\curlywedge}$ finds this same answer. $\mathrm{I}_{\curlywedge}$ extracts the common random variable within X_1 and X_2, the `r`/`R` bit, and calculates the mutual information between the common random variable and Y to arrive at $\mathrm{I}_{\curlywedge}(X_1, X_2 : Y) = 1$ bit.

Example ImperfectRdn (Figure 3). ImperfectRdn highlights the foremost shortcoming of $\mathrm{I}_{\curlywedge}$; $\mathrm{I}_{\curlywedge}$ does not detect "imperfect" or "lossy" correlations between X_1 and X_2. Given ($\mathbf{LP_0}$), we can determine the desired decomposition analytically. First, $\mathrm{I}(X_1 \curlyvee X_2 : Y) = \mathrm{I}(X_1 : Y) = 1$ bit; therefore, $\mathrm{I}\big(X_2 : Y | X_1\big) = \mathrm{I}(X_1 \curlyvee X_2 : Y) - \mathrm{I}(X_1 : Y) \ = \ 0$ bits. This determines two of the partial informations—the synergistic information $\mathrm{I}_\partial(X_1 \curlyvee X_2 : Y)$ and the unique information $\mathrm{I}_\partial(X_2 : Y)$ are both zero. Then, the redundant information $\mathrm{I}_\partial(X_1, X_2 : Y) = \mathrm{I}(X_2 : Y) - \mathrm{I}_\partial(X_2 : Y) = \mathrm{I}(X_2 : Y) = 0.99$ bits. Having determined three of the partial informations, we compute the final unique information $\mathrm{I}_\partial(X_1 : Y) = \mathrm{I}(X_1 : Y) - 0.99 = 0.01$ bits.

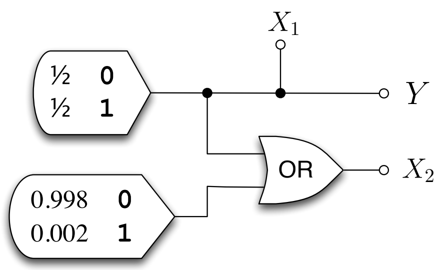
How well do $\mathrm{I}_{\min}$ and $\mathrm{I}_{\curlywedge}$ match the desired decomposition of ImperfectRdn? We see that $\mathrm{I}_{\min}$ calculates the desired decomposition (Figure 3c); however, $\mathrm{I}_{\curlywedge}$ does not (Figure 3d). Instead, $\mathrm{I}_{\curlywedge}$ calculates zero redundant information, that $\mathrm{I}_\cap(X_1, X_2 : Y) = 0$ bits. This unpleasant answer arises from $\Pr(X_1 = \mathtt{1}, X_2 = \mathtt{0}) > 0$. If this were zero, ImperfectRdn reverts to the example Rdn (Figure 5 in Appendix E) where both $\mathrm{I}_{\curlywedge}$ and $\mathrm{I}_{\min}$ reach the desired one bit of redundant information. Due to the nature of the common random variable, $\mathrm{I}_{\curlywedge}$ only sees the "deterministic" correlations between X_1 and X_2—add even an iota of noise between X_1 and X_2 and $\mathrm{I}_{\curlywedge}$ plummets to zero. This highlights a related issue with $\mathrm{I}_{\curlywedge}$; it is not continuous—an arbitrarily small change in the probability distribution can result in a discontinuous jump in the value of $\mathrm{I}_{\curlywedge}$. As with traditional information measures such as the entropy and the mutual information, it may be desirable to have an $\mathrm{I}_\cap$ measure that is continuous over the simplex.

To summarize, ImperfectRdn shows that when there are additional "imperfect" correlations between A and B, i.e. $\mathrm{I}(A : B | A \curlywedge B) > 0$, $\mathrm{I}_{\curlywedge}$ sometimes *underestimates* the ideal $\mathrm{I}_\cap(A, B : Y)$.

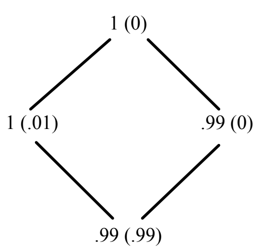
X_1	X_2	Y	
0	0	0	0.499
0	1	0	0.001
1	1	1	0.500

(a) $\Pr(x_1, x_2, y)$

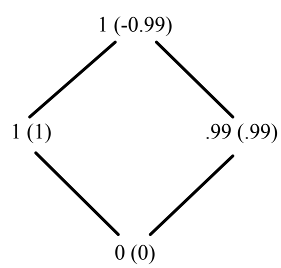
$$\mathrm{I}(X_1 \curlyvee X_2 : Y) = 1$$
$$\mathrm{I}(X_1 : Y) = 1$$
$$\mathrm{I}(X_2 : Y) = 0.99$$
$$\mathrm{I}_{\min}(X_1, X_2 : Y) = 0.99$$
$$\mathrm{I}_{\curlywedge}(X_1, X_2 : Y) = 0$$



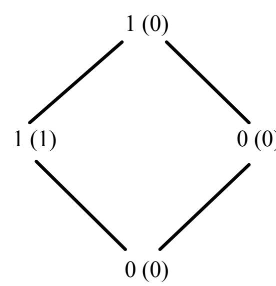
(b) circuit diagram



(c) $\mathrm{I}_{\min}$



(d) $\mathrm{I}_{\curlywedge}$



(e) $\mathrm{I}^0_{\curlywedge}$

Figure 3: Example ImperfectRdn. $\mathrm{I}_{\curlywedge}$ is blind to the noisy correlation between X_1 and X_2 and calculates zero redundant information. An ideal $\mathrm{I}_{\cap}$ measure would detect that all of the information X_2 specifies about Y is also specified by X_1 to calculate $\mathrm{I}_{\cap}(X_1, X_2 : Y) = 0.99$ bits.

6 Negative synergy and state-dependent (GP)

In ImperfectRdn we saw $I_\curlywedge$ calculate a synergy of -0.99 bits (Figure 3d). What does this mean? Could negative synergy be a "real" property of Shannon information? When $n = 2$, it's fairly easy to diagnose the cause of negative synergy from the equation for $I_\partial(X_1, X_2 : Y)$ in eq. (7). Given (**GP**) and (**SR**), negative synergy occurs if and only if,

$$\begin{aligned} I(X_1 \curlyvee X_2 : Y) &< I(X_1 : Y) + I(X_2 : Y) - I_\cap(X_1, X_2 : Y) \\ &= I_\cup(X_1, X_2 : Y) \ . \end{aligned} \tag{9}$$

From eq. (9), we see negative synergy occurs when $I_\cap$ is small, perhaps *too small*. Equivalently, negative synergy occurs when the joint r.v. conveys *less* about Y than the two r.v.'s X_1 and X_2 convey separately—mathematically, when $I(X_1 \curlyvee X_2 : Y) < I_\cup(X_1, X_2 : Y)$.[10] On the face of it this sounds strange. No usable structure in X_1 or X_2 "disappears" after they are combined by $Z = X_1 \curlyvee X_2$. By the definition of $\curlyvee$, there are always functions f_1 and f_2 such that $X_1 \cong f_1(Z)$ and $X_2 \cong f_2(Z)$. Therefore, if your favorite $I_\cap$ measure does not satisfy ($\mathbf{LP_0}$), it is likely too strict.

This means that, to our surprise, our measure $I^0_\curlywedge$ does not account for the full zero-information overlap between $I^0(X_1 : Y)$ and $I^0(X_2 : Y)$. This is shown in example Subtle (Figure 4) where $I^0_\curlywedge$ calculates a synergy of -0.252 bits. Defining a zero-error $I_\cap$ that satisfies ($\mathbf{LP_0}$) is a matter of ongoing research.

6.1 Consequences of state-dependent (GP)

In [2] it's argued that $I_{\min}$ upperbounds the ideal $I_\cap$. Inspired by $I_{\min}$ assuming state-dependent (**SR**) and ($\mathbf{M_0}$) to achieve a tighter upperbound on $I_\cap$, we assume state-dependent (**GP**) to achieve a tighter lowerbound on $I_\cap$ for $n = 2$. Our bound, denoted I_{smp} for "sum minus pair", is defined as,

$$I_{\text{smp}}(X_1, X_2 : Y) \equiv \sum_{y \in Y} \Pr(y) \max\left[0, I(X_1 : y) + I(X_2 : y) - I(X_1 \curlyvee X_2 : y)\right] \ , \tag{10}$$

where $I(\bullet : y)$ is the same Kullback-Liebler divergence from eq. (1).

For example Subtle, the target $Y \cong X_1 \curlyvee X_2$, therefore per (**Id**), $I_\cap(X_1, X_2 : Y) = I(X_1 : X_2) = 0.252$ bits. However, given state-dependent (**GP**), applying I_{smp} yields $I_\cap(X_1, X_2 : Y) \geq 0.390$. Therefore, (**Id**) and state-dependent (**GP**) are incompatible. Secondly, given state-dependent (**GP**), example Subtle additionally illustrates a conjecture from [4] that the intersection information two predictors have about a target can exceed the mutual information between them, i.e., $I_\cap(X_1, X_2 : Y) \not\leq I(X_1 : X_2)$.

7 Conclusion and Path Forward

We've made incremental progress on several fronts towards the ideal Shannon $I_\cap$.

Desired Properties. We have tightened, expanded, and pruned the desired properties for $I_\cap$. Particularly,

- (**LB**) is a non-contentious yet tighter lower-bound on $I_\cap$ than (**GP**).
- Motivated by the natural equality $I_\cap(X_1, \ldots, X_n : Y) = I_\cap(X_1, \ldots, X_n, Y : Y)$, we introduce ($\mathbf{M_1}$) as a desired property.
- What was before an implicit assumption, we introduce (**Eq**) to better ground one's thinking.

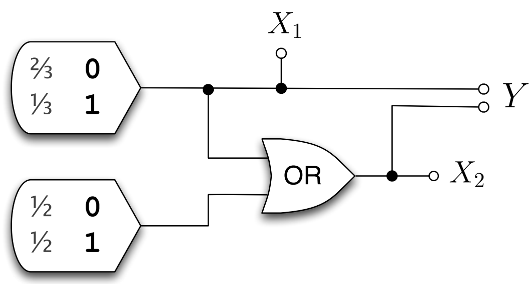
[10] $I_\cap$ and $I_\cup$ are duals related by the inclusion–exclusion principle. For arbitrary n, this is $I_\cup(X_1, \ldots, X_n : Y) = \sum_{\mathbf{S} \subseteq \{X_1, \ldots, X_n\}} (-1)^{|\mathbf{S}|+1} I_\cap\big(S_1, \ldots, S_{|\mathbf{S}|} : Y\big)$.

X_1	X_2	Y	
0	0	00	1/3
1	1	11	1/3
0	1	01	1/3

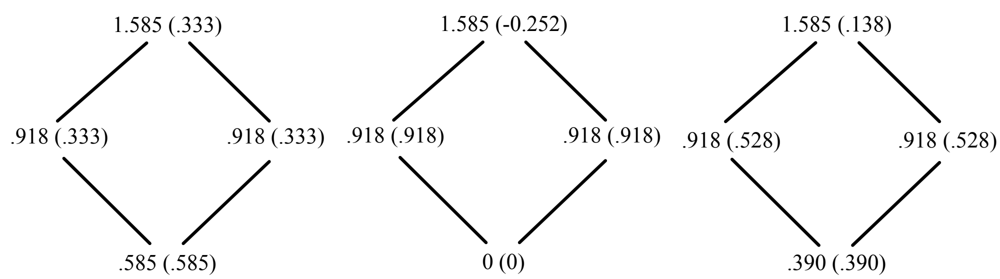
(a) $\Pr(x_1, x_2, y)$

$$\mathrm{I}(X_1 \curlyvee X_2 : Y) = 1.585$$
$$\mathrm{I}(X_1 : Y) = 0.918$$
$$\mathrm{I}(X_2 : Y) = 0.918$$
$$\mathrm{I}(X_1 : X_2) = 0.252$$

$$\mathrm{I}_{\min}(X_1, X_2 : Y) = 0.585$$
$$\mathrm{I}_{\curlywedge}(X_1, X_2 : Y) = 0.0$$
$$\mathrm{I}_{\mathrm{smp}}(X_1, X_2 : Y) = 0.390$$



(b) circuit diagram



(c) $\mathrm{I}_{\min}$ (d) $\mathrm{I}_{\curlywedge}$ and $\mathrm{I}^0_{\curlywedge}$ (e) $\mathrm{I}_{\mathrm{smp}}$

Figure 4: Example SUBTLE. In this example both $\mathrm{I}_{\curlywedge}$ and $\mathrm{I}^0_{\curlywedge}$ calculate a synergy of -0.252 bits of synergy. What kind of redundancy must be captured for a nonnegative decomposition for this example?

- A separate chain-rule property is superfluous. Any desirable properties of conditional $\mathrm{I}_\cap$ are simply consequences of (**GP**) and (**TM**).

A new measure. Based on the Gács-Körner common random variable, we introduced a new Shannon $\mathrm{I}_\cap$ measure. Our measure, $\mathrm{I}_\curlywedge$, is theoretically principled and the first to satisfy (**TM**).

How to improve. We identified where $\mathrm{I}_\curlywedge$ fails; it does not detect "imperfect" correlations between X_1 and X_2. One next step is to develop a less stringent $\mathrm{I}_\cap$ measure that satisfies ($\mathbf{LP_0}$) for simple nondeterministic examples like ImperfectRdn while still satisfying (**TM**).

To our surprise, example Subtle shows that $\mathrm{I}_\curlywedge^0$ does not satisfy ($\mathbf{LP_0}$)! This suggests that $\mathrm{I}_\curlywedge^0$ is too strict—what kind of zero-error informational overlap is $\mathrm{I}_\curlywedge^0$ not capturing? A separate next step is to formalize what exactly is required for a zero-error $\mathrm{I}_\cap$ to satisfy ($\mathbf{LP_0}$).

Finally, we showed that state-dependent (**GP**), a seemingly reasonable property, is incompatible with (**Id**) and moreover entails that $\mathrm{I}_\cap(X_1, X_2 : Y)$ can exceed $\mathrm{I}(X_1 : X_2)$.

Acknowledgements. VG thanks Tracey Ho and EKPC thanks Hua Li for valuable discussions. While intrepidly pulling back the veil of ignorance, VG was funded by a DOE CSGF fellowship. EKPC was funded by the CSU Information Science and Technology Center (ISTeC). RGJ and JPC were funded by ARO grant W911NF-12-1-0234, and CJE by a grant from the John Templeton Foundation.

References

[1] Williams PL, Beer RD (2010) Nonnegative decomposition of multivariate information. CoRR abs/1004.2515.

[2] Griffith V, Koch C (2014) Quantifying synergistic mutual information. In: Prokopenko M, editor, Guided Self-Organization: Inception. Springer.

[3] Harder M, Salge C, Polani D (2012) A bivariate measure of redundant information. CoRR abs/1207.2080.

[4] Bertschinger N, Rauh J, Olbrich E, Jost J (2012) Shared information – new insights and problems in decomposing information in complex systems. CoRR abs/1210.5902.

[5] Lizier JT, Flecker B, Williams PL (2013) Towards a synergy-based approach to measuring information modification. CoRR abs/1303.3440.

[6] Li H, Chong EKP (2011) On a connection between information and group lattices. Entropy 13: 683–708.

[7] Gács P, Körner J (1973) Common information is far less than mutual information. Problems of Control and Informaton Theory 2: 149–162.

[8] Wyner AD (1975) The common information of two dependent random variables. IEEE Transactions in Information Theory 21: 163–179.

[9] Cerf NJ, Adami C (1997) Negative entropy and information in quantum mechanics. Phys Rev Lett 79: 5194–5197.

[10] Wolf S, Wullschleger J (2004) Zero-error information and applications in cryptography. Proc IEEE Information Theory Workshop 04: 1–6.

Appendix

A Algorithm for Computing Common Random Variable

Given n random variables $X_1, \ldots, X_n$, the common random variable $X_1 \curlywedge \cdots \curlywedge X_n$ is computed by steps 1–3 in Appendix B.

B Algorithm for Computing $\mathrm{I}_\curlywedge$

1. For each X_i for $i = 1, \ldots, n$, take its states x_i and place them as nodes on a graph. At the end of this process there will be $\sum_{i=1}^{n} |X_i|$ nodes on the graph.
2. For each pair of RVs X_i, X_j $(i \neq j)$, draw an undirected edge connecting nodes x_i and x_j if $\Pr\big(x_i, x_j\big) > 0$. At the end of this process the undirected graph will consist of k connected components $1 \leq k \leq \min_i |X_i|$. Denote these k disjoint components as $\mathbf{c_1}, \ldots, \mathbf{c_k}$.
3. Each connected component of the graph constitutes a distinct state of the common random variable Q, i.e., $|Q| = k$. Denote the states of the common random variable Q by $q_1, \ldots, q_k$.
4. Construct the joint probability distribution $\Pr(Q, Y)$ as follows. For every state $(q_i, y) \in Q \times Y$, the joint probability is created by summing over the entries of $\Pr(x_1, \ldots, x_n, y)$ in component i. More precisely,

$$\Pr(Q = q_i, Y = y) = \sum_{x_1,\ldots,x_n} \Pr(x_1, \ldots, x_n, y) \qquad \text{if } \{x_1, \ldots, x_n\} \subseteq \mathbf{c_i} \,.$$

5. Using $\Pr(Q, Y)$, compute $\mathrm{I}_\curlywedge(X_1, \ldots, X_n : Y)$ simply by computing the Shannon mutual information between Q and Y, i.e., $\mathrm{I}(Q : Y) = \mathrm{D_{KL}}\big[\Pr(Q, Y) \big\| \Pr(Q) \Pr(Y)\big]$.

C Lemmas and Proofs

C.1 Lemmas on Desired Properties

Lemma 2. *If* (**LB**) *holds, then* $\mathrm{I}_\cap(X_1, \ldots, X_n : Y) \geq \mathcal{I}\left(X_1 \curlywedge \cdots \curlywedge X_n : Y\right)$.

Proof. Assume that (**LB**) holds. By definition, $X_1 \curlywedge \cdots \curlywedge X_n \preceq X_i$ for $i = 1, \ldots, n$. So, by (**LB**), we immediately conclude that $\mathrm{I}_\cap(X_1, \ldots, X_n : Y) \geq \mathcal{I}\left(X_1 \curlywedge \cdots \curlywedge X_n : Y\right)$, which is the desired result. ☐

For the converse, we need the following assumption:

(**IM**) If $X_1 \preceq X_2$, then $\mathcal{I}\left(X_1 : Y\right) \leq \mathcal{I}\left(X_2 : Y\right)$.

Lemma 3. *Suppose that* (**IM**) *holds, and that* $\mathrm{I}_\cap(X_1, \ldots, X_n : Y) \geq \mathcal{I}\left(X_1 \curlywedge \cdots \curlywedge X_n : Y\right)$. *Then* (**LB**) *holds.*

Proof. Assume that $\mathrm{I}_\cap(X_1, \ldots, X_n : Y) \geq \mathcal{I}\left(X_1 \curlywedge \cdots \curlywedge X_n : Y\right)$. Let $Q \preceq X_i$ for $i = 1, \ldots, n$. Because $X_1 \curlywedge \cdots \curlywedge X_n$ is the largest (informationally richest) random variable that is informationally poorer than X_i for $i = 1, \ldots, n$, it follows that $Q \preceq X_1 \curlywedge \cdots \curlywedge X_n$. Therefore, by (**IM**), $\mathcal{I}\left(X_1 \curlywedge \cdots \curlywedge X_n : Y\right) \geq \mathcal{I}\left(Q : Y\right)$. Hence, $\mathrm{I}_\cap(X_1, \ldots, X_n : Y) \geq \mathcal{I}\left(Q : Y\right)$ also, which completes the proof. ☐

Remark: Assumption (**IM**) is satisfied by zero-error information and Shannon mutual information.

Lemma 4. *Given* $\mathrm{I}_\cap$, $X_1, \ldots, X_n$, Y, *and* Z, *consider the conditional intersection information*

$$\mathrm{I}_\cap(X_1, \ldots, X_n : Z | Y) = \mathrm{I}_\cap(X_1, \ldots, X_n : Y \curlyvee Z) - \mathrm{I}_\cap(X_1, \ldots, X_n : Y) \,.$$

Suppose that (**GP**), (**Eq**), *and* (**TM**) *hold. Then, the following properties hold:*

- $\mathrm{I}_\cap(X_1, \ldots, X_n : Z | Y) \geq 0$.
- $\mathrm{I}_\cap(X_1, \ldots, X_n : Z | Y) = \mathrm{I}_\cap(X_1, \ldots, X_n : Z)$ *if* Y *is a constant.*

Proof. We have $Y \preceq Y \curlyvee Z$. Therefore, by (**TM**), it immediately follows that $\mathrm{I}_\cap(X_1, \ldots, X_n \!:\! Z|Y) \geq 0$.

Next, suppose that Y is a constant. Then $Y \preceq Z$, and hence $Y \curlyvee Z \cong Z$. By (**Eq**), $\mathrm{I}_\cap(X_1, \ldots, X_n \!:\! Y \curlyvee Z) = \mathrm{I}_\cap(X_1, \ldots, X_n \!:\! Z)$. Moreover, by (**GP**), $\mathrm{I}_\cap(X_1, \ldots, X_n \!:\! Y) = 0$. Thus, $\mathrm{I}_\cap(X_1, \ldots, X_n \!:\! Z|Y) = \mathrm{I}_\cap(X_1, \ldots, X_n \!:\! Z)$ as desired. □

C.2 Properties of $\mathrm{I}^0_\curlywedge$

Lemma 5. *The measure of intersection information $\mathrm{I}^0_\curlywedge(X_1, \ldots, X_n \!:\! Y)$ satisfies* (**GP**)*,* (**Eq**)*,* (**TM**)*,* ($\mathbf{M_0}$)*, and* ($\mathbf{S_0}$)*, but not* ($\mathbf{LP_0}$)*.*

Proof. (**GP**): The inequality $\mathrm{I}^0_\curlywedge(X_1, \ldots, X_n : Y) \geq 0$ follows immediately from the identity $\mathrm{I}^0_\curlywedge(X_1, \ldots, X_n : Y) = \mathrm{H}(X_1 \curlywedge \cdots \curlywedge X_n \curlywedge Y)$ and the nonnegativity of $\mathrm{H}(\cdot)$. Next, if Y is a constant, then by the generalized absorption law, $X_1 \curlywedge \cdots \curlywedge X_n \curlywedge Y \cong Y$. Thus, by the invariance of $\mathrm{H}(\cdot)$ (Lemma 1(a)), $\mathrm{H}(X_1 \curlywedge \cdots \curlywedge X_n \curlywedge Y) = \mathrm{H}(Y) = 0$.

(**Eq**): Consider $X_1 \curlywedge \cdots \curlywedge X_n \curlywedge Y$. The equivalence class (with respect to $\cong$) in which this random variable resides is closed under substitution of X_i (for $i = 1, \ldots, n$) or Y by an informationally equivalent random variable. Hence, because $\mathrm{I}^0_\curlywedge(X_1, \ldots, X_n : Y) = \mathrm{H}(X_1 \curlywedge \cdots \curlywedge X_n \curlywedge Y)$ and $\mathrm{H}(\cdot)$ is invariant over the equivalence class of random variables that are informationally equivalent to $X_1 \curlywedge \cdots \curlywedge X_n \curlywedge Y$ (by Lemma 1(a)), the desired result holds.

(**TM**): Suppose that $Y \preceq Z$. Then, $X_1 \curlywedge \cdots \curlywedge X_n \curlywedge Y \preceq X_1 \curlywedge \cdots \curlywedge X_n \curlywedge Z$. Then, we have

$$\begin{aligned}
\mathrm{I}^0_\curlywedge(X_1, \ldots, X_n : Y) &= \mathrm{H}(X_1 \curlywedge \cdots \curlywedge X_n \curlywedge Y) \\
&\leq \mathrm{H}(X_1 \curlywedge \cdots \curlywedge X_n \curlywedge Z) \quad \text{by monotonicity of } \mathrm{H}(\cdot) \text{ (Lemma 1(b))} \\
&= \mathrm{I}^0_\curlywedge(X_1, \ldots, X_n : Z)\,,
\end{aligned}$$

as desired.

($\mathbf{M_0}$): By the generalized absorption law, $X_1 \curlywedge \cdots \curlywedge X_n \curlywedge W \curlywedge Y \preceq X_1 \curlywedge \cdots \curlywedge X_n \curlywedge Y$. Hence,

$$\begin{aligned}
\mathrm{I}^0_\curlywedge(X_1, \ldots, X_n, W : Y) &= \mathrm{H}(X_1 \curlywedge \cdots \curlywedge X_n \curlywedge W \curlywedge Y) \\
&\leq \mathrm{H}(X_1 \curlywedge \cdots \curlywedge X_n \curlywedge Y) \quad \text{by monotonicity of } \mathrm{H}(\cdot) \text{ (Lemma 1(b))} \\
&= \mathrm{I}^0_\curlywedge(X_1, \ldots, X_n : Y)\,,
\end{aligned}$$

as desired.

Next, suppose that there exists $Z \in \{X_1, \ldots, X_n\}$ such that $Z \preceq W$. Then, by the generalized absorption law, $X_1 \curlywedge \cdots \curlywedge X_n \curlywedge W \curlywedge Y \cong X_1 \curlywedge \cdots \curlywedge X_n \curlywedge Y$. Hence,

$$\begin{aligned}
\mathrm{I}^0_\curlywedge(X_1, \ldots, X_n, W : Y) &= \mathrm{H}(X_1 \curlywedge \cdots \curlywedge X_n \curlywedge W \curlywedge Y) \\
&= \mathrm{H}(X_1 \curlywedge \cdots \curlywedge X_n \curlywedge Y) \quad \text{by invariance of } \mathrm{H}(\cdot) \text{ (Lemma 1(a))} \\
&= \mathrm{I}^0_\curlywedge(X_1, \ldots, X_n : y)\,,
\end{aligned}$$

as desired.

($\mathbf{S_0}$): By the commutativity law, $X_1 \curlywedge \cdots \curlywedge X_n \curlywedge Y$ is invariant (with respect to $\cong$) under reordering of $X_1, \ldots, X_n$. Hence, the desired result follows immediately from the identity $\mathrm{I}^0_\curlywedge(X_1, \ldots, X_n : Y) = \mathrm{H}(X_1 \curlywedge \cdots \curlywedge X_n \curlywedge Y)$ and the invariance of $\mathrm{H}(\cdot)$ (Lemma 1(a)).

($\mathbf{LP_0}$): For $\mathrm{I}^0_\curlywedge$, ($\mathbf{LP_0}$) relative to zero-error information can be written as

$$\mathrm{H}(X_1 \curlywedge X_2 \curlywedge Y) \geq \mathrm{H}(X_1 \curlywedge Y) + \mathrm{H}(X_2 \curlywedge Y) - \mathrm{H}\big((X_1 \curlyvee X_2) \curlywedge Y\big)\,. \tag{11}$$

However, this inequality does not hold in general. To see this, suppose that it does hold for arbitrary X_1, X_2, and Y. Note that $(X_1 \curlyvee X_2) \curlywedge Y \preceq Y$, which implies that $\mathrm{H}\big((X_1 \curlyvee X_2) \curlywedge Y\big) \leq \mathrm{H}(Y)$ (by monotonicity of $\mathrm{H}(\cdot)$). Hence, the inequality (11) implies that

$$\mathrm{H}(X_1 \curlywedge X_2 \curlywedge Y) \geq \mathrm{H}(X_1 \curlywedge Y) + \mathrm{H}(X_2 \curlywedge Y) - \mathrm{H}(Y)\,.$$

Rewriting this, we get

$$\mathrm{H}(X_1 \curlywedge Y) + \mathrm{H}(Y \curlywedge X_2) \le \mathrm{H}(X_1 \curlywedge Y \curlywedge X_2) + \mathrm{H}(Y)\,.$$

But this is the supermodularity law for common information, which is known to be false in general; see [6], Section 5.4.

□

Lemma 6. *With respect to zero-error information, the measure of intersection information* $\mathrm{I}^0_\curlywedge(X_1,\ldots,X_n\!:\!Y)$ *satisfies* (**LB**)*,* (**SR**)*, and* (**Id**)*.*

Proof. (**LB**): Suppose that $Q \preceq X_i$ for $i = 1,\ldots,n$. Because $X_1 \curlywedge \cdots \curlywedge X_n$ is the largest (informationally richest) random variable that is informationally poorer than X_i for $i = 1,\ldots,n$, it follows that $Q \preceq X_1 \curlywedge \cdots \curlywedge X_n$. This implies that $X_1 \curlywedge \cdots \curlywedge X_n \curlywedge Y \succeq Q \curlywedge Y$. Therefore,

$$\begin{aligned}
\mathrm{I}^0_\curlywedge(X_1,\ldots,X_n : Y) &= \mathrm{H}(X_1 \curlywedge \cdots \curlywedge X_n \curlywedge Y) \\
&\ge \mathrm{H}(Q \curlywedge Y) \quad \text{by monotonicity of } \mathrm{H}(\cdot) \text{ (Lemma 1(b))} \\
&= \mathrm{I}^0(Q\!:\!Y)\,,
\end{aligned}$$

as desired.

(**SR**): We have $\mathrm{I}^0_\curlywedge(X_1 : Y) = \mathrm{H}(X_1 \curlywedge Y) = \mathrm{I}^0(X_1\!:\!Y)$.

(**Id**): By the associative and absorption laws, we have $X \curlywedge Y \curlywedge (X \curlyvee Y) \cong X \curlywedge Y$. Thus,

$$\begin{aligned}
\mathrm{I}^0_\curlywedge(X, Y : X \curlyvee Y) &= \mathrm{H}\big(X \curlywedge Y \curlywedge (X \curlyvee Y)\big) \\
&= \mathrm{H}(X \curlywedge Y) \quad \text{by invariance of } \mathrm{H}(\cdot) \text{ (Lemma 1(a))} \\
&= \mathrm{I}^0(X\!:\!Y)\,,
\end{aligned}$$

as desired.

□

Lemma 7. *The measure of intersection information* $\mathrm{I}^0_\curlywedge(X_1,\ldots,X_n\!:\!Y)$ *satisfies* ($\mathbf{M_1}$) *and* ($\mathbf{S_1}$)*, but not* ($\mathbf{LP_1}$)*.*

Proof. ($\mathbf{M_1}$): The desired inequality is identical to ($\mathbf{M_0}$), so it remains to prove the sufficient condition for equality. Suppose that there exists $Z \in \{X_1,\ldots,X_n,Y\}$ such that $Z \preceq W$. Then, by the generalized absorption law, $X_1 \curlywedge \cdots \curlywedge X_n \curlywedge W \curlywedge Y \cong X_1 \curlywedge \cdots \curlywedge X_n \curlywedge Z$. Hence,

$$\begin{aligned}
\mathrm{I}^0_\curlywedge(X_1,\ldots,X_n, W : Y) &= \mathrm{H}(X_1 \curlywedge \cdots \curlywedge X_n \curlywedge W \curlywedge Y) \\
&= \mathrm{H}(X_1 \curlywedge \cdots \curlywedge X_n \curlywedge Z) \quad \text{by invariance of } \mathrm{H}(\cdot) \text{ (Lemma 1(a))} \\
&= \mathrm{I}^0_\curlywedge(X_1,\ldots,X_n : Z)\,,
\end{aligned}$$

as desired.

($\mathbf{S_1}$): By the commutativity law, $X_1 \curlywedge \cdots \curlywedge X_n \curlywedge Y$ is invariant (with respect to $\cong$) under reordering of $X_1,\ldots,X_n,Y$. Hence, the desired result follows immediately from the identity $\mathrm{I}^0_\curlywedge(X_1,\ldots,X_n : Y) = \mathrm{H}(X_1 \curlywedge \cdots \curlywedge X_n \curlywedge Y)$ and the invariance of $\mathrm{H}(\cdot)$ (Lemma 1(a)).

($\mathbf{LP_1}$): This follows from not satisfying ($\mathbf{LP_0}$).

□

C.3 Properties of $\mathrm{I}_\curlywedge$

Lemma 8. *The measure of intersection information* $\mathrm{I}_\curlywedge(X_1,\ldots,X_n\!:\!Y)$ *satisfies* (**GP**)*,* (**Eq**)*,* (**TM**)*,* ($\mathbf{M_0}$)*, and* ($\mathbf{S_0}$)*, but not* ($\mathbf{LP_0}$)*.*

Proof. (**GP**): The inequality $\mathrm{I}_{\curlywedge}(X_1, \ldots, X_n : Y) \geq 0$ follows immediately from the identity $\mathrm{I}_{\curlywedge}(X_1, \ldots, X_n : Y) = \mathrm{I}(X_1 \curlywedge \cdots \curlywedge X_n \!:\! Y)$ and the nonnegativity of mutual information. Next, suppose that Y is a constant. Then $\mathrm{H}(Y) = 0$. Moreover, $Y \preceq X_1 \curlywedge \cdots \curlywedge X_n$ by definition of $\curlywedge$. Thus, by Lemma 1(c), $\mathrm{H}\big(Y|X_1 \curlywedge \cdots \curlywedge X_n\big) = 0$, and

$$\begin{aligned}
\mathrm{I}_{\curlywedge}(X_1, \ldots, X_n : Y) &= \mathrm{I}(X_1 \curlywedge \cdots \curlywedge X_n \!:\! Y) \\
&= \mathrm{I}(Y \!:\! X_1 \curlywedge \cdots \curlywedge X_n) \\
&= \mathrm{H}(Y) - \mathrm{H}\big(Y|X_1 \curlywedge \cdots \curlywedge X_n\big) \\
&= 0.
\end{aligned}$$

(**Eq**): Consider $X_1 \curlywedge \cdots \curlywedge X_n \curlywedge Y$. The equivalence class (with respect to $\cong$) in which this random variable resides is closed under substitution of X_i (for $i = 1, \ldots, n$) or Y by an informationally equivalent random variable. Hence, because

$$\begin{aligned}
\mathrm{I}_{\curlywedge}(X_1, \ldots, X_n : Y) &= \mathrm{H}(Y) - \mathrm{H}\big(Y|X_1 \curlywedge \cdots \curlywedge X_n\big) \\
&= \mathrm{H}(X_1 \curlywedge \cdots \curlywedge X_n) - \mathrm{H}\big(X_1 \curlywedge \cdots \curlywedge X_n|Y\big) \,,
\end{aligned}$$

by Lemma 1(a), the desired result holds.

(**TM**): Suppose that $Y \preceq Z$. For simplicity, let $Q = X_1 \curlywedge \cdots \curlywedge X_n$. Then,

$$\begin{aligned}
\mathrm{I}_{\curlywedge}(X_1, \ldots, X_n : Y) &= \mathrm{H}(Q) - \mathrm{H}\big(Q|Y\big) \\
&\leq \mathrm{H}(Q) - \mathrm{H}\big(Q|Z\big) \quad \text{by Lemma 1(b)} \\
&= \mathrm{I}_{\curlywedge}(X_1, \ldots, X_n : Z) \,,
\end{aligned}$$

as desired.

($\mathbf{M_0}$): By definition of $\curlywedge$, we have $X_1 \curlywedge \cdots \curlywedge X_n \curlywedge W \preceq X_1 \curlywedge \cdots \curlywedge X_n$. Hence,

$$\begin{aligned}
\mathrm{I}_{\curlywedge}(X_1, \ldots, X_n, W : Y) &= \mathrm{H}(X_1 \curlywedge \cdots \curlywedge X_n \curlywedge W) - \mathrm{H}\big(X_1 \curlywedge \cdots \curlywedge X_n \curlywedge W|Y\big) \\
&\leq \mathrm{H}(X_1 \curlywedge \cdots \curlywedge X_n) - \mathrm{H}\big(X_1 \curlywedge \cdots \curlywedge X_n|Y\big) \quad \text{by Lemma 1(b)} \\
&= \mathrm{I}_{\curlywedge}(X_1, \ldots, X_n : Y) \,,
\end{aligned}$$

as desired.

Next, suppose that there exists $Z \in \{X_1, \ldots, X_n\}$ such that $Z \preceq W$. Then, by the algebraic laws of $\curlywedge$, we have $X_1 \curlywedge \cdots \curlywedge X_n \curlywedge W \cong X_1 \curlywedge \cdots \curlywedge X_n$. Hence,

$$\begin{aligned}
\mathrm{I}_{\curlywedge}(X_1, \ldots, X_n, W : Y) &= \mathrm{H}(X_1 \curlywedge \cdots \curlywedge X_n \curlywedge W) - \mathrm{H}\big(X_1 \curlywedge \cdots \curlywedge X_n \curlywedge W|Y\big) \\
&= \mathrm{H}(X_1 \curlywedge \cdots \curlywedge X_n) - \mathrm{H}\big(X_1 \curlywedge \cdots \curlywedge X_n|Y\big) \quad \text{by Lemma 1(a)} \\
&= \mathrm{I}_{\curlywedge}(X_1, \ldots, X_n : Y) \,,
\end{aligned}$$

as desired.

($\mathbf{S_0}$): By the commutativity law, $X_1 \curlywedge \cdots \curlywedge X_n$ is invariant (with respect to $\cong$) under reordering of $X_1, \ldots, X_n$. Hence, the desired result follows immediately from the identity $\mathrm{I}_{\curlywedge}(X_1, \ldots, X_n : Y) = \mathrm{H}(X_1 \curlywedge \cdots \curlywedge X_n) - \mathrm{H}\big(X_1 \curlywedge \cdots \curlywedge X_n|Y\big)$ and Lemma 1(a).

($\mathbf{LP_0}$): A counterexample is provided by ImperfectRdn (Figure 3).

□

Lemma 9. *With respect to mutual information, the measure of intersection information* $\mathrm{I}_{\curlywedge}(X_1, \ldots, X_n \!:\! Y)$ *satisfies* (**LB**) *and* (**SR**)*, but not* (**Id**)*.*

Proof. (**LB**): Suppose that $Q \preceq X_i$ for $i = 1, \ldots, n$. Because $X_1 \curlywedge \cdots \curlywedge X_n$ is the largest (informationally richest) random variable that is informationally poorer than X_i for $i = 1, \ldots, n$, it follows that $Q \preceq X_1 \curlywedge \cdots \curlywedge X_n$. Therefore,

$$\begin{aligned}
\mathrm{I}_{\curlywedge}(X_1, \ldots, X_n : Y) &= \mathrm{H}(X_1 \curlywedge \cdots \curlywedge X_n) - \mathrm{H}\big(X_1 \curlywedge \cdots \curlywedge X_n|Y\big) \\
&\geq \mathrm{H}(Q) - \mathrm{H}\big(Q|Y\big) \quad \text{by Lemma 1(b)} \\
&= \mathrm{I}(Q \!:\! Y) \,,
\end{aligned}$$

as desired.

($\mathbf{SR}$): By definition, $I_\curlywedge(X_1 : Y) = I(X_1 \!:\! Y)$.

($\mathbf{Id}$): We have $X \curlywedge Y \preceq X \curlyvee Y$ by definition of $\curlywedge$ and $\curlyvee$. Thus,

$$\begin{aligned} I_\curlywedge(X, Y : X \curlyvee Y) &= I(X \curlywedge Y \!:\! X \curlyvee Y) \\ &= H(X \curlywedge Y) - H\big(X \curlywedge Y | X \curlyvee Y\big) \\ &= H(X \curlywedge Y) \quad \text{by Lemma 1(a)} \\ &= I^0(X \!:\! Y) \\ &\neq I(X \!:\! Y) \ . \end{aligned}$$

□

Lemma 10. *The measure of intersection information* $I_\curlywedge(X_1, \ldots, X_n \!:\! Y)$ *does not satisfy* ($\mathbf{M_1}$)*,* ($\mathbf{S_1}$)*, and* ($\mathbf{LP_1}$)*.*

Proof. ($\mathbf{M_1}$): A counterexample is provided in ImperfectRdn (Figure 3), where $I_\curlywedge(X_1 : Y) = 0.99$ bits, yet $I_\curlywedge(X_1, Y : Y) = 0$ bits.

($\mathbf{S_1}$): A counterexample. We show $I_\curlywedge(X, X \!:\! Y) \neq I_\curlywedge(X, Y \!:\! X)$.

$$\begin{aligned} I_\curlywedge(X, X \!:\! Y) - I_\curlywedge(X, Y \!:\! X) &= I(X \!:\! Y) - I_\curlywedge(X, Y \!:\! X) \\ &= I(X \!:\! Y) - I(X \curlywedge Y \!:\! X) \\ &= I(X \!:\! Y) - H(X \curlywedge Y) - H\big(X \curlywedge Y | X\big) \\ &= I(X \!:\! Y) - H(X \curlywedge Y) \\ &\neq 0 \ . \end{aligned}$$

($\mathbf{LP_1}$): This follows from not satisfying ($\mathbf{LP_0}$).

□

D Miscellaneous Results

Simplification of $I^0_\curlywedge$

Lemma 11. *We have* $I^0_\curlywedge(X_1, \ldots, X_n \!:\! Y) = H(X_1 \curlywedge \cdots \curlywedge X_n \curlywedge Y)$*.*

Proof. By definition,

$$\begin{aligned} I^0_\curlywedge(X_1, \ldots, X_n \!:\! Y) &\equiv \max_{\Pr(Q|Y)} I^0(Q \!:\! Y) \\ &\qquad \text{subject to } Q \preceq X_i \ \forall i \in \{1, \ldots, n\} \\ &= \max_{\Pr(Q|Y)} H(Q \curlywedge Y) \\ &\qquad \text{subject to } Q \preceq X_i \ \forall i \in \{1, \ldots, n\} \end{aligned}$$

Let Q be an arbitrary random variable satisfying the constraint $Q \preceq X_i$ for $i = 1, \ldots, n$. Because $X_1 \curlywedge \cdots \curlywedge X_n$ is the largest random variable (in the sense of the partial order $\preceq$) that is informationally poorer than X_i for $i = 1, \ldots, n$, we have $Q \preceq X_1 \curlywedge \cdots \curlywedge X_n$. By the property of $\curlywedge$ pointed out before, we also have $Q \curlywedge Y \preceq X_1 \curlywedge \cdots \curlywedge X_n \curlywedge Y$. By Lemma 1(b), this implies that $H(Q \curlywedge Y) \leq H(X_1 \curlywedge \cdots \curlywedge X_n \curlywedge Y)$. Therefore, $I^0_\curlywedge(X_1, \ldots, X_n \!:\! Y) = H(X_1 \curlywedge \cdots \curlywedge X_n \curlywedge Y)$. □

Simplification of $I_\curlywedge$

Lemma 12. *We have* $I_\curlywedge(X_1, \ldots, X_n \!:\! Y) = I(X_1 \curlywedge \cdots \curlywedge X_n \!:\! Y)$*.*

Proof. By definition,

$$\begin{aligned}\mathrm{I}_{\curlywedge}(X_1,\ldots,X_n\!:\!Y) &\equiv \max_{\Pr(Q|Y)} \mathrm{I}(Q\!:\!Y)\\ &\qquad \text{subject to } Q \preceq X_i \ \forall i \in \{1,\ldots,n\}\\ &= \mathrm{H}(Y) - \min_{\Pr(Q|Y)} \mathrm{H}\big(Y|Q\big)\\ &\qquad \text{subject to } Q \preceq X_i \ \forall i \in \{1,\ldots,n\}\end{aligned}$$

Let Q be an arbitrary random variable satisfying the constraint $Q \preceq X_i$ for $i = 1,\ldots,n$. Because $X_1 \curlywedge \cdots \curlywedge X_n$ is the largest random variable (in the sense of the partial order $\preceq$) that is informationally poorer than X_i for $i = 1,\ldots,n$, we have $Q \preceq X_1 \curlywedge \cdots \curlywedge X_n$. By Lemma 1(b), this implies that $\mathrm{H}\big(Y|Q\big) \geq \mathrm{H}\big(Y|X_1 \curlywedge \cdots \curlywedge X_n \curlywedge Y\big)$. Therefore, $\mathrm{I}_{\curlywedge}(X_1,\ldots,X_n\!:\!Y) = \mathrm{I}(X_1 \curlywedge \cdots \curlywedge X_n\!:\!Y)$. ☐

Proof that $\mathrm{I}_{\curlywedge}(X_1,\ldots,X_n\!:\!Y) \leq \mathrm{I}_{\min}(X_1,\ldots,X_n : Y)$

Lemma 13. *We have* $\mathrm{I}_{\curlywedge}(X_1,\ldots,X_n\!:\!Y) \leq \mathrm{I}_{\min}(X_1,\ldots,X_n : Y)$

Proof. Starting from the definitions,

$$\begin{aligned}\mathrm{I}_{\curlywedge}(X_1,\ldots,X_n\!:\!Y) &\equiv \mathrm{I}(X_1 \curlywedge \cdots \curlywedge X_n\!:\!Y)\\ &= \sum_y \Pr(y)\,\mathrm{I}(X_1 \curlywedge \cdots \curlywedge X_n\!:\!y)\\ \mathrm{I}_{\min}(X_1,\ldots,X_n\!:\!Y) &\equiv \sum_y \Pr(y) \min_i \mathrm{I}(X_i\!:\!y)\ .\end{aligned}$$

For a particular state y, without loss of generality we define the minimizing predictor X_m by $X_m \equiv \operatorname{argmin}_{X_i} \mathrm{I}(X_i\!:\!y)$ and the common random variable $Q \equiv X_1 \curlywedge \cdots \curlywedge X_n$. It then remains to show that $\mathrm{I}(Q\!:\!y) \leq \mathrm{I}(X_m\!:\!y)$.

By definition of $\curlywedge$, we have $Q \preceq X_m$. Hence,

$$\begin{aligned}\mathrm{I}(X_m\!:\!y) &= \mathrm{H}(X_m) - \mathrm{H}\big(X_m|Y=y\big)\\ &\geq \mathrm{H}(Q) - \mathrm{H}\big(Q|Y=y\big) \quad \text{by Lemma 1(b)}\\ &= \mathrm{I}(Q\!:\!y)\ .\end{aligned}$$

☐

State-dependent zero-error information

We define the state-dependent zero-error information, $\mathrm{I}^0(X\!:\!Y=y)$ as,

$$\mathrm{I}^0(X\!:\!Y=y) \equiv \log \frac{1}{\Pr(Q=q)}\ ,$$

where the random variable $Q \equiv X \curlywedge Y$ and $\Pr(Q=q)$ is the probability of the connected component containing state $y \in Y$. This entails that $\Pr(y) \leq \Pr(q) \leq 1$. Similar to the state-dependent information, $\mathbb{E}_Y\,\mathrm{I}^0(X\!:\!y) = \mathrm{I}^0(X\!:\!Y)$, where $\mathbb{E}_Y$ is the expectation value over Y.

Proof. We define two functions f and g:

- $f : y \to q$ s.t. $\Pr\big(q|y\big) = 1$ where $q \in Q$ and $y \in Y$.
- $g : q \to \{y_1,\ldots,y_k\}$ s.t. $\Pr\big(q|y_i\big) = 1$ where $q \in Q$ and $y \in Y$.

Now we have,

$$\mathbb{E}_Y \, \mathrm{I}^0(X:y) \equiv \sum_{y \in Y} \Pr(y) \log \frac{1}{\Pr\big(f(y)\big)} \ .$$

Since each y is associated with exactly one q, we can reindex the $\sum_{y \in Y}$. We then simplify to achieve the result.

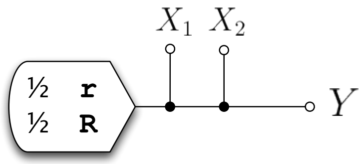
$$\begin{aligned}
\sum_{y \in Y} \Pr(y) \log \frac{1}{\Pr\big(f(y)\big)} &= \sum_{q \in Q} \sum_{y \in g(q)} \Pr(y) \log \frac{1}{\Pr\big(f(y)\big)} \\
&= \sum_{q \in Q} \sum_{y \in g(q)} \Pr(y) \log \frac{1}{\Pr(q)} = \sum_{q \in Q} \log \frac{1}{\Pr(q)} \sum_{y \in g(q)} \Pr(y) \\
&= \sum_{q \in Q} \log \frac{1}{\Pr(q)} \Pr(q) = \sum_{q \in Q} \Pr(q) \log \frac{1}{\Pr(q)} \\
&= \mathrm{H}(Q) = \mathrm{I}^0(X:Y) \ .
\end{aligned}$$

□

E Misc Figures

X_1 X_2	Y	
r r	r	$1/2$
R R	R	$1/2$

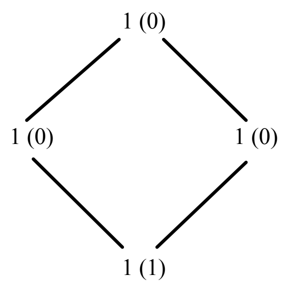
(a) $\Pr(x_1, x_2, y)$



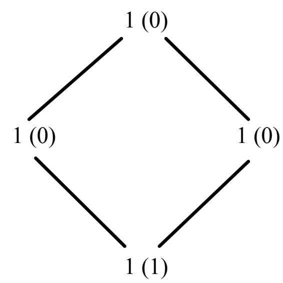
(b) circuit diagram

$$\mathrm{I}(X_1 X_2 : Y) = 1$$
$$\mathrm{I}(X_1 : Y) = 1$$
$$\mathrm{I}(X_2 : Y) = 1$$

$$\mathrm{I}_{\min}(X_1, X_2 : Y) = 1$$
$$\mathrm{I}_{\curlywedge}(X_1, X_2 : Y) = 1$$



(c) $\mathrm{I}_{\min}$



(d) $\mathrm{I}_{\curlywedge}$ and $\mathrm{I}^0_{\curlywedge}$

Figure 5: Example Rdn. In this example $\mathrm{I}_{\min}$ and $\mathrm{I}_{\curlywedge}$ reach the same answer yet diverge drastically for example ImperfectRdn.