Max-Planck-Institut
für Mathematik
in den Naturwissenschaften
Leipzig

On extractable shared information

by

Johannes Rauh, Pradeep Kumar Banerjee,
Eckehard Olbrich, Jürgen Jost, and Nils Bertschinger

On extractable shared information

Johannes Rauh*, Pradeep Kr. Banerjee*, Eckehard Olbrich*, Jürgen Jost*, and Nils Bertschinger[†]

*Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany

{jrauh,pradeep,olbrich,jjost}@mis.mpg.de

[†]Frankfurt Institute for Advanced Studies, Frankfurt, Germany

bertschinger@fias.uni-frankfurt.de

Abstract

We consider the problem of quantifying the information shared by a pair of random variables X_1, X_2 about another variable S. We propose a new measure of shared information, called *extractable shared information*, that is left monotonic; that is, the information shared about S is bounded from below by the information shared about $f(S)$ for any function f. We show that our measure leads to a new nonnegative decomposition of the mutual information $I(S; X_1 X_2)$ into shared, complementary and unique components. We study properties of this decomposition and show that a left monotonic shared information is not compatible with a Blackwell interpretation of unique information. We also discuss whether it is possible to have a decomposition in which both shared and unique information are left monotonic.

Keywords: Information decomposition; multivariate mutual information; left monotonicity; Blackwell order

I. INTRODUCTION

A series of recent papers have focused on the bivariate information decomposition problem [1]–[6]. Consider three random variables S, X_1, X_2 with finite alphabets $\mathcal{S}$, $\mathcal{X}_1$ and $\mathcal{X}_2$, respectively. The total information that the pair (X_1, X_2) convey about the target S can have aspects of *shared* or *redundant* information (conveyed by both X_1 and X_2), of *unique* information (conveyed exclusively by either X_1 or X_2), and of *complementary* or *synergistic* information (retrievable only from the the joint variable (X_1, X_2)). In general, all three kinds of information may be present concurrently. One would like to express this by decomposing the mutual information $I(S; X_1 X_2)$ into a sum of nonnegative components with a well-defined operational interpretation. One possible application area is in the neurosciences. In [7], it is argued that such a decomposition can provide a framework to analyze neural information processing using information theory that can integrate and go beyond previous attempts.

For the general case of k finite source variables $(X_1, \ldots, X_k)$, Williams and Beer [3] proposed the partial information lattice framework that specifies how the total information about the target S is shared across the singleton sources and their disjoint or overlapping coalitions. The lattice is a consequence of certain natural properties of

shared information (sometimes called the *Williams–Beer axioms*). In the bivariate case ($k = 2$), the decomposition has the form

$$I(S; X_1 X_2) = \underbrace{SI(S; X_1, X_2)}_{\text{shared}} + \underbrace{CI(S; X_1, X_2)}_{\text{complementary}} + \underbrace{UI(S; X_1 \backslash X_2)}_{\text{unique } (X_1 \text{ wrt } X_2)} + \underbrace{UI(S; X_2 \backslash X_1)}_{\text{unique } (X_2 \text{ wrt } X_1)}, \tag{1}$$

$$I(S; X_1) = SI(S; X_1, X_2) + UI(S; X_1 \backslash X_2), \tag{2}$$

$$I(S; X_2) = SI(S; X_1, X_2) + UI(S; X_2 \backslash X_1), \tag{3}$$

where $SI(S; X_1, X_2)$, $UI(S; X_1 \backslash X_2)$, $UI(S; X_2 \backslash X_1)$, and $CI(S; X_1, X_2)$ are nonnegative functions that depend continuously on the joint distribution of (S, X_1, X_2). The difference between shared and complementary information is the familiar co-information [8] (or interaction information [9]), a symmetric generalization of the mutual information for three variables,

$$CoI(S; X_1, X_2) = I(S; X_1) - I(S; X_1 | X_2) = SI(S; X_1, X_2) - CI(S; X_1, X_2).$$

Equations (1) to (3) leave only a single degree of freedom, i.e., it suffices to specify either a measure for SI, for CI or for UI.

Williams and Beer not only introduced the general partial information framework, but also proposed a measure of SI to fill this framework. While their measure has subsequently been criticized for "not measuring the right thing" [4]–[6], there has been no successful attempt to find better measures, except for the bivariate case ($k = 2$) [1], [4]. One problem seems to be the lack of a clear consensus on what an ideal measure of shared (or unique or complementary) information should look like and what properties it should satisfy. In particular, the Williams–Beer axioms only put crude bounds on the values of the functions SI, UI and CI. Therefore, additional axioms have been proposed by various authors [4]–[6]. Unfortunately, some of these properties contradict each other [5], and the question for the right axiomatic characterization is still open.

The Williams–Beer axioms do not say anything about what should happen when the target variable S undergoes a local transformation. In this context, the following *left monotonicity* property was proposed in [5]:

(LM) $SI(S; X_1, X_2) \geq SI(f(S); X_1, X_2)$ for any function f. *(left monotonicity)*

Left monotonicity for unique or complementary information can be defined similarly. The property captures the intuition that shared information should only decrease if the target performs some *local* operation (e.g., coarse graining) on her variable S. As argued in [2], left monotonicity of shared and unique information are indeed desirable properties. Unfortunately, none of the measures of shared information proposed so far satisfy left monotonicity.

In this contribution, we study a construction that enforces left monotonicity. Namely, given a measure of shared information SI, define

$$\overline{SI}(S; X_1, X_2) := \sup_{f: \mathcal{S} \to \mathcal{S}'} SI(f(S); X_1, X_2), \tag{4}$$

where the supremum runs over all functions $f : \mathcal{S} \to \mathcal{S}'$ from the domain of S to an arbitrary finite set $\mathcal{S}'$. By construction, $\overline{SI}$ satisfies left monotonicity, and $\overline{SI}$ is the smallest function bounded from below by SI that satisfies left monotonicity.

Changing the definition of shared information in the information decomposition framework Equations (1)–(3) leads to new definitions of unique and complementary information:

$$\overline{UI}^*(S; X_1 \backslash X_2) := I(S; X_1) - \overline{SI}(S; X_1, X_2),$$

$$\overline{UI}^*(S; X_2 \backslash X_1) := I(S; X_2) - \overline{SI}(S; X_1, X_2),$$

$$\overline{CI}^*(S; X_1, X_2) := I(S; X_1 X_2) - \overline{SI}(S; X_1, X_2) - \overline{UI}^*(S; X_1 \backslash X_2) - \overline{UI}^*(S; X_2 \backslash X_1).$$

In general, $\overline{UI}^*(S; X_1 \backslash X_2) \neq \overline{UI}(S; X_1 \backslash X_2) := \sup_{f:\mathcal{S} \to \mathcal{S}'} UI(f(S); X_1 \backslash X_2)$. Thus, our construction cannot enforce left monotonicity for both UI and SI in parallel.

Lemma 2 shows that $\overline{SI}$, $\overline{UI}^*$ and $\overline{CI}^*$ are nonnegative and thus define a nonnegative bivariate decomposition. We study this decomposition in Section IV. In Theorem 1, we show that our construction is not compatible with a decision-theoretic interpretation of unique information proposed in [1]. In Section V, we ask whether it is possible to find an information decomposition in which both shared and unique information measures are left monotonic. Our construction cannot directly be generalized to ensure left monotonicity of two functions simultaneously. Nevertheless, it is possible that such a decomposition exists, and in Proposition 5, we prove bounds on the corresponding shared information measure.

Our original motivation for the definition of $\overline{SI}$ was to find a bivariate decomposition in which the shared information satisfies left monotonicity. However, one could also ask whether left monotonicity is a required property of shared information, as put forward in [2]. In contrast, [4] argue that redundancy can also arise by means of a mechanism. Applying a function to S corresponds to such a mechanism that singles out a certain aspect from S. Even if all the X_i share nothing about the whole S, they might still share information about this aspect of S, which means that the shared information will increase. With this intuition, we can interpret $\overline{SI}$ not as an improved measure of shared information, but as a measure of *extractable shared information*, because it asks for the maximal amount of shared information that can be extracted from S by further processing S by a local mechanism. More generally, one can apply a similar construction to arbitrary information measures. We explore this idea in Section III and discuss probabilistic generalizations and relations to other information measures. In Section VI, we apply our construction to existing measures of shared information.

II. PROPERTIES OF INFORMATION DECOMPOSITIONS

A. The Williams–Beer Axioms

Although we are mostly concerned with the case $k = 2$, let us first recall the three axioms that Williams and Beer [3] proposed for a measure of shared information for arbitrarily many arguments:

(S) $SI(S; X_1, \ldots, X_k)$ is symmetric under permutations of $X_1, \ldots, X_k$, *(Symmetry)*

(SR) $SI(S; X_1) = I(S; X_1)$, *(Self-redundancy)*

(M) $SI(S; X_1, \ldots, X_{k-1}, X_k) \leq SI(S; X_1, \ldots, X_{k-1})$,

with equality if $X_i = f(X_k)$ for some $i < k$ and some function f. *(Monotonicity)*

Any measure of SI satisfying these axioms is nonnegative. Moreover, the axioms imply the following:

(RM) $SI(S; X_1, \ldots, X_k) \geq SI(S; f_1(X_1), \ldots, f_k(X_k))$ for all functions $f_1, \ldots, f_k$. *(right monotonicity)*

Williams and Beer also defined a function

$$I_{\min}(S; X_1, \ldots, X_k) = \sum_s P_S(s) \ \min_i \left\{ \sum_{x_i} P_{X_i|S}(x_i|s) \log \frac{P_{S|X_i}(s|x_i)}{P_S(s)} \right\} \tag{5}$$

and showed that $I_{\min}$ satisfies their axioms.

B. The COPY *example and the Identity Axiom*

Let X_1, X_2 be independent uniformly distributed binary random variables, and consider the copy function $\text{COPY}(X_1, X_2) := (X_1, X_2)$. One point of criticism of $I_{\min}$ is the fact that X_1 and X_2 share $I_{\min}(\text{COPY}(X_1, X_2); X_1, X_2) = 1\,\text{bit}$ about $\text{COPY}(X_1, X_2)$ according to $I_{\min}$, even though they are independent. [4] argue that the shared information about the copied pair should equal the mutual information:

(Id) $SI(\text{COPY}(X_1, X_2); X_1, X_2) = I(X_1; X_2)$. *(Identity)*

Ref. [4] also proposed a bivariate measure of shared information that satisfies **(Id)**. Similarly, the measures of bivariate shared information proposed in [1] satisfies **(Id)**. However, **(Id)** is incompatible with a nonnegative information decomposition according to the Williams–Beer axioms for $k \geq 3$ [2].

On the other hand, Ref. [5] uses an example from game theory to give an intuitive explanation how even independent variables X_1 and X_2 can have nontrivial shared information. However, in any case the value of $1\,\text{bit}$ assigned by $I_{\min}$ is deemed to be too large.

C. The Blackwell property and property (∗)

One of the reasons that it is so difficult to find good definitions of shared, unique or synergistic information is that a clear operational idea behind these notions is missing. Starting from an operational idea about decision problems, Ref. [1] proposed the following property for the unique information, which we now propose to call *Blackwell property*:

(BP) For a given joint distribution $P_{SX_1X_2}$, $UI(S; X_1 \backslash X_2)$ vanishes if and only if there exists a random variable X_1' such that $S - X_2 - X_1'$ is a Markov chain and $P_{SX_1'} = P_{SX_1}$. *(Blackwell property)*

In other words, the channel $S \to X_1$ is a *garbling* or *degradation* of the channel $S \to X_2$. Blackwell's theorem [10] implies that this garbling property is equivalent to the fact that any decision problem in which the task is to predict S can be solved just as well with the knowledge of X_2 as with the knowledge of X_1. We refer to Section 2 in [1] for the details.

Ref. [1] also proposed the following property:

(∗) SI and UI depend only on the marginal distributions P_{SX_1} and P_{SX_2} of the pairs (S, X_1) and (S, X_2).

This property was in part motivated by **(BP)**, which also depends only on the channels $S \to X_1$ and $S \to X_2$ and thus on P_{SX_1} and P_{SX_2}. Most information decompositions proposed so far satisfy property (∗).

III. EXTRACTABLE INFORMATION MEASURES

One can interpret $\overline{SI}$ as a measure of *extractable shared information*. We explain this idea in a more general setting.

For fixed k, let $IM(S; X_1, \ldots, X_k)$ be an arbitrary information measure that measures one aspect of the information that $X_1, \ldots, X_k$ contain about S. At this point, we do not specify what precisely an information measure is, except that it is a function that assigns a real number to any joint distributions of $S, X_1, \ldots, X_k$. The notation is, of course, suggestive of the fact that we mostly think about one of the measures SI, UI or CI, in which the first argument plays a special role. However, IM could also be the mutual information $I(S; X_1)$, the entropy $H(S)$, or the coinformation $CoI(S; X_1, X_2)$. We define the corresponding *extractable* information measure as

$$\overline{IM}(S; X_1, \ldots, X_k) := \sup_f \ IM(f(S); X_1, \ldots, X_k), \tag{6}$$

where the supremum runs over all functions $f : \mathcal{S} \mapsto \mathcal{S}'$ from the domain of S to an arbitrary finite set $\mathcal{S}'$. The intuition is that $\overline{IM}$ is the maximal possible amount of IM one can "extract" from $(X_1, \ldots, X_k)$ by transforming S. Clearly, the precise interpretation depends on the interpretation of IM.

This construction has the following general properties:

1) Most information measures satisfy $IM(O; X_1, \ldots, X_k) = 0$ when O is a constant random variable. Thus, in this case, $\overline{IM}(S; X_1, \ldots, X_k) \geq 0$. Thus, for example, even though the coinformation can be negative, the *extractable coinformation* is never negative.

2) Suppose that IM satisfies left monotonicity. Then, $\overline{IM} = IM$. For example, entropy H and mutual information I satisfy left monotonicity, and so $\overline{H} = H$ and $\overline{I} = I$. Similarly, as shown in [2], the measure of unique information $\widetilde{UI}$ defined in [1] satisfies left monotonicity, and so $\overline{\widetilde{UI}} = \widetilde{UI}$.

3) In fact, $\overline{IM}$ is the smallest left monotonic information measure that is at least as large as IM.

The next result shows that our construction preserves monotonicity properties of the other arguments of IM. It follows that, by iterating this construction, one can construct an information measure that is monotonic in all arguments.

Lemma 1. *Let $f_1, \ldots, f_k$ be fixed functions. If IM satisfies $IM(S; f_1(X_1), \ldots, f_k(X_k)) \leq IM(S; X_1, \ldots, X_k)$ for all S, then $\overline{IM}(S; f_1(X_1), \ldots, f_k(X_k)) \leq \overline{IM}(S; X_1, \ldots, X_k)$ for all S.*

Proof. Let $f^* = \arg\max_f \left\{ IM(f(S); f_1(X_1), \ldots, f_k(X_k)) \right\}$. Then,

$$\overline{IM}(S; f_1(X_1), \ldots, f_k(X_k)) = IM(f^*(S); f_1(X_1), \ldots, f_k(X_k))$$

$$\stackrel{(a)}{\leq} IM(f^*(S); X_1, \ldots, X_k) \leq \sup_f \ IM(f(S); X_1, \ldots, X_k) = \overline{IM}(S; X_1, \ldots, X_k),$$

where (a) follows from the assumptions. $\qquad\square$

As a generalization to the construction, instead of looking at "deterministic extractability," one can also look at "probabilistic extractability" and replace f by a stochastic matrix. This leads to the definition

$$\overset{..}{I}\overset{..}{M}(S; X_1, \ldots, X_k) := \sup_{P_{S'|S}} IM(S'; X_1, \ldots, X_k), \tag{7}$$

where the supremum now runs over all random variables S' that are independent of $X_1, \ldots, X_k$ given S. The function $\overset{..}{I}\overset{..}{M}$ is the smallest function bounded from below by IM that satisfies

(PLM) $IM(S; X_1, X_2) \geq IM(S'; X_1, X_2)$ whenever S' is independent of X_1, X_2 given S.

(probabilistic left monotonicity)

An example of this construction is the intrinsic conditional information $I(X; Y \downarrow Z) := \min_{P_{Z'|Z}} I(X; Y|Z')$, which was defined in [11] to study the secret-key rate, which is the maximal rate at which a secret can be generated by two agents knowing X or Y, respectively, such that a third agent who knows Z has arbitrarily small information about this key. The min instead of the max in the definition implies that $I(X; Y \downarrow Z)$ is "anti-monotone" in Z.

In this paper, we restrict ourselves to the deterministic notions, since many of the properties we want to discuss can already be explained using deterministic extractability. Moreover, the optimization problem (6) is a finite optimization problem and thus much easier to solve than Equation (7).

IV. EXTRACTABLE SHARED INFORMATION

We now specialize to the case of shared information. The first result is that when we apply our construction to a measure of shared information that belongs to a bivariate information decomposition, we again obtain a bivariate information decomposition.

Lemma 2. *Suppose that SI is a measure of shared information, coming from a nonnegative bivariate information decomposition (satisfying Equations* (1) *to* (3)*). Then, $\overline{SI}$ defines a nonnegative information decomposition; that is, the derived functions*

$$\overline{UI}^*(S; X_1 \backslash X_2) := I(S; X_1) - \overline{SI}(S; X_1, X_2),$$

$$\overline{UI}^*(S; X_2 \backslash X_1) := I(S; X_2) - \overline{SI}(S; X_1, X_2),$$

and $\quad \overline{CI}^*(S; X_1, X_2) := I(S; X_1 X_2) - \overline{SI}(S; X_1, X_2) - \overline{UI}^*(S; X_1 \backslash X_2) - \overline{UI}^*(S; X_2 \backslash X_1)$

are nonnegative. These quantities relate to the original decomposition by

a) $\overline{SI}(S; X_1, X_2) \geq SI(S; X_1, X_2),$

b) $\overline{CI}^*(S; X_1, X_2) \geq CI(S; X_1, X_2),$

c) $UI(f^*(S); X_1 \backslash X_2) \leq \overline{UI}^*(S; X_1 \backslash X_2) \leq UI(S; X_1 \backslash X_2),$

where f^* is a function that achieves the supremum in Equation (4)*.*

Proof.

    a) $\overline{SI}(S; X_1, X_2) \geq SI(S; X_1, X_2) \geq 0,$

    b) $\overline{CI}^*(S; X_1, X_2) = \overline{SI}(S; X_1, X_2) - CoI(S; X_1, X_2) \geq SI(S; X_1, X_2) - CoI(S; X_1, X_2)$

        $\geq CI(S; X_1, X_2) \geq 0,$

    c) $\overline{UI}^*(S; X_1 \backslash X_2) = I(S; X_1) - \overline{SI}(S; X_1, X_2) \leq I(S; X_1) - SI(S; X_1, X_2) = UI(S; X_1 \backslash X_2),$

       $\overline{UI}^*(S; X_1 \backslash X_2) = I(S; X_1) - \overline{SI}(S; X_1, X_2) \geq I(f^*(S); X_1) - SI(f^*(S); X_1, X_2)$

        $= UI(f^*(S); X_1 \backslash X_2) \geq 0,$

where we have used the data processing inequality. $\qquad\square$

Lemma 3.   *1) If SI satisfies (∗), then $\overline{SI}$ also satisfies (∗).*

  2) If SI is right monotonic, then $\overline{SI}$ is also right monotonic.

Proof. (1) is direct, and (2) follows from Lemma 1. $\qquad\square$

    Without further assumptions on SI, we cannot say much about when $\overline{SI}$ vanishes. However, the condition that $\overline{UI}^*$ vanishes has strong consequences.

Lemma 4. *Suppose that $\overline{UI}^*(S; X_1 \backslash X_2)$ vanishes, and let f^* be a function that achieves the supremum in Equation* (4)*. Then, there is a Markov chain $X_1 — f^*(S) — S$. Moreover, $UI(f^*(S); X_1 \backslash X_2) = 0$.*

Proof. Suppose that $\overline{UI}^*(S; X_1 \backslash X_2) = 0$. Then, $I(S; X_1) = \overline{SI}(S; X_1, X_2) = SI(f^*(S); X_1, X_2) \leq I(f^*(S); X_1) \leq I(S; X_1)$. Thus, the data processing inequality holds with equality. This implies that $X_1 - f^*(S) - S$ is a Markov chain. The identity $UI(f^*(S); X_1 \backslash X_2) = 0$ follows from the same chain of inequalities. $\qquad\square$

Theorem 1. *If UI has the Blackwell property, then $\overline{UI}^*$ does not have the Blackwell property.*

Proof. As shown in the example in the appendix, there exist random variables S, X_1, X_2 and a function f that satisfy

  1) S and X_1 are independent given $f(S)$.

  2) The channel $f(S) \to X_1$ is a garbling of the channel $f(S) \to X_2$.

  3) The channel $S \to X_1$ is not a garbling of the channel $S \to X_2$.

We claim that f solves the optimization problem (4). Indeed, for an arbitrary function f',

$$SI(f'(S); X_1, X_2) \leq I(f'(S); X_1) \leq I(S; X_1) = I(f(S); X_1) = SI(f(S); X_1, X_2).$$

Thus, f solves the maximization problem (4).

If UI satisfies the Blackwell property, then (2) and (3) imply $UI(f(S); X_1\backslash X_2) = 0$ and $UI(S; X_1\backslash X_2) > 0$. On the other hand,

$$\overline{UI}^*(S; X_1 \setminus X_2) = I(S; X_1) - \overline{SI}(S; X_1, X_2) = I(S; X_1) - SI(f(S); X_1, X_2)$$

$$= I(S; X_1) - I(f(S); X_1) + UI(f(S); X_1\backslash X_2) = 0.$$

Thus, $\overline{UI}^*$ does not satisfy the Blackwell property. $\qquad\square$

Corollary 2. *There is no bivariate information decomposition in which UI satisfies the Blackwell property and SI satisfies left monotonicity.*

Proof. If SI satisfies left monotonicity, then $\overline{SI} = SI$. Thus, $UI = \overline{UI}^*$ cannot satisfy the Blackwell property by Theorem 1. $\qquad\square$

V. Left Monotonic Information Decompositions

Is it possible to have an extractable information decomposition? More precisely, is it possible to have an information decomposition in which all information measures are left monotonic? The obvious strategy of starting with an arbitrary information decomposition and replacing each partial information measure by its extractable analogue does not work, since this would mean increasing all partial information measures (unless they are extractable already), but then their sum would also increase. For example, in the bivariate case, when SI is replaced by a larger function $\overline{SI}$, then UI needs to be replaced by a smaller function, due to the constraints (2) and (3).

As argued in [2], it is intuitive that UI be left monotonic. As argued above (and in [5]), it is also desirable that SI be left monotonic. The intuition for synergy is much less clear. In the following, we restrict our focus to the bivariate case and study the implications of requiring both SI and UI to be left monotonic. Proposition 5 gives bounds on the corresponding SI measure.

Proposition 5. *Suppose that SI, UI and CI define a bivariate information decomposition, and suppose that SI and UI are left monotonic. Then,*

$$SI(f(X_1, X_2); X_1, X_2) \le I(X_1; X_2) \tag{8}$$

for any function f.

Before proving the proposition, let us make some remarks. Inequality (8) is related to the identity axiom. Indeed, it is easy to derive Inequality (8) from the identity axiom and from the assumption that SI is left monotonic. Although Inequality (8) may not seem counterintuitive at first sight, none of the information decompositions proposed so far satisfy this property (the function I_λ from [12] satisfies left monotonicity and has been proposed as a measure of shared information, but it does not lead to a nonnegative information decomposition).

Proof. If SI is left monotonic, then

$$SI(f(X_1, X_2); X_1, X_2) \leq SI(\text{COPY}(X_1, X_2); X_1, X_2)$$

$$= I(\text{COPY}(X_1, X_2); X_1) - UI(\text{COPY}(X_1, X_2); X_1 \backslash X_2).$$

If UI is left monotonic, then

$$UI(\text{COPY}(X_1, X_2); X_1 \backslash X_2) \geq UI(X_1; X_1 \backslash X_2) = I(X_1; X_1) - SI(X_1; X_1, X_2).$$

Note that $I(X_1; X_1) = H(X_1) = I(\text{COPY}(X_1, X_2); X_1)$ and

$$SI(X_1; X_1, X_2) = I(X_1; X_2) - UI(X_1; X_2 \backslash X_1) = I(X_1; X_2).$$

Putting these inequalities together, we obtain $SI(f(X_1, X_2); X_1, X_2) \leq I(X_1; X_2).$ □

VI. EXAMPLES

In this section, we apply our construction to Williams and Beer's measure, $I_{\min}$ [3], and to the bivariate measure of shared information, $\widetilde{SI}$, proposed in [1].

First, we make some remarks on how to compute the extractable information measure (under the assumption that one knows how to compute the underlying information measure itself). The optimization problem (4) is a discrete optimization problem. The search space is the set of functions from the support $\mathcal{S}$ of S to some finite set $\mathcal{S}'$. For the information measures that we have in mind, we may restrict to surjective functions f, since the information measures only depend on events with positive probabilities. Thus, we may restrict to sets $\mathcal{S}'$ with $|\mathcal{S}'| \leq |\mathcal{S}|$. Moreover, the information measures are invariant under permutations of the alphabet $\mathcal{S}$. Therefore, the only thing that matters about f is which elements from $\mathcal{S}$ are mapped to the same element in $\mathcal{S}'$. Thus, any function $f : \mathcal{S} \to \mathcal{S}'$ corresponds to a partition of $\mathcal{S}$, where $s, s' \in \mathcal{S}$ belong to the same block if and only if $f(s) = f(s')$, and it suffices to look at all such partitions. The number of partitions of a finite set $\mathcal{S}$ is the *Bell number* $B_{|\mathcal{S}|}$.

The Bell numbers increase super-exponentially, and for larger sets $\mathcal{S}$, the search space of the optimization problem (4) becomes quite large. For smaller problems, enumerating all partitions in order to find the maximum is still feasible. For larger problems, one would need a better understanding about the optimization problem. For reference, some Bell numbers include:

n	3	4	6	10
B_n	5	15	203	115975

As always, symmetries may help, and so in the COPY example discussed below, where $|\mathcal{S}| = 4$, it suffices to study six functions instead of $B_4 = 15$.

We now compare the measure $\overline{I}_{\min}$, an extractable version of Williams and Beer's measure $I_{\min}$ (see Equation (5) above), to the measure $\overline{\widetilde{SI}}$, an extractable version of the measure $\widetilde{SI}$ proposed in [1]. For the latter, we briefly recall the definitions. Let Δ be the set of all joint distributions of random variables (S, X_1, X_2) with given state

TABLE I

Shared information about $f(X_1, X_2)$ for various functions f (in bits).

f	$I_{\min}$	$\bar{I}_{\min}$	$\widetilde{SI}$	$\overline{\widetilde{SI}}$
COPY	1	1	0	1/2
AND/OR	3/4 log 4/3	3/4 log 4/3	3/4 log 4/3	3/4 log 4/3
XOR	0	0	0	0
SUM	1/2	1/2	1/2	1/2
X_1	0	0	0	0
f_1	1/2	1/2	0	0

spaces $\mathcal{S}$, $\mathcal{X}_1$, $\mathcal{X}_2$. Fix $P = P_{SX_1X_2} \in \Delta$. Define Δ_P as the set of all distributions $Q_{SX_1X_2}$ that preserves the marginals of the pairs (S, X_1) and (S, X_2), that is,

$$\Delta_P := \left\{ Q_{SX_1X_2} \in \Delta : Q_{SX_1} = P_{SX_1},\ Q_{SX_2} = P_{SX_2}, \forall\ (S, X_1, X_2) \in \Delta \right\}.$$

Then, define the functions

$$\widetilde{UI}(S; X_1 \backslash X_2) := \min_{Q \in \Delta_P} I_Q(S; X_1 | X_2),$$

$$\widetilde{UI}(S; X_2 \backslash X_1) := \min_{Q \in \Delta_P} I_Q(S; X_2 | X_1),$$

$$\widetilde{SI}(S; X_1, X_2) := \max_{Q \in \Delta_P} CoI_Q(S; X_1, X_2),$$

$$\widetilde{CI}(S; X_1, X_2) := I(S; X_1 X_2) - \min_{Q \in \Delta_P} I_Q(S; X_1 X_2),$$

where the index Q in I_Q or CoI_Q indicates that the corresponding quantity is computed with respect to the joint distribution Q. The decomposition corresponding to $\widetilde{SI}$ satisfies the Blackwell property and the identity axiom [1]. $\widetilde{UI}$ is left monotonic, but $\widetilde{SI}$ is not [2]. In particular, $\overline{\widetilde{SI}} \neq \widetilde{SI}$. $\widetilde{SI}$ can be characterized as the smallest measure of shared information that satisfies property (∗). Therefore, $\overline{\widetilde{SI}}$ is the smallest left monotonic measure of shared information that satisfies property (∗).

Let $\mathcal{X}_1 = \mathcal{X}_2 = \{0, 1\}$ and let X_1, X_2 be independent uniformly distributed random variables. Table I collects values of shared information about $f(X_1, X_2)$ for various functions f (in bits). The function $f_1 : \{00, 01, 10, 11\} \rightarrow \{0, 1, 2\}$ is defined as

$$f_1(X_1, X_2) := \begin{cases} X_1, & \text{if } X_2 = 1, \\ 2, & \text{if } X_2 = 0. \end{cases}$$

The SUM function is defined as $f(X_1, X_2) := X_1 + X_2$. Table I contains (up to symmetry) all possible non-trivial functions f. The values for the extractable measures are derived from the values of the corresponding non-extractable measures. Note that the values for the extractable versions differ only for COPY from the original ones. In these examples, $\bar{I}_{\min} = I_{\min}$, but as shown in [5], $I_{\min}$ is not left monotonic in general.

VII. Conclusions

We introduced a new measure of shared information that satisfies the left monotonicity property with respect to local operations on the target variable. Left monotonicity corresponds to the idea that local processing will remove information in the target variable and thus should lead to lower values of measures which quantify information about the target variable. Our measure fits the bivariate information decomposition framework; that is, we also obtain corresponding measures of unique and synergistic information. However, we also have shown that left monotonicity for the shared information contradicts the Blackwell property of the unique information, which limits the value of a left monotonic measure of shared information for information decomposition.

We also presented an alternative interpretation of the construction used in this paper. Starting from an arbitrary measure of shared information SI (which need not be left monotonic), we interpret the left monotonic measure $\overline{SI}$ as the amount of shared information that can be extracted from S by local processing.

Our initial motivation for the construction of $\overline{SI}$ was the question to which extent shared information originates from the redundancy between the predictors X_1 and X_2 or is created by the mechanism that generated S. These two different flavors of redundancy were called *source redundancy* and *mechanistic redundancy*, respectively, in [4]. While $\overline{SI}$ cannot be used to completely disentangle source and mechanistic redundancy, it can be seen as a measure of the maximum amount of redundancy that can be created from S using a (deterministic) mechanism. In this sense, we believe that it is an important step forward towards a better understanding of this problem and related questions.

Appendix: Counterexample in Theorem 1

Consider the joint distribution

$f(s)$	s	x_1	x_2	$P_{f(S)SX_1X_2}$
0	0	0	0	1/4
0	1	0	1	1/4
0	0	1	0	1/8
0	1	1	0	1/8
1	2	1	1	1/4

and the function $f : \{0, 1, 2\} \to \{0, 1\}$ with $f(0) = f(1) = 0$ and $f(2) = 1$. Then, X_1 and X_2 are independent uniform binary random variables, and $f(S) = \text{AND}(X_1, X_2)$. In addition, $S - f(S) - X_1$ is a Markov chain. By symmetry, the joint distributions of the pairs $(f(S), X_1)$ and $(f(S), X_2)$ are identical, and so the two channels $f(S) \to X_1$ and $f(S) \to X_2$ are identical, and, hence, trivially, one is a garbling of the other. However, one can check that the channel $S \to X_1$ is not a garbling of the channel $S \to X_2$.

This example is discussed in more detail in [13].

References

[1] N. Bertschinger, J. Rauh, E. Olbrich, J. Jost, and N. Ay, "Quantifying unique information," *Entropy*, vol. 16, no. 4, pp. 2161–2183, 2014.

[2] J. Rauh, N. Bertschinger, E. Olbrich, and J. Jost, "Reconsidering unique information: Towards a multivariate information decomposition," in *Proc. IEEE ISIT*, 2014, pp. 2232–2236.

[3] P. Williams and R. Beer, "Nonnegative decomposition of multivariate information," *arXiv:1004.2515v1*, 2010.

[4] M. Harder, C. Salge, and D. Polani, "A bivariate measure of redundant information," *Phys. Rev. E*, vol. 87, p. 012130, Jan 2013.

[5] N. Bertschinger, J. Rauh, E. Olbrich, and J. Jost, "Shared information — new insights and problems in decomposing information in complex systems," in *Proc. ECCS 2012*.   Springer, 2013, pp. 251–269.

[6] V. Griffith and C. Koch, "Quantifying synergistic mutual information," in *Guided Self-Organization: Inception*, M. Prokopenko, Ed. Springer Berlin Heidelberg, 2014, vol. 9, pp. 159–190.

[7] M. Wibral, V. Priesemann, J. W. Kay, J. T. Lizier, and W. A. Phillips, "Partial information decomposition as a unified approach to the specification of neural goal functions," *Brain and Cognition*, vol. 112, pp. 25 – 38, 2017, perspectives on Human Probabilistic Inferences and the 'Bayesian Brain'. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S027826261530021X

[8] A. J. Bell, "The co-information lattice," in *Proc. Fourth Int. Symp. Independent Component Analysis and Blind Signal Separation (ICA 03)*, 2003.

[9] W. McGill, "Multivariate information transmission," *IRE Trans. Inf. Theory*, vol. 4, no. 4, pp. 93–111, 1954.

[10] D. Blackwell, "Equivalent comparisons of experiments," *The Annals of Mathematical Statistics*, vol. 24, no. 2, pp. 265–272, 1953.

[11] U. Maurer and S. Wolf, "The intrinsic conditional mutual information and perfect secrecy," in *Proc. IEEE ISIT*, 1997.

[12] V. Griffith, E. K. P. Chong, R. G. James, C. J. Ellison, and J. P. Crutchfield, "Intersection information based on common randomness," *Entropy*, vol. 16, no. 4, pp. 1985–2000, 2014.

[13] J. Rauh, P. K. Banerjee, E. Olbrich, J. Jost, N. Bertschinger, and D. Wolpert, "Coarse-graining and the blackwell order," *arXiv:1701.07805*, 2017.