Max-Planck-Institut
für Mathematik
in den Naturwissenschaften
Leipzig

Quantifying unique information

Nils Bertschinger[1], Johannes Rauh[1], Eckehard Olbrich[1], Jürgen Jost[1,2], Nihat Ay[1,2]

[1]Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany

[2]Santa Fe Institute, Santa Fe, USA

October 22, 2013

We propose new measures of shared information, unique information and synergistic information that can be used to decompose the multi-information of a pair of random variables (Y, Z) with a third random variable X. Our measures are motivated by an operational idea of unique information which suggests that shared information and unique information should depend only on the pair marginal distributions of (X, Y) and (X, Z). Although this invariance property has not been studied before, it is satisfied by other proposed measures of shared information. The invariance property does not uniquely determine our new measures, but it implies that the functions that we define are bounds to any other measures satisfying the same invariance property. We study properties of our measures and compare them to other candidate measures.

1. Introduction

Consider three random variables X, Y, Z with finite state spaces $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$. Suppose that we are interested in the value of X, but we can only observe Y or Z. If the tuple (Y, Z) is not independent of X, then the values of Y or Z or both of them contain information about X. The information about X contained in the tuple (Y, Z) can be distributed in different ways. For example, it may happen that Y contains information about X, but Z does not, or vice versa. In this case, it would suffice to observe one of the two variables Y, Z. It may also happen, that both X and Y contain different information, so it would be worthwhile to observe both of the variables. If both X and Y contain the same information about Z, we could chose to observe either X and Y. Finally, it is possible that neither X nor Y taken for itself contains any information about Z, but together they contain information about Z. This effect is called *synergy*, and it occurs, for example, if all variables X, Y, Z are binary, and $X = Y \operatorname{XOR} Z$. In general, all effects may be present at the same time. That is, the information that (Y, Z) has

about X is a mixture of *shared information* (that is, information contained both in X and in Y), *unique information* (that is, information that only one of X and Y has) and *synergistic* or *complementary information* (that is, information that can only be retrieved when considering Y and Z together). It is often assumed that these three types of information are everything there is, but one may ask, of course, whether there are further types of information.

The total information that (Y, Z) has about X can be quantified by the mutual information $MI(X : (Y, Z))$. Decomposing $MI(X : (Y, Z))$ into shared information, unique information and synergistic information leads to four terms, as

$$MI(X : (Y, Z)) = SI(X : Y; Z) + UI(X : Y \setminus Z) + UI(X : Z \setminus Y) + CI(X : Y; Z). \quad (1)$$

The interpretation of the four terms as informations demands that they should all be positive. Furthermore, it suggests that the following identities also hold:

$$\begin{aligned} MI(X : Y) &= SI(X : Y; Z) + UI(X : Y \setminus Z), \\ MI(X : Z) &= SI(X : Y; Z) + UI(X : Z \setminus Y). \end{aligned} \quad (2)$$

In the following, when we talk about a *binary information decomposition* we mean a set of three functions SI, UI and CI that satisfy (1) and (2).

Combining the three equalities in (1) and (2) and using the chain rule of mutual information
$$MI(X : (Y, Z)) = MI(X : Y) + MI(X : Z|Y)$$

yields the identity

$$\begin{aligned} CoI(X; Y; Z) &:= MI(X : Y) - MI(X : Y|Z) \\ &= MI(X : Y) + MI(X : Z) - MI(X : (Y, Z)) = SI(X : Y; Z) - CI(X : Y; Z), \quad (3) \end{aligned}$$

which identifies the *co-information*[1] with the difference of shared information and synergistic information. It has been known for a long time, that a positive co-information is a sign of redundancy, while a negative co-information expresses synergy [8]. However, although there have been many attempts, as of currently, there has been no fully satisfactory solution to separate the redundant and synergistic contributions to the co-information, and also a fully satisfying definition of the function UI is still missing. Observe that, since we have three equations relating the four functions $SI(X : Y; Z)$, $UI(X : Y \setminus Z)$, $UI(X : Z \setminus Y)$ and $CI(X : Y; Z)$, it suffices to specify one of them to compute the others. When defining a solution for the unique information UI, this leads to the consistency equation

$$MI(X : Z) + UI(X : Y \setminus Z) = MI(X : Y) + UI(X : Z \setminus Y). \quad (4)$$

The value of (4) can be interpreted as the *union information*, that is, the union of the informations contained in Y and in Z without the synergy.

---

[1] The co-information was originally called *interaction information* in [8].

The problem to separate the contributions of shared information and synergistic information to the co-information is probably as old as the definition of co-information itself. Nevertheless, the co-information has been widely used as a measure of synergy in the neurosciences; see, for example, [9, 7] and references therein. The first general attempt to construct a consistent information decomposition into terms corresponding to different combinations of shared and synergistic information is due to Williams and Beer [10]. See also the references in [10] for other approaches to study multivariate information. While the general approach of [10] is intriguing, the proposed measure of shared information $I_{\min}$ suffers from serious flaws, which prompted a series of other papers trying to improve these results [4, 5, 2].

In our current contribution, we propose to define the unique information as follows: For any set $\mathcal{S}$ let $\Delta(\mathcal{S})$ be the set of probability distributions on $\mathcal{S}$, and let $\Delta = \Delta(\mathcal{X} \times \mathcal{Y} \times \mathcal{Z})$ be the set of all joint distributions of X, Y and Z. Define

$$\Delta_P = \Big\{ Q \in \Delta : Q(X = x, Y = y) = P(X = x, Y = y)$$
$$\text{and } Q(X = x, Z = z) = P(X = x, Z = z) \text{ for all } x \in \mathcal{X}, y \in \mathcal{Y}, z \in \mathcal{Z} \Big\}$$

as the set of all joint distributions which have the same marginal distributions on the pairs (X, Y) and (X, Z). Then we define

$$\widetilde{UI}(X : Z \setminus Y) = \min_{Q \in \Delta_P} MI_Q(X : Y|Z),$$

where $MI_Q(X : Y|Z)$ denotes the conditional multi-information of X and Y given Z, computed with respect to the joint distribution Q. Equation (3) implies

$$\widetilde{SI}(X : Y; Z) = \max_{Q \in \Delta_P} CoI_Q(X; Y; Z),$$
$$\widetilde{CI}(X : Y; Z) = MI(X : (Y, Z)) - \min_{Q \in \Delta_P} MI_Q(X : (Y, Z)).$$

In Section 3 we show that the four functions $\widetilde{UI}$, $\widetilde{SI}$ and $\widetilde{CI}$ are non-negative, and we study further properties. In Appendix A we describe the set Δ_P in terms of a parametrization.

Our approach is motivated by the idea that unique and shared information should only depend on the marginal distribution of the pairs (X, Z) and (X, Y). This idea can be explained from an operational interpretation of unique information: Namely, if Y has unique information about X (with respect to Z), then there must be some way to extract this information. More precisely, there must be a situation in which Y can use this information to perform better at predicting the outcome of X. We make this idea precise in Section 2 and show how it naturally leads to the definition of the functions $\widetilde{UI}$, $\widetilde{SI}$ and $\widetilde{CI}$, as defined above. Section 3 contains basic properties of these three functions. In particular, Lemma 5 shows that all three functions are non-negative. Corollary 7 proves that the function $\widetilde{UI}$ is consistent with the operational idea

put forward in Section 2. In Section 4 we compare our function with other proposed information decompositions. Some examples are studied in Section 5. Remaining open problems are discussed in Section 6. The appendix contains some more technical aspects that help to compute the functions $\widetilde{UI}$, $\widetilde{SI}$ and $\widetilde{CI}$.

2. Operational interpretation

Let X, Y, Z be three random variables, let p be the marginal distribution of X, and let $\kappa \in [0,1]^{\mathcal{X} \times \mathcal{Y}}$ and $\mu \in [0,1]^{\mathcal{X} \times \mathcal{Z}}$ be (row) stochastic matrices describing the conditional distribution of Y and Z, respectively, given X. In other words, p, κ and μ satisfy

$$P(X = x, Y = y) = p(x)\kappa(x; y) \quad \text{and} \quad P(X = x, Z = y) = p(x)\mu(x; z).$$

Observe that, if $p(x) > 0$, then $\kappa(x; y)$ and $\mu(x; z)$ are uniquely defined. Otherwise, $\kappa(x; y)$ and $\mu(x; z)$ can be chosen arbitrarily. In this section, we will assume that the random variable X has full support. If this is not the case, our discussion will remain valid after replacing $\mathcal{X}$ by the support of X. In fact, the information quantities that we consider later will not depend on those matrix elements $\kappa(x; y)$ and $\mu(x; z)$ which are not uniquely defined.

Suppose that an agent has a finite set of possible actions $\mathcal{A}$. After the agent chooses her action $a \in \mathcal{A}$, she receives a reward $u(x, a)$, which not only depends on the chosen action $a \in \mathcal{A}$, but also on the value $x \in \mathcal{X}$ of the random variable X. The tuple $(p, \mathcal{A}, u)$, consisting of the prior distribution p, the set of possible actions A and the reward function u is called a *decision problem*. If the agent can observe the value x of X before choosing her action, her best strategy is to chose a such that $u(x, a) = \max_{a' \in \mathcal{A}} u(x, a')$. Suppose now, that the agent cannot observe X directly, but the agent knows the probability distribution p of X. Moreover, the agent observes a random variable Y with conditional distribution described by the row-stochastic matrix $\kappa \in [0,1]^{\mathcal{X} \times \mathcal{Y}}$. In this context, κ will also be called a *channel* from $\mathcal{X}$ to $\mathcal{Y}$. When using a channel κ, the agent's optimal strategy is to choose her action such that her expected reward

$$\sum_x P(X = x | Y = y)u(x, a) = \frac{\sum_x p(x)\kappa(x; y)u(x, a)}{\sum_{x \in \mathcal{X}} p(x)\kappa(x; y)} \tag{5}$$

is maximal. Note that, in order to maximize (5), the agent has to know (or estimate) the prior distribution of X as well as the channel κ. Often, the agent is allowed to play a stochastic strategy. However, in the present setting, the agent cannot increase her expected reward by randomizing her actions, and therefore, we only consider deterministic strategies here.

Let $R(\kappa, p, u, y)$ be the maximum of (5) (over $a \in \mathcal{A}$), and let

$$R(\kappa, p, u) = \sum_y P(Y = y)R(\kappa, p, u, y).$$

be the maximal expected reward that the agent can achieve by always choosing the optimal action.

In this setting we make the following definition:

Definition 1. Let X, Y, Z be three random variables, and let p be the marginal distribution of X.

- Y has *unique information* about X (with respect to Z), if there is a set $\mathcal{A}$ and a reward function $u \in \mathbb{R}^{\mathcal{X} \times \mathcal{A}}$ such that $R(\kappa, p, u) > R(\mu, p, u)$.

- Z has *no unique information* about X (with respect to Y), if for any set $\mathcal{A}$ and reward function $u \in \mathbb{R}^{\mathcal{X} \times \mathcal{A}}$ the inequality $R(\kappa, p, u) \geq R(\mu, p, u)$ holds. In this situation we also say that Y knows everything that Z knows about X, and we write $Y \sqsupseteq_X Z$.

This operational idea allows to distinguish when the unique information vanishes, but, unfortunately, does not allow to quantify the unique information.

As shown recently in [1], the question whether $Y \sqsupseteq_X Z$ or not, does not depend on the prior distribution p (but just on the support of p, which we assume to be $\mathcal{X}$). In fact, if p has full support, then, in order to check whether $Y \sqsupseteq Z$, it suffices to know the stochastic matrices κ, μ representing the conditional distributions of Y and Z given X.

Consider the case $\mathcal{Y} = \mathcal{Z}$ and $\kappa = \mu \in K(\mathcal{X}; \mathcal{Y})$, i.e. Y and Z use a similar channel. In this case, Y has no unique information with respect to Z, and Z has no unique information with respect to Y. Hence, in the decomposition (1) only the shared information and the synergistic information may be larger than zero. The shared information may be computed from

$$SI(X : Y; Z) = MI(X : Y) - UI(X : Y \setminus Z) = MI(X : Y) = MI(X : Z);$$

and so the synergistic information is

$$CI(X : Y; Z) = MI(X : (Y, Z)) - SI(X : Y; Z) = MI(X : (Y, Z)) - MI(X : Y).$$

Observe that in this case, the shared information can be computed from the marginal distribution of X and Y. Only the synergistic information depends on the joint distribution of X, Y and Z.

We argue that this should be the case in general: By what was said above, whether the unique information $UI(X : Y \setminus Z)$ is greater than zero only depends on the two channels κ and μ. Even more is true: The set of decision problems $(p, \mathcal{A}, u)$ such that $R(\kappa, p, u) > R(\mu, p, u)$ only depends on κ and μ (and the support of p). To quantify the unique information, this set of decision problems must be measured in some way. It is reasonable to expect that this quantification can be achieved by taking into account only the marginal distribution p of X. Therefore, we believe that a sensible measure UI for unique information should satisfy the following property:

$$UI(X : Y \setminus Z) \text{ only depends on } p, \kappa \text{ and } \mu. \qquad (*)$$

Although this condition seems to have not been considered before, many candidate measures of unique information satisfy this property; for example those defined in [10, 5]. In the following, we explore the consequences of assumption $(*)$.

Lemma 2. *Under assumption* $(*)$, *the shared information only depends on* p, κ *and* μ.

Proof. This follows from $SI(X:Y;Z) = MI(X:Y) - UI(X:Y \setminus Z)$. ☐

Let Δ be the set of all joint distributions of X, Y and Z. Fix $P \in \Delta$, and assume that the marginal distribution of X, denoted by p, has full support. Denote by κ and μ the stochastic matrices corresponding to the conditional distributions of Y and Z given X. Let

$$\Delta_P = \Big\{ Q \in \Delta_P : Q(X = x, Y = y) = P(X = x, Y = y)$$

$$\text{and } Q(X = x, Z = z) = P(X = x, Z = z) \text{ for all } x \in \mathcal{X}, y \in \mathcal{Y}, z \in \mathcal{Z} \Big\}$$

be the set of all joint distributions which have the same marginal distributions on the pairs (X, Y) and (X, Z), and let

$$\Delta_P^* = \big\{ Q \in \Delta_P : Q(x) > 0 \text{ for all } x \in \mathcal{X} \big\}$$

be the subset of distributions with full support. Lemma 2 says that, under assumption $(*)$, the functions $UI(X:Y \setminus Z)$, $UI(X:Z \setminus Y)$ and $SI(X:Y;Z)$ are constant on Δ_P^*, and only the function $CI(X:Y;Z)$ depends on the joint distribution $Q \in \Delta_P^*$. If we further assume continuity, the same statement holds true for all $Q \in \Delta_P$. To make clear that we now consider the synergistic information and the mutual information as a function of the joint distribution $Q \in \Delta$, we write $CI_Q(X:Y;Z)$ and $MI_Q(X:(Y,Z))$ in the following; and we omit this subscript, if these information theoretic quantities are computed with respect to the "true" joint distribution P.

Consider the following functions:

$$\widetilde{UI}(X:Z \setminus Y) = \min_{Q \in \Delta_P} MI_Q(X:Y|Z),$$

$$\widetilde{UI}(X:Y \setminus Z) = \min_{Q \in \Delta_P} MI_Q(X:Z|Y),$$

$$\widetilde{SI}(X:Y;Z) = \max_{Q \in \Delta_P} CoI_Q(X;Y;Z),$$

$$\widetilde{CI}(X:Y;Z) = MI(X:(Y,Z)) - \min_{Q \in \Delta_P} MI_Q(X:(Y,Z)).$$

Observe that these minima and maxima are well-defined, since the set Δ_P is compact and the mutual-informations and the co-information are continuous. The next lemma says that, under assumption $(*)$, the quantities $\widetilde{UI}$, $\widetilde{SI}$ and $\widetilde{CI}$ bound the unique, shared and synergistic information.

Lemma 3. *Let* $UI(X:Y \setminus Z)$, $UI(X:Z \setminus Y)$, $SI(X:Y;Z)$ *and* $CI(X:Y;Z)$ *be non-negative continuous functions on* Δ *satisfying equations* (1) *and* (2) *and assumption* $(*)$. *Then*

$$UI(X:Z \setminus Y) \leq \widetilde{UI}(X:Z \setminus Y),$$

$$UI(X:Y \setminus Z) \leq \widetilde{UI}(X:Y \setminus Z),$$

$$SI(X:Y;Z) \geq \widetilde{SI}(X:Y;Z),$$

$$CI(X:Y;Z) \geq \widetilde{CI}(X:Y;Z).$$

If $P \in \Delta$ and if there exists $Q \in \Delta_P$ such that $CI_Q(X : Y; Z) = 0$, then equality holds in all four inequalities. Conversely, if equality holds in one of the inequalities for a joint distribution $P \in \Delta$, then there exists $Q \in \Delta_P$ such that $CI_Q(X : Y; Z) = 0$.

Proof. Fix a joint distribution $P \in \Delta$. By Lemma 2, assumption $(*)$ and continuity, the functions $UI(X : Y \setminus Z)$, $UI(X : Z \setminus Y)$ and $SI(X : Y; Z)$ are constant on Δ_P, and only the function $CI(X : Y; Z)$ depends on the joint distribution $Q \in \Delta_P$. The decomposition (1) rewrites to

$$CI_Q(X : Y; Z) = MI_Q(X : (Y, Z)) - UI(X : Y \setminus Z) - UI(X : Z \setminus Y) - SI(X : Y; Z).$$

Using the non-negativity of synergistic information, this implies

$$UI(X : Y \setminus Z) + UI(X : Z \setminus Y) + SI(X : Y; Z) \leq \min_{Q \in \Delta_P} MI_Q(X : (Y, Z)).$$

In total, this shows

$$CI(X : Y; Z) \geq MI(X : (Y, Z)) - \min_{Q \in \Delta_P} MI_Q(X : (Y, Z)) = \widetilde{CI}(X : Y; Z).$$

The chain rule of mutual information says

$$MI_Q(X : (Y, Z)) = MI_Q(X : Z) + MI_Q(X : Y|Z).$$

Now, $Q \in \Delta_P$ implies $MI_Q(X : Z) = MI(X : Z)$, and therefore,

$$\widetilde{CI}(X : Y; Z) = MI(X : Y|Z) - \min_{Q \in \Delta_P} MI_Q(X : Y|Z).$$

Moreover,

$$MI_Q(X : Y|Z) = H_Q(X|Z) - H_Q(X|Y, Z),$$

where $H_Q(X|Z) = H(X|Z)$ for $Q \in \Delta_P$, and so

$$\widetilde{CI}(X : Y; Z) = \max_{Q \in \Delta_P} H_Q(X|Y, Z) - H(X|Y, Z).$$

By (3), the shared information satisfies

$$
\begin{aligned}
SI(X : Y; Z) &= CI(X : Y; Z) + MI(X : Y) + MI(X : Z) - MI(X : (Y, Z)) \\
&\geq \widetilde{CI}(X : Y; Z) + MI(X : Y) + MI(X : Z) - MI(X : (Y, Z)) \\
&= MI(X : Y) + MI(X : Z) - \min_{Q \in \Delta_P} MI_Q(X : (Y, Z)) \\
&= \max_{Q \in \Delta_P} (MI_Q(X : Y) + MI_Q(X : Z) - MI_Q(X : (Y, Z))) \\
&= \max_{Q \in \Delta_P} CoI_Q(X; Y; Z) = \widetilde{SI}(X : Y; Z).
\end{aligned}
$$

By (2), the unique information satisfies

$$UI(X : Y \setminus Z) = MI(X : Y) - SI(X : Y; Z)$$
$$\leq \min_{Q \in \Delta_P} (MI_Q(X : (Y, Z) - MI(X : Z))$$
$$= \min_{Q \in \Delta_P} (MI_Q(X : Y|Z)) = \widetilde{UI}(X : Y \setminus Z).$$

The inequality for $UI(X : Z \setminus Y)$ follows similarly.

If there exists $Q_0 \in \Delta_P$ such that $CI_{Q_0}(X : Y; Z) = 0$, then

$$0 = CI_{Q_0}(X : Y; Z) \geq \widetilde{CI}(X : Y; Z) = MI(X : (Y, Z)) - \min_{Q \in \Delta_P} MI_Q(X : (Y, Z)) \geq 0.$$

Hence, in this case, all inequalities are tight. Conversely, assume that one of the inequalities is tight for some $P \in \Delta$. The proof above shows that all four inequalities hold with equality. By assumption $(*)$, the functions $\widetilde{UI}$ and $\widetilde{SI}$ are constant on Δ_P. Therefore, the inequalities are tight for all $Q \in \Delta_P$. Now, if $Q_0 \in \Delta_P$ minimizes $MI_Q(X : (Y, Z))$ over Δ_P, then $CI_{Q_0}(X : Y; Z) = \widetilde{CI}_{Q_0}(X : Y; Z) = 0$. $\qquad \square$

The proof of Lemma 3 shows that the optimization problems defining $\widetilde{UI}$, $\widetilde{SI}$ and $\widetilde{CI}$ are in fact equivalent; that is, it suffices to solve one of them. Lemma 4 in Section 3 gives yet another formulation and shows that the solution is actually unique.

In the following, we interpret the functions $\widetilde{UI}$, $\widetilde{SI}$ and $\widetilde{CI}$ as measures of unique, shared and complementary information. Under assumption $(*)$, Lemma 3 says that choosing those measures is equivalent to saying that in each set Δ_P there exists a measure Q such that $CI_Q(X : Y; Z) = 0$. In other words, $\widetilde{UI}$, $\widetilde{SI}$ and $\widetilde{CI}$ are the only measures of unique, shared and complementary information that satisfy the following property:

It is not possible to decide whether or not there is synergistic information,

when only the marginal distributions of (X, Y) and (X, Z) are known. $\quad (**)$

For any other combination of measures different from $\widetilde{UI}$, $\widetilde{SI}$ and $\widetilde{CI}$ that satisfy assumption $(*)$ there are combinations of (p, μ, κ) such that the existence of non-vanishing complementary information can be deduced. Since complementary information should capture precisely the information that is carried by the joint dependencies between X, Y and Z we find assumption $(**)$ natural, and we consider this observation as evidence in favour of our interpretation of the functions $\widetilde{UI}$, $\widetilde{SI}$ and $\widetilde{CI}$.

3. Properties

3.1. Characterization and Positivity

The next lemma shows that the optimization problems involved in the definitions of $\widetilde{UI}$, $\widetilde{SI}$ and $\widetilde{CI}$ are easy to solve numerically, in the sense that they are convex optimization problems on convex sets. As always, theory is easier than practice, as discussed in Example 31 in the appendix.

Lemma 4. *Let $P \in \Delta$ and $Q_P \in \Delta_P$. The following conditions are equivalent:*

1. $MI_{Q_P}(X : Y|Z) = \min_{Q \in \Delta_P} MI_Q(X : Y|Z)$.

2. $MI_{Q_P}(X : Z|Y) = \min_{Q \in \Delta_P} MI_Q(X : Z|Y)$.

3. $MI_{Q_P}(X; (Y, Z)) = \min_{Q \in \Delta_P} MI_Q(X : (Y, Z))$.

4. $CoI_{Q_P}(X; Y; Z) = \max_{Q \in \Delta_P} CoI_Q(X; Y; Z)$.

5. $H_{Q_P}(X|Y, Z) = \max_{Q \in \Delta_P} H_Q(X|Y, Z)$.

Moreover, the functions $MI_Q(X : Y|Z)$, $MI_Q(X : Z|Y)$ and $MI_Q(X : (Y, Z))$ are convex on Δ_P; and $CoI_Q(X; Y; Z)$ and $H_Q(X|Y, Z)$ are concave. Therefore, for fixed $P \in \Delta$, the set of all $Q_P \in \Delta_P$ satisfying any of these conditions is convex.

Proof. The conditional entropy $H_Q(X|Y, Z)$ is a concave function on Δ; therefore, the set of maxima is convex. To show the equivalence of the five optimization problems and the convexity properties, it suffices to show that the difference of any two minimized functions and the sum of a minimized and a maximized function is constant on Δ_p. Except for $H_Q(X|Y, Z)$ this follows from the proof of Lemma 3. For $H_Q(X|Y, Z)$, this follows from the chain rule:

$$\begin{aligned} MI_Q(X : (Y, Z)) &= MI_P(X : Y) + MI_Q(X : Z|Y) \\ &= MI_P(X : Y) + H_P(X|Y) - H_Q(X|Y, Z) = H(X) - H_Q(X|Y, Z). \end{aligned} \quad \square$$

The optimization problems mentioned in Lemma 4 will be studied more closely in Appendix A.

Lemma 5 (Nonnegativity). *$\widetilde{UI}$, $\widetilde{SI}$ and $\widetilde{CI}$ are non-negative functions.*

Proof. $\widetilde{CI}$ is non-negative by definition. The functions $\widetilde{UI}$ are non-negative, because they are obtained by minimizing mutual-informations, which are non-negative.

Consider the real function

$$Q_0(X = x, Y = y, Z = z) = \begin{cases} \frac{P(X=x, Y=y)P(X=x, Z=z)}{P(X=x)}, & \text{if } P(X = x) > 0, \\ 0, & \text{else.} \end{cases}$$

It is easy to check $Q_0 \in \Delta_P$. Moreover, with respect to Q_0, the two random variables Y and Z are conditionally independent given X, that is, $MI_{Q_0}(Y : Z|X) = 0$. But this implies

$$CoI_{Q_0}(X; Y; Z) = MI_{Q_0}(Y : Z) - MI_{Q_0}(Y : Z|X) = MI_{Q_0}(Y : Z) \geq 0.$$

Therefore, $\widetilde{SI}(X : Y; Z) = \max_{Q \in \Delta_P} CoI_Q(X; Y; Z) \geq CoI_{Q_0}(X; Y; Z) \geq 0$, showing that $\widetilde{SI}$ is a non-negative function. $\square$

In general, the measure Q_0 constructed in the proof of Lemma 5 does not satisfy the conditions of Lemma 4.

9

3.2. Vanishing shared and unique information

In this section we study when $\widetilde{SI} = 0$ and when $\widetilde{UI} = 0$. In particular, in Corollary 7 we show that $\widetilde{UI}$ conforms with the operational idea put forward in Section 2.

Lemma 6. $\widetilde{UI}(X : Y \setminus Z)$ *vanishes if and only if there exists a row-stochastic matrix* $\lambda \in [0,1]^{\mathcal{Z} \times \mathcal{Y}}$ *such that*

$$P(X = x, Y = y) = \sum_{z \in \mathcal{Z}} P(X = x, Z = z)\lambda(z; y).$$

Proof. If $MI_Q(X : Y|Z) = 0$ for some $Q \in \Delta_P$, then X and Y are independent given Z with respect to Q. Therefore, there exists a stochastic matrix $\lambda \in [0,1]^{\mathcal{Z} \times \mathcal{Y}}$ satisfying

$$P(X = x, Y = y) = Q(X = x, Y = y) = \sum_{z \in \mathcal{Z}} Q(X = x, Z = z)\lambda(z; y)$$

$$= \sum_{z \in \mathcal{Z}} P(X = x, Z = z)\lambda(z; y).$$

Conversely, if such a matrix λ exists, then the equality

$$Q(X = x, Y = y, Z = z) = P(X = x, Z = z)\lambda(z; y)$$

defines a probability distribution Q which lies in Δ_P. Then

$$\widetilde{UI}(X : Y \setminus Z) \le MI_Q(X : Y|Z) = 0. \qquad \square$$

The last result can be translated into the language of our motivational Section 2 and says that $\widetilde{UI}$ is consistent with our operational idea of unique information:

Corollary 7. $\widetilde{UI}(X : Z \setminus Y) = 0$ *if and only if Z has no unique information about X with respect to Y (according to Definition 1).*

Proof. We need to show that decision problems can be solved with the channel κ at least as well as with the channel μ if and only if $\mu = \kappa\lambda$ for some stochastic matrix λ. This result is known as Blackwell's theorem [3]; see also [1]. $\qquad \square$

Corollary 8. *Suppose that $\mathcal{Y} = \mathcal{Z}$ and that the marginal distributions of the pairs (X, Y) and (X, Z) are identical. Then*

$$\widetilde{UI}(X : Y \setminus Z) = \widetilde{UI}(X : Z \setminus Y) = 0,$$
$$\widetilde{SI}(X : Y; Z) = MI(X : Y) = MI(X : Z),$$
$$\widetilde{CI}(X : Y; Z) = MI(X : Y|Z) = MI(X : Z|Y).$$

In particular, under assumption $(*)$, there is no unique information in this situation.

Proof. Apply Lemma 6 with the identity matrix in the place of λ. $\qquad \square$

Lemma 9. $\widetilde{SI}(X : Y; Z) = 0$ *if and only if* $MI_{Q_0}(Y : Z) = 0$, *where* $Q_0 \in \Delta$ *is the distribution constructed in the proof of Lemma 5.*

The proof of the lemma will be given in Appendix A.3, since it relies on some technical results from Appendix A, where Δ_P is characterized and the critical equations corresponding to the optimization problems in Lemma 4 are computed.

Corollary 10. *If both* $Y \perp\!\!\!\perp Z \,|\, X$ *and* $Y \perp\!\!\!\perp Z$, *then* $\widetilde{SI}(X : Y; Z) = 0$.

Proof. By assumption, $P = Q_0$. Thus the statement follows from Lemma 9. $\qquad \square$

3.3. The bivariate PI axioms

In [10], Williams and Beer proposed axioms that a measure of shared information should satisfy. We call these axioms the *PI axioms* after the partial information decomposition framework derived from these axioms in [10]. In fact, the PI axioms apply to a measure of shared information that is defined for arbitrarily many random variables, while our function $\widetilde{SI}$ only measures the shared information of two random variables (about a third variable). The PI axioms are as follows:

1. The shared information of $Y_1, \ldots, Y_n$ about X is symmetric under permutations of $Y_1, \ldots, Y_n$. (symmetry)

2. The shared information of Y_1 about X is equal to $MI(X : Y_1)$. (self-redundancy)

3. The shared information of $Y_1, \ldots, Y_n$ about X is less than the shared information of $Y_1, \ldots, Y_{n-1}$ about X, with equality if Y_{n-1} is a function of Y_n. (monotonicity)

Any measure $\widetilde{SI}$ of bivariate shared information that is consistent with the PI axioms must obviously satisfy the following two properties, which we call the *bivariate PI axioms:*

A) $\widetilde{SI}(X : Y; Z) = \widetilde{SI}(X : Z; Y)$. (symmetry)

B) $\widetilde{SI}(X : Y; Z) \leq MI(X : Y)$, with equality if Z is a function of Y.

(bivariate monotonicity)

We do not claim that any function $\widetilde{SI}$ that satisfies A) and B) can be extended to a measure of multivariate shared information satisfying the PI axioms. In fact, such a claim is false, and as discussed in Section 6, our bivariate function $\widetilde{SI}$ is not extendable in this way.

The following two lemmas show that $\widetilde{SI}$ satisfies the bivariate PI axioms, and they show corresponding properties of $\widetilde{UI}$ and $\widetilde{CI}$.

Lemma 11 (Symmetry)**.**

$$\widetilde{SI}(X : Y; Z) = \widetilde{SI}(X : Z; Y),$$
$$\widetilde{CI}(X : Y; Z) = \widetilde{CI}(X : Z; Y),$$
$$MI(X : Z) + \widetilde{UI}(X : Y \setminus Z) = MI(X : Y) + \widetilde{UI}(X : Z \setminus Y).$$

Proof. The first two equalities follow since the definitions of $\widetilde{SI}$ and $\widetilde{CI}$ are symmetric in Y and Z. The third equality follows from

$$MI(X:Z) + \widetilde{UI}(X:Y \setminus Z) = \min_{Q \in \Delta_P} \left(MI(X:Z) + MI_Q(X:Y|Z) \right)$$
$$= \min_{Q \in \Delta_P} \left(MI(X:Y) + MI_Q(X:Z|Y) \right) = MI(X:Y) + \widetilde{UI}(X:Z \setminus Y),$$

where the chain rule of mutual information was used. $\qquad\square$

The third equality from Lemma 11 is the consistency condition (4).

The following lemma is the inequality condition of the monotonicity axiom.

Lemma 12 (Bounds).

$$\widetilde{SI}(X:Y;Z) \leq MI(X:Y),$$
$$\widetilde{CI}(X:Y;Z) \leq MI(X:Y|Z),$$
$$\widetilde{UI}(X:Y \setminus Z) \geq MI(X:Y) - MI(X:Z).$$

Proof. The first inequality follows from

$$\widetilde{SI}(X:Y;Z) = \max_{Q \in \Delta_P} CoI_Q(X;Y;Z)$$
$$= \max_{Q \in \Delta_P} \left(MI(X:Y) - MI_Q(X:Y|Z) \right) \leq MI(X:Y),$$

the second from

$$\widetilde{CI}(X:Y;Z) = MI(X:Y|Z) - \min_{Q \in \Delta_P} MI_Q(X:Y|Z),$$

using the chain rule again. The last inequality follows from the first inequality, equality (2) and the symmetry of Lemma 11. $\qquad\square$

To finish the study of the bivariate PI axioms, only the equality condition in the monotonicity axiom is missing. We show that $\widetilde{SI}$ satisfies $\widetilde{SI}(X:Y;Z) = MI(X:Y)$ not only if Z is a deterministic function of Y, but also more generally, when Z is independent of X given Y. In this case, Z can be interpreted as a stochastic function of Y, independent of X.

Lemma 13. *If X is independent of Z given Y, then P solves the optimization problems of Lemma 4. In particular,*

$$\widetilde{UI}(X:Z \setminus Y) = MI(X:Y|Z),$$
$$\widetilde{UI}(X:Y \setminus Z) = 0,$$
$$\widetilde{SI}(X:Y;Z) = MI(X;Z),$$
$$\widetilde{CI}(X:Y;Z) = 0.$$

Proof. If X is independent of Z given Y, then

$$MI(X:Z|Y) = 0 \le \min_{Q \in \Delta_P} MI_Q(X:Y|Z),$$

so P minimizes $MI_Q(X:Y|Z)$ over Δ_P. $\qquad\square$

Remark 14. In fact, Lemma 13 can be generalized as follows: In any binary information decomposition, equations (1) and (2) and the chain rule imply

$$MI(X:Z|Y) = MI(X:(Y,Z)) - MI(X:Y) = UI(X:Z \setminus Y) + CI(X:Y;Z).$$

Therefore, if $MI(X:Z|Y) = 0$, then $UI(X:Z \setminus Y) = 0 = CI(X:Y;Z)$.

3.4. Probability distributions with structure

In this section we compute the values of $\widetilde{SI}$, $\widetilde{CI}$ and $\widetilde{UI}$ for probability distributions with special structure. If two of the variables are identical, then $\widetilde{CI} = 0$ as a consequence of Lemma 13 (see Corollaries 15 and 16). When $X = (Y, Z)$, then the same is true (Proposition 18). Moreover, in this case, $\widetilde{SI}((Y,Z):Y;Z) = MI(Y:Z)$. This equation has been postulated as an additional axiom, called *identity axiom*, in [5].

Corollary 15.

$$\widetilde{CI}(X:Y;Y) = 0,$$
$$\widetilde{SI}(X:Y;Y) = CoI(X;Y;Y) = MI(X:Y),$$
$$\widetilde{UI}(X:Y \setminus Y) = 0.$$

Proof. If $Y = Z$, then X is independent of Z given Y. $\qquad\square$

Corollary 16.

$$\widetilde{CI}(X:X;Z) = 0,$$
$$\widetilde{SI}(X:X;Z) = CoI(X;X;Z) = MI(X:Z) - MI(X:Z|X) = MI(X:Z),$$
$$\widetilde{UI}(X:X \setminus Z) = MI(X:X|Z) = H(X|Z),$$
$$\widetilde{UI}(X:Z \setminus X) = MI(X:Z|X) = 0.$$

Proof. If $X = Y$, then X is independent of Z given Y. $\qquad\square$

Remark 17. Remark 14 implies that Corollaries 15 and 16 hold for any bivariate information decomposition.

Proposition 18 (Identity property). *Suppose that $\mathcal{X} = \mathcal{Y} \times \mathcal{Z}$, and $X = (Y,Z)$. Then P solves the optimization problems of Lemma 4. In particular,*

$$\widetilde{CI}((Y,Z):Y;Z) = 0,$$
$$\widetilde{SI}((Y,Z):Y;Z) = MI(Y:Z),$$
$$\widetilde{UI}((Y,Z):Y \setminus Z) = H(Y|Z),$$
$$\widetilde{UI}((Y,Z):Z \setminus Y) = H(Z|Y).$$

Proof. If $X = (Y, Z)$, then, by Corollary 28 in the Appendix, $\Delta_P = \{P\}$, and therefore

$$\widetilde{SI}((Y,Z) : Y; Z) = MI((Y,Z) : Y) - MI((Y,Z) : Y|Z)$$
$$= H(Y) - H(Y|Z) = MI(Y : Z)$$

and

$$\widetilde{UI}((Y : Z) : Y \setminus Z) = MI((Y,Z) : Y|Z) = H(Y|Z),$$

and similarly for $\widetilde{UI}((Y : Z) : Y \setminus Z)$. $\qquad\qquad\qquad\qquad\qquad\square$

The following Lemma shows that the quantities $\widetilde{UI}$, $\widetilde{SI}$ and $\widetilde{CI}$ are additive when considering systems that can be decomposed into independent subsystems.

Lemma 19. *Let $X_1, X_2, Y_1, Y_2, Z_1, Z_2$ be random variables such that (X_1, Y_1, Z_1) is independent of (X_2, Y_2, Z_2). Then*

$$\widetilde{SI}((X_1, X_2) : (Y_1, Y_2); (Z_1, Z_2)) = \widetilde{SI}(X_1 : Y_1; Z_1) + \widetilde{SI}(X_1 : Y_1; Z_1),$$
$$\widetilde{CI}((X_1, X_2) : (Y_1, Y_2); (Z_1, Z_2)) = \widetilde{CI}(X_1 : Y_1; Z_1) + \widetilde{CI}(X_1 : Y_1; Z_1),$$
$$\widetilde{UI}((X_1, X_2) : (Y_1, Y_2) \setminus (Z_1, Z_2)) = \widetilde{UI}(X_1 : Y_1 \setminus Z_1) + \widetilde{UI}(X_1 : Y_1 \setminus Z_1),$$
$$\widetilde{UI}((X_1, X_2) : (Z_1, Z_2) \setminus (Y_1, Y_2)) = \widetilde{UI}(X_1 : Z_1 \setminus Y_1) + \widetilde{UI}(X_1 : Z_1 \setminus Y_1).$$

The proof of the last lemma is given in Appendix A.3.

4. Comparison with other measures

In this section we compare our information decomposition using $\widetilde{UI}$, $\widetilde{SI}$ and $\widetilde{CI}$ with similar functions proposed in other papers; in particular, the function $I_{\min}$ of [10] and the bivariate redundancy measure I_{red} of [5]. We do not repeat their definitions here, since they are rather technical.

The first observation is that both I_{red} and $I_{\min}$ satisfy assumption $(*)$. Therefore, $I_{\text{red}} \geq \widetilde{SI}$ and $I_{\min} \geq \widetilde{SI}$. According to [5], $I_{\min}$ tends to be larger than I_{red}, but there are some exceptions.

It is easy to find examples where $I_{\min}$ is unreasonably large [5, 2]. It is much more difficult to distinguish I_{red} and $\widetilde{SI}$. In fact, in many special cases the two measures I_{red} and $\widetilde{SI}$ agree, as the following results show.

Theorem 20. $I_{red}(X : Y; Z) = 0$ *if and only if* $\widetilde{SI}(X : Y; Z) = 0$.

The proof of the theorem builds on the following lemma:

Lemma 21. *If both $Y \perp\!\!\!\perp Z \,|\, X$ and $Y \perp\!\!\!\perp Z$, then $I_{red}(X : Y; Z) = 0$.*

The proof of the lemma is deferred to Appendix A.3.

Proof of Theorem 20. By Lemma 3, if $I_{\mathrm{red}}(X : Y; Z) = 0$, then $\widetilde{SI}(X : Y; Z) = 0$. Now assume that $\widetilde{SI}(X : Y; Z) = 0$. Since both $\widetilde{SI}$ and I_{red} are constant on Δ_P, we may assume that $P = Q_0$; that is, we may assume that $Y \perp\!\!\!\perp Z \,|\, X$. Then $Y \perp\!\!\!\perp Z$ by Lemma 9. Therefore, Lemma 21 implies that $I_{\mathrm{red}}(X : Y; Z) = 0$. $\qquad\square$

Denote by UI_{red} the unique information defined from I_{red} and (2). Then:

Theorem 22. $UI_{red}(X : Y \setminus Z) = 0$ *if and only if* $\widetilde{UI}(X : Y \setminus Z) = 0$

Proof. By Lemma 3, if $\widetilde{UI}$ vanishes, then so does UI_{red}. Conversely, $UI_{\mathrm{red}}(X : Y \setminus Z) = 0$ if and only if $I_{\mathrm{red}}(X : Y; Z) = MI(X : Y)$. By (20) in [5], this is equivalent to $p(x|y) = p_{y \setminus Z}(x)$ for all x, y. In this case, $p(x|y) = \sum_z p(x|z)\lambda(z; y)$ for some $\lambda(z; y)$ with $\sum_z \lambda(z; y) = 1$. Hence, Lemma 6 implies that $\widetilde{UI}(X : Y \setminus Z) = 0$. $\qquad\square$

Theorem 22 implies that I_{red} does not contradict our operational ideas introduced in Section 2.

Corollary 23. *Suppose that one of the following conditions is satisfied:*

1. *X is independent of Y given Z.*

2. *X is independent of Z given Y.*

3. *$\mathcal{X} = \mathcal{Y} \times \mathcal{Z}$, and $X = (Y, Z)$.*

Then $I_{red}(X : Y; Z) = \widetilde{SI}(X : Y; Z)$.

Proof. If X is independent of Z given Y, then, by Remark 14, for any binary information decomposition, $UI(X : Z \setminus Y) = 0$. In particular, $\widetilde{UI}(X : Z \setminus Y) = 0 = UI_{\mathrm{red}}(X : Z \setminus Y)$ (compare also Lemma 13 and Theorem 22). Therefore, $I_{\mathrm{red}}(X : Y; Z) = \widetilde{SI}(X : Y; Z)$. If $\mathcal{X} = \mathcal{Y} \times \mathcal{Z}$ and $X = (Y, Z)$, then $\widetilde{SI}(X : Y; Z) = MI(Y : Z) = I_{\mathrm{red}}(X : Y; Z)$ by Proposition 18 and the identity axiom in [5]. $\qquad\square$

Corollary 24. *If the two pairs (X, Y) and (X, Z) have the same marginal distribution, then $I_{red}(X : Y; Z) = \widetilde{SI}(X : Y; Z)$.*

Proof. In this case, $\widetilde{UI}(X : Y; Z) = 0 = UI_{\mathrm{red}}(X : Y; Z)$. $\qquad\square$

Although $\widetilde{SI}$ and I_{red} often agree, they are different functions. An example where $\widetilde{SI}$ and I_{red} have different values is the dice example given at the end of the next section. In particular, it follows that I_{red} does not satisfy property $(**)$.

Example	$\widetilde{CI}$	$\widetilde{SI}$	$I_{\min}$	Note
RDN	0	1	1	$X = Y = Z$ uniformly distributed
UNQ	1	0	1	$X = (Y, Z)$, Y, Z i.i.d.
XOR	1	0	0	$X = Y \, \mathrm{XOR} \, Z$, Y, Z i.i.d.
AND	1/2	0.311	0.311	$X = Y \, \mathrm{AND} \, Z$, Y, Z i.i.d.
RDNXOR	1	1	1	$X = (Y_1 \, \mathrm{XOR} \, Z_1, W)$, $Y = (Y_1, W)$, $Z = (Z_1, W)$, Y_1, Z_1, W i.i.d.
RDNUNQXOR	1	1	2	$X = (Y_1 \, \mathrm{XOR} \, Z_1, (Y_2, Z_2), W)$, $Y = (Y_1, Y_2, W)$, $Z = (Z_1, Z_2, W)$, Y_1, Y_2, Z_1, Z_2, W i.i.d.
XORAND	1	1/2	1/2	$X = (Y \, \mathrm{XOR} \, Z, Y \, \mathrm{AND} \, Z)$, Y, Z i.i.d.
Copy	0	$MI(X : Y)$	1	$X = (Y, Z)$

Table 1: The value of $\widetilde{SI}$ in some examples. The note is a short explanation of the example; see [5] for the details.
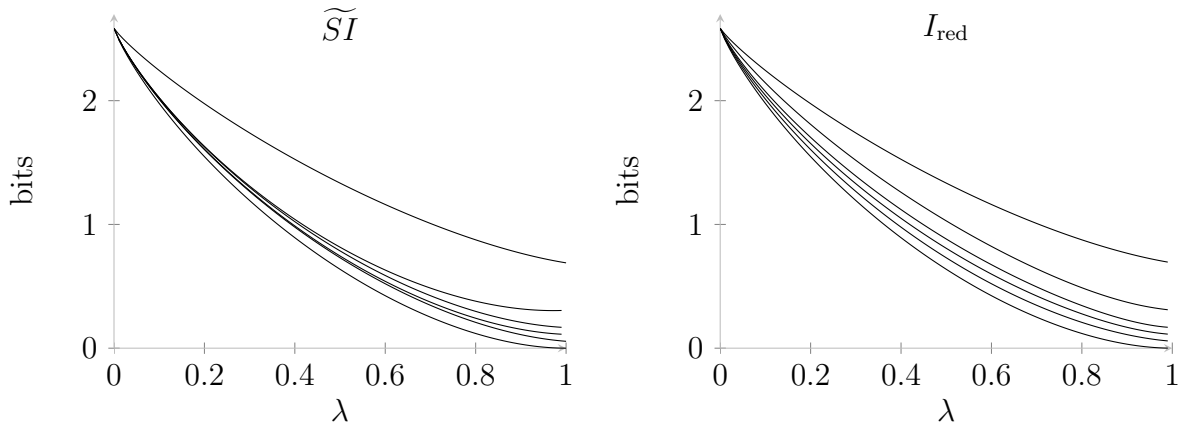

Figure 1: The shared information measures $\widetilde{SI}$ or I_{red} in the dice example depending on the correlation parameter λ. The summation parameter α varies from 1 (uppermost line) to 6 (lowest line).

5. Examples

Table 1 contains the values of $\widetilde{CI}$ and $\widetilde{SI}$ for some paradigmatic examples. The list of examples is taken from [5]; see also [4]. In all these examples, $\widetilde{SI}$ agrees with I_{red}. In particular, in these examples the values of $\widetilde{SI}$ agree with the intuitively plausible values called "expected values" in [5].

As a more complicated example we treated the following system with two parameters $\lambda \in [0, 1]$, $\alpha \in \{1, 2, 3, 4, 5, 6\}$, also proposed by [5]. Let Y and Z be two dices, and define $X = Y + \alpha Z$. To change the degree of dependence of the two dices, assume that they are distributed according to

$$P(Y = i, Z = j) = \frac{\lambda}{36} + (1 - \lambda)\frac{\delta_{i,j}}{6}.$$

For $\lambda = 0$ the two dice are completely correlated, while for $\lambda = 1$ they are independent. The resulting shared information is shown in Figure 1. As a comparison, we reproduce Figure 8 from [5] showing the function I_{red} in the same example. In fact, for $\alpha = 1$, $\alpha = 5$ and $\alpha = 6$ the two functions agree. Moreover, they agree for $\lambda = 0$ and $\lambda = 1$. In all other cases, $\widetilde{SI} \leq I_{\text{red}}$, in agreement with Lemma 3. For $\alpha = 1$ and $\alpha = 6$ and $\lambda = 0$ the fact that $I_{\text{red}} = \widetilde{SI}$ follows from the results in Section 4; in the other cases we do not know a simple reason for this coincidence.

It is interesting to note that for small λ and $\alpha > 1$ the function $\widetilde{SI}$ depends only weakly on α. In contrast, the dependence of I_{red} on α is stronger. At the moment we do not have an argument that tells us which of these two behaviours is more intuitive.

6. Outlook

We defined a decomposition of the mutual information $MI(X : (Y, Z))$ of a random variable X with a pair of random variables (Y, Z) into non-negative terms which have an interpretation in terms of shared information, unique information and synergistic information. We have shown that the quantities $\widetilde{SI}$, $\widetilde{CI}$ and $\widetilde{UI}$ have many properties that such a decomposition should intuitively fulfil; among them the PI axioms and the identity axiom. It is a natural question whether the same can be done when further random variables are added to the system.

The first question in this context is how the decomposition of $MI(X : Y_1, \ldots, Y_n)$ should look like. How many terms do we need? In the bivariate case $n = 2$, many people agree that shared, unique and synergistic information should provide a complete decomposition (but it may well be worth to look for other types of decompositions). For $n > 2$, there is no universal agreement of this kind.

Williams and Beer proposed a framework that suggests to construct an information decomposition only in terms of shared information [10]. Their ideas naturally lead to a decomposition according to a lattice, called PI lattice. For example, in this framework, $MI(X : Y_1, Y_2, Y_3)$ has to be decomposed into 18 terms with well-defined interpretation. The approach is very appealing, since it is only based on very natural properties of shared

information (the PI axioms) and the idea that all information can be "localized," in the sense that, in an information decomposition, it suffices to classify information according to "who knows what," that is, which information is shared by which subsystems.

Unfortunately, as shown in [2], our function $\widetilde{SI}$ cannot be generalized to the case $n = 3$ in the framework of the PI lattice. The problem is that the identity axiom is incompatible with a non-negative decomposition according to the PI lattice.

Even though we currently cannot extend our decomposition to $n > 2$, our bivariate decomposition can be useful for the analysis of larger systems consisting of more than two parts. For example, the quantity

$$\widetilde{UI}(X : Y_i \setminus (Y_1, \ldots, Y_{i-1}, Y_{i+1}, \ldots, Y_n))$$

can still be interpreted as the unique information of Y_i about X with respect to all other variables, and it can be used to assess the value of the ith variable, when synergistic contributions can be ignored. Furthermore the measure has the intuitive property that the unique information cannot grow when additional variables are taken into account:

Lemma 25. $\widetilde{UI}(X : Y \setminus (Z_1, \ldots, Z_k)) \geq \widetilde{UI}(X : Y \setminus (Z_1, \ldots, Z_{k+1}))$.

Proof. Let P^k be the joint distribution of $X, Y, Z_1, \ldots, Z_k$, and let P^{k+1} be the joint distribution of $X, Y, Z_1, \ldots, Z_k, Z_{k+1}$. By definition, P^k is a marginal of P^{k+1}. For any $Q \in \Delta_{P^k}$, the distribution Q' defined by

$$Q'(x, y, z_1, \ldots, z_k, z_{k+1}) := \begin{cases} \frac{Q(x, y, z_1, \ldots, z_k) P^{k+1}(x, z_1, \ldots, z_k, z_{k+1})}{P^k(x, z_1, \ldots, z_k)}, & \text{if } P^k(x, z_1, \ldots, z_k) > 0, \\ 0, & \text{else,} \end{cases}$$

lies in $\Delta_{P^{k+1}}$. Moreover, Q is the $(X, Y, Z_1, \ldots, Z_k)$-marginal of Q', and Z_{k+1} is independent of Y given $X, Z_1, \ldots, Z_k$. Therefore,

$$\begin{aligned} MI_{Q'}(X : Y | Z_1, \ldots, Z_k, Z_{k+1}) &\leq MI_{Q'}(X, Z_{k+1} : Y | Z_1, \ldots, Z_k) \\ &= MI_{Q'}(X : Y | Z_1, \ldots, Z_k) + MI_{Q'}(Z_{k+1} : Y | X, Z_1, \ldots, Z_k) \\ &\leq MI_{Q'}(X : Y | Z_1, \ldots, Z_k) = MI_Q(X : Y | Z_1, \ldots, Z_k). \end{aligned}$$

The statement now follows by taking the minimum over $Q \in \Delta_{P^k}$. $\qquad\square$

Thus, we believe that our measure, which is well-motivated in operational terms, can serve as a good starting point towards a general decomposition of multi-variate information.

A. Computing $\widetilde{UI}$, $\widetilde{SI}$ and $\widetilde{CI}$

A.1. The optimization domain Δ_P

By Lemma 4, to compute the quantities $\widetilde{UI}$, $\widetilde{SI}$ and $\widetilde{CI}$, we need to solve a convex optimization problem. In this section we study some aspects of this problem.

First we describe Δ_P. Let A be the map $\Delta \to \Delta(\mathcal{X} \times \mathcal{Y}) \times \Delta(\mathcal{X} \times \mathcal{Z})$ that takes a joint probability distribution of X, Y and Z and computes the marginal distributions of the pairs (X, Y) and (X, Z). Then A is a linear map, and $\Delta_P = (P + \ker A) \cap \Delta$. In particular, Δ_P is the intersection of an affine space and a simplex; hence Δ_P is a polytope.

The matrix describing A (and denoted by the same symbol in the following) is a well-studied object. For example, A describes the graphical model associated with the graph Y—X—Z. The columns of A define a polytope, called *marginal polytope*. Moreover, the kernel of A is known: Let $\delta_{x,y,z} \in \mathbb{R}^{\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}}$ be the characteristic function of the point $(x, y, z) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$, and let

$$\gamma_{x;y,y';z,z'} = \delta_{x,y,z} + \delta_{x,y',z'} - \delta_{x,y',z} - \delta_{x,y,z'}.$$

Lemma 26. *The defect of A (that is, the dimension of $\ker A$) is $|\mathcal{X}|(|\mathcal{Y}| - 1)(|\mathcal{Z}| - 1)$.*

- *The functions $\gamma_{x;y,y';z,z'}$ for all $x \in \mathcal{X}$, $y, y' \in \mathcal{Y}$ and $z, z' \in \mathcal{Z}$ span $\ker A$.*

- *For any fixed $y_0 \in \mathcal{Y}$, $z_0 \in \mathcal{Z}$, the functions $\gamma_{x;y_0,y;z_0,z}$ for all $x \in \mathcal{X}$, $y \in \mathcal{Y} \setminus \{y_0\}$ and $z \in \mathcal{Z} \setminus \{z_0\}$ form a basis of $\ker A$.*

Proof. See [6]. $\qquad\square$

The vectors $\gamma_{x;y,y';z,z'}$ for different values of $x \in \mathcal{X}$ have disjoint supports. As the next lemma shows, this can be used to write Δ_P as a Cartesian product of simpler polytopes. Unfortunately, the function $MI(X : (Y, Z))$ does not respect this product structures. In fact, the diagonal directions are important (see Example 31 below).

Lemma 27. *Let $P \in \Delta$. For all $x \in \mathcal{X}$ with $P(x) > 0$ denote by*

$$\Delta_{P,x} = \Big\{ Q \in \Delta(\mathcal{Y} \times \mathcal{Z}) : Q(Y = y) = P(Y = y | X = x)$$
$$\text{and } Q(Z = z) = P(Z = z | X = x) \Big\}$$

the set of joint distributions of Y and Z such that the marginal distributions of Y and Z agree with the conditional distributions of Y and Z given $X = x$. Then the map $\pi_P : \Delta_P \mapsto \times_{x \in \mathcal{X}:P(x)>0} \Delta_{P,x}$ that maps each $Q \in \Delta_P$ to the family $(Q(\cdot | X = x))_{x \in \mathcal{X}:P(x)>0}$ of conditional distributions of Y and Z given $X = x$ for those $x \in \mathcal{X}$ with $P(X = x) > 0$ is a linear bijection.

Proof. The image of π_P is contained in $\times_{x \in \mathcal{X}:P(x)>0} \Delta_{P,x}$ by definition of Δ_P. The relation
$$Q(X = x, Y = y, Z = z) = P(X = x)Q(Y = y, Z = z | X = x)$$
shows that π_P is injective and surjective. Since π_P is in fact a linear map, the domain and the codomain of π_P are affinely equivalent. $\qquad\square$

Each Cartesian factor $\Delta_{P,x}$ of Δ_P is a fibre polytope of the independence model.

Corollary 28. *If $X = (Y, Z)$, then $\Delta_P = \{P\}$.*

Proof. By assumption, both conditional probability distributions $P(Y|X = x)$ and $P(Z|X = x)$ are point measures. Therefore, each factor $\Delta_{P,x}$ consists of a single point; namely the conditional distribution $P(Y, Z|X = x)$ of Y and Z given X. Hence, Δ_P is a singleton. $\qquad\square$

A.2. The critical equations

Lemma 29. *The derivative of $MI_Q(X : (Y, Z))$ in the direction $\gamma_{x;y,y';z,z'}$ is*

$$\log \frac{Q(x, y, z)Q(x, y', z')}{Q(x, y', z)Q(x, y, z')} \frac{Q(y', z)Q(y, z')}{Q(y, z)Q(y', z')}.$$

Therefore, Q solves the optimization problems of Lemma 4 if and only if

$$\log \frac{Q(x, y, z)Q(x, y', z')}{Q(x, y', z)Q(x, y, z')} \frac{Q(y', z)Q(y, z')}{Q(y, z)Q(y', z')} \geq 0 \tag{6}$$

for all x, y, y', z, z' with $Q + \epsilon\gamma_{x;y,y';z,z'} \in \Delta_P$ for $\epsilon > 0$ small enough.

Proof. The proof is by direct computation. $\qquad\square$

Example 30 (The AND-example). Consider the binary case $\mathcal{X} = \mathcal{Y} = \mathcal{Z} = \{0, 1\}$, assume that Y and Z are independent and uniformly distributed, and suppose that $X = Y \text{ AND } Z$. The underlying distribution P is uniformly distributed on the four states $\{000, 001, 010, 111\}$. In this case, $\Delta_{P,1} = \{\delta_{Y=1,Z=1}\}$ is a singleton, and $\Delta_{P,0}$ consists of all probability distributions Q of the form

$$Q(Y = y, Z = z) = \begin{cases} \frac{1}{3} + \alpha', & \text{if } (y, z) = (0, 0), \\ \frac{1}{3} - \alpha', & \text{if } (y, z) = (0, 1), \\ \frac{1}{3} - \alpha', & \text{if } (y, z) = (1, 0), \\ \alpha', & \text{if } (y, z) = (1, 1), \end{cases}$$

for some $0 \leq \alpha' \leq \frac{1}{3}$. Therefore, Δ_P is a one-dimensional polytope consisting of all probability distributions of the form

$$Q_\alpha(Y = y, Z = z) = \begin{cases} \frac{1}{4} + \alpha, & \text{if } (x, y, z) = (0, 0, 0), \\ \frac{1}{4} - \alpha, & \text{if } (x, y, z) = (0, 0, 1), \\ \frac{1}{4} - \alpha, & \text{if } (x, y, z) = (0, 1, 0), \\ \alpha, & \text{if } (x, y, z) = (0, 1, 1), \\ \frac{1}{4}, & \text{if } (x, y, z) = (1, 1, 1), \\ 0, & \text{else}, \end{cases}$$

for some $0 \leq \alpha \leq \frac{1}{4}$. To compute the minimum of $MI_{Q_\alpha}(X : (Y, Z))$ over Δ_P, we compute the derivative with respect to α (which equals the directional derivative of $MI_Q(X : (Y, Z))$ in the direction $\gamma_{0;0,1;0,1}$ at Q_α) and obtain:

$$\log \frac{(\frac{1}{4} + \alpha)\alpha}{(\frac{1}{4} - \alpha)^2} \frac{(\frac{1}{4} - \alpha)^2}{(\frac{1}{4} + \alpha)^2} = \log \frac{\alpha}{\frac{1}{4} + \alpha}.$$

Since $\frac{\alpha}{\frac{1}{4}+\alpha} < 1$ for all $\alpha > 0$, the function $MI_{Q_\alpha}(X : (Y, Z))$ has a unique minimum at $\alpha = \frac{1}{4}$. Therefore,

$$\widetilde{UI}(X : Z \setminus Y) = MI_{Q_{1/4}}(X : Y|Z) = 0 = \widetilde{UI}(X : Y \setminus Z),$$

$$\widetilde{SI}(X : Y; Z) = CoI_{Q_{1/4}}(X; Y; Z) = MI_{Q_{1/4}}(X : Y) = \frac{3}{4} \log \frac{4}{3},$$

$$\widetilde{CI}(X : Y; Z) = MI(X : (Y, Z)) - MI_{Q_{1/4}}(X : (Y, Z)) = \frac{1}{2} \log 2.$$

In other words, in the AND-example there is no unique information, but only shared and synergistic information. This follows, of course, also from Corollary 8.

Example 31. The optimization problems in Lemma 4 can be very ill-conditioned, in the sense that there are directions in which the function varies fast, and other directions in which the function varies slowly. As an example, consider the example where P is the distribution of three i.i.d. uniform binary random variables. In this case, Δ_P is a square. Figure 2 contains a heat map of the function CoI_Q on Δ_P, where Δ_P is parametrized by

$$Q(a, b) = P + a\gamma_{0;0,1;0,1} + b\gamma_{1;0,1;0,1}, \qquad -\frac{1}{8} \leq a \leq \frac{1}{8}, -\frac{1}{8} \leq b \leq \frac{1}{8}.$$

Clearly, the function varies very little along one of the diagonals. In fact, along this diagonal, X is independent of (Y, Z), corresponding to a very low synergy.

Although in this case the optimising probability distribution Q_P is unique, it can be difficult to find. For example, Mathematica's function `FindMinimum` does not always find the true optimum out of the box (apparently, `FindMinimum` cannot make use of the convex structure in the presence of constraints) [11].

A.3. Technical proofs

Proof of Lemma 9. Since $\widetilde{SI}(X : Y; Z) \geq CoI_{Q_0}(X; Y; Z) \geq 0$, if $\widetilde{SI}(X : Y; Z) = 0$, then

$$0 = CoI_{Q_0}(X; Y; Z) = MI_{Q_0}(Y : Z) - MI_{Q_0}(Y : Z|X) = MI_{Q_0}(Y : Z).$$

To show that $MI_{Q_0}(Y : Z) = 0$ is also sufficient, observe that

$$Q_0(x, y, z)Q_0(x, y', z') = Q_0(x, y, z')Q_0(x, y', z),$$

by construction of Q_0, and that
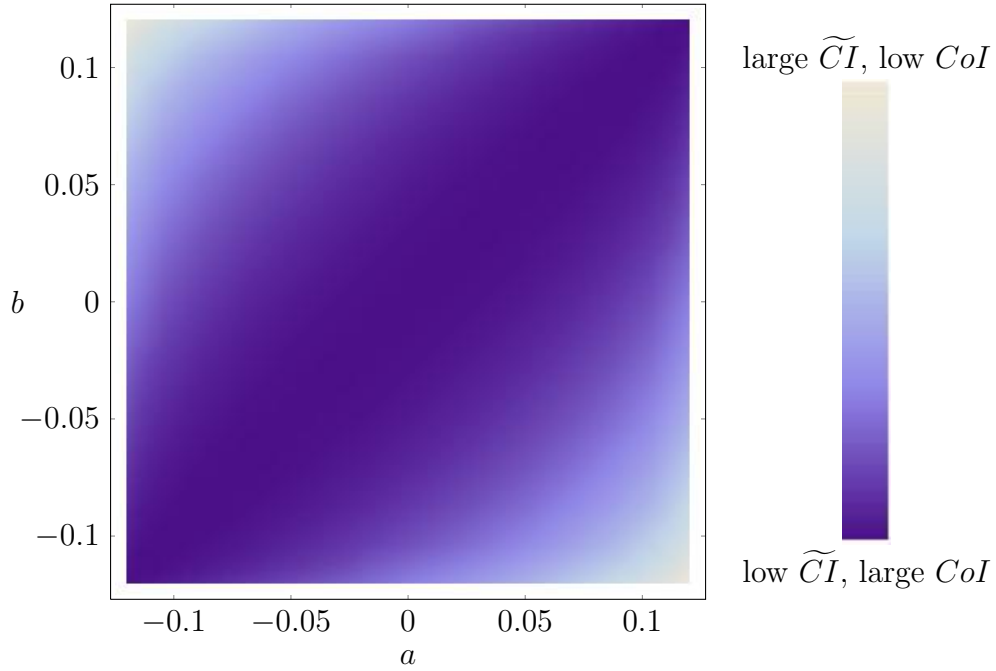
$$Q_0(y, z)Q_0(y', z') = Q_0(y, z')Q_0(y', z)$$

21

Figure 2: The function CoI_Q in Example 31 (figure created with the help of Mathematica [11]). Darker colours indicate larger values of CoI_Q. In this example, Δ_P is a square. The uniform distribution lies at the centre of this square and is the maximum of CoI_Q. In the two dark corners, X is independent of Y and Z, and either $Y = Z$ or $Y = \neg Z$. In the two light corners, Y and Z are independent, and either $X = Y\,\mathrm{XOR}\,Z$ or $X = \neg(Y\,\mathrm{XOR}\,Z)$.

by the assumption that $MI_{Q_0}(Y:Z) = 0$. Therefore, by Lemma 29, all partial derivatives vanish at Q_0. Therefore, Q_0 solves the optimization problems in Lemma 4, and $\widetilde{SI}(X:Y;Z) = CoI_{Q_0}(X;Y;Z) = 0$. $\qquad\square$

Proof of Lemma 19. Let Q_1 and Q_2 be solutions of the optimization problems from Lemma 4 for (X_1, Y_1, Z_1) and (X_2, Y_2, Z_2) in the place of (X, Y, Z), respectively. Consider the probability distribution Q defined by

$$Q(x_1, x_2, y_1, y_2, z_1, z_2) = Q_1(x_1, y_1, z_1)Q_2(x_2, y_2, z_2).$$

Since (X_1, Y_1, Z_1) is independent of (X_2, Y_2, Z_2) (under P), $Q \in \Delta_P$. We show that Q solves the optimization problems from Lemma 4 for $X = (X_1, X_2)$, $Y = (Y_1, Y_2)$ and $Z = (Z_1, Z_2)$. We use the notation from Appendix A.

If $Q + \epsilon\gamma_{x_1 x_2; y_1 y_2, y_1' y_2'; z_1 z_2, z_1' z_2'} \in \Delta_Q$, then

$$Q_1 + \epsilon\gamma_{x_1; y_1, y_1'; z_1, z_1'} \in \Delta_{Q_1} \quad \text{and} \quad Q_2 + \epsilon\gamma_{x_2; y_2, y_2'; z_2, z_2'} \in \Delta_{Q_2}.$$

Therefore, by Lemma 29,

$$
\begin{aligned}
\log &\frac{Q(x_1 x_2, y_1 y_2, z_1 z_2)Q(x_1 x_2, y_1' y_2', z_1' z_2')}{Q(x_1 x_2, y_1' y_2', z_1 z_2)Q(x_1 x_2, y_1 y_2, z_1' z_2')} \frac{Q(y_1' y_2', z_1 z_2)Q(y_1 y_2, z_1' z_2')}{Q(y_1 y_2, z_1 z_2)Q(y_1' y_2', z_1' z_2')} \\
&= \log \frac{Q(x_1, y_1, z_1)Q(x_1, y_1', z_1')}{Q(x_1, y_1', z_1)Q(x_1, y_1, z_1')} \frac{Q(y_1', z_1)Q(y_1, z_1')}{Q(y_1, z_1)Q(y_1', z_1')} \\
&+ \log \frac{Q(x_2, y_2, z_2)Q(x_2, y_2', z_2')}{Q(x_2, y_2', z_2)Q(x_2, y_2, z_2')} \frac{Q(y_2', z_2)Q(y_2, z_2')}{Q(y_2, z_2)Q(y_2', z_2')} \geq 0,
\end{aligned}
$$

and hence, again by Lemma 29, Q is a critical point and solves the optimization problems. $\qquad\square$

Proof of Lemma 21. We use the notation from [5]. The information divergence is jointly convex. Therefore, any critical point of the divergence restricted to a convex set is a global minimizer. Let $y \in \mathcal{Y}$. Then it suffices to show: If P satisfies the two conditional independence statements, then the marginal distribution P_X of X is a critical point of $D(P(\cdot|y)\|Q)$ for Q restricted to $C_{\mathrm{cl}}(\langle Z\rangle_X)$; for if this statement is true, then $P_{y\searrow Z} = P_X$, thus $I_X^\pi(Y \searrow Z) = 0$, and finally $I_{\mathrm{red}}(X:Y;Z) = 0$.

Let $z, z' \in \mathcal{Z}$. The derivative of $D(P(\cdot|y)\|Q)$ at $Q = P_X$ in the direction $P(X|z) - P(X|z')$ is

$$\sum_{x \in \mathcal{X}} (P(x|z) - P(x|z'))\frac{P(x|y)}{P(x)} = \sum_{x \in \mathcal{X}} \left(\frac{P(x,z)P(x,y)}{P(x)P(y)P(z)} - \frac{P(x,z')P(x,y)}{P(x)P(y)P(z')} \right).$$

Now, $Y \perp\!\!\!\perp Z | X$ implies

$$\sum_{x \in \mathcal{X}} \frac{P(x,z)P(x,y)}{P(x)} = P(y,z) \quad \text{and} \quad \sum_{x \in \mathcal{X}} \frac{P(x,z')P(x,y)}{P(x)} = P(y,z'),$$

and $Y \perp\!\!\!\perp Z$ implies $P(y)P(z) = P(y,z)$ and $P(y)P(z') = P(y,z')$. Together, this shows that P_X is a critical point. $\qquad\square$

Acknowledgements

References

[1] Nils Bertschinger and Johannes Rauh. The Blackwell relation and unique information. manuscript.

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

[3] David Blackwell. Equivalent comparisons of experiments. *The Annals of Mathematical Statistics*, 24(2):265–272, 1953.

[4] Virgil Griffith and Christof Koch. Quantifying synergistic mutual information. *arXiv:1205.4265*, 2013.

[5] Malte Harder, Christoph Salge, and Daniel Polani. Bivariate measure of redundant information. *Phys. Rev. E*, 87:012130, 2013.

[6] Serkan Hoşten and Seth Sullivant. Gröbner bases and polyhedral geometry of reducible and cyclic models. *Journal of Combinatorial Theory: Series A*, 100(2):277–301, 2002.

[7] Peter E. Latham and Sheila Nirenberg. Synergy, redundancy, and independence in population codes, revisited. *Journal of Neuroscience*, 25(21):5195–5206, 2005.

[8] William J. McGill. Multivariate information transmission. *Psychometrika*, 19(2):97–116, 1954.

[9] Elad Schneidman, William Bialek, and Michael J. II Berry. Synergy, redundancy, and independence in population codes. *Journal of Neuroscience*, 23(37):11539–11553, 2003.

[10] Paul Williams and Randall Beer. Nonnegative decomposition of multivariate information. *arXiv:1004.2515v1*, 2010.

[11] Wolfram Research, Inc. *Mathematica*. Wolfram Research, Inc., Champaign, Illinois, 8.0 edition, 2003.