Max-Planck-Institut für Mathematik in den Naturwissenschaften Leipzig

Secret Sharing and Shared Information

by

Johannes Rauh

Preprint no.: 72 2017

SECRET SHARING AND SHARED INFORMATION

JOHANNES RAUH

Abstract. Secret sharing is a cryptographic discipline in which the goal is to distribute information about a secret over a set of participants in such a way that only specific authorized combinations of participants together can reconstruct the secret. Thus, secret sharing schemes are systems of variables in which it is very clearly specified which subsets have information about the secret. As such, they provide perfect model systems for information decompositions. However, following this intuition too far leads to an information decomposition with negative partial information terms, which are difficult to interpret. One possible explanation is that the partial information lattice proposed by Williams and Beer is incomplete and has to be extended to incorporate terms corresponding to higher order redundancy. These results put bounds on information decompositions that follow the partial information framework, and they hint at where the partial information lattice needs to be improved.

1. Introduction

Williams and Beer (2010) have proposed a general framework to decompose the multivariate mutual information $I(S; X_1, \dots, X_n)$ between a target random variable S and predictor random variables $X_1, \dots, X_n$ into different terms (called *partial information terms*) according to different ways in which combinations of the variables $X_1, \dots, X_n$ provide unique, shared or synergistic information about S. Williams and Beer argue that such a decomposition can be based on a measure of shared information. The underlying idea is that any information can be classified according to "who knows what." But is this true?

A situation where the question "who knows what" is easy to answer very precisely is secret sharing, a part of cryptography in which the goal is to distribute information (the *secret*) over a set of participants such that the secret can only be reconstructed if certain *authorized* combinations of the participants join their information (see Beimel (2011) for a survey). The set of authorized combinations is called the *access structure*. Formally, the secret is modelled as a random variable S, and a *secret sharing scheme* assigns a random variable X_i to each participant i in such a way that, if $\{i_1, \dots, i_k\}$ is an authorized set of participants, then S is a function of $X_{i_1}, \dots, X_{i_k}$; that is, $H(S|X_{i_1}, \dots, X_{i_k}) = 0$; and, conversely, if $\{i_1, \dots, i_k\}$ is not authorized, then $H(S|X_{i_1}, \dots, X_{i_k}) > 0$. It is assumed that the participants know the scheme, and so any authorized combination of participants can reconstruct the secret if they join their information. A secret sharing scheme is *perfect* if non-authorized sets of participants know nothing about the secret; i.e., $H(S|X_{i_1}, \dots, X_{i_k}) = H(S)$. Thus, in a perfect secret sharing scheme, it is very clearly specified "who knows what." In this sense, perfect secret sharing schemes provide model systems for which it should be easy to write down an information decomposition.

One connection between secret sharing and information decompositions is that the set of access structures of secret sharing schemes with n participants is in

2000 *Mathematics Subject Classification.* 94A17; 94A62.

Key words and phrases. Information decomposition, partial information lattice, shared information, secret sharing.

one-to-one correspondence with the partial information terms of Williams and Beer. This correspondence makes it possible to give another interpretation to all partial information terms: Namely, the partial information term is a measure of how similar a given system of random variables is to a secret sharing scheme with a given access structure.

This correspondence also allows to introduce the *secret sharing property* that makes precise the above intuition: An information decomposition satisfies this property if and only if any perfect secret sharing scheme has just a single partial information term (which corresponds to its access structure). Lemma 2 states the secret sharing property is implied by the Williams and Beer axioms, which shows that the secret sharing property plays well together with the ideas of Williams and Beer. Proposition 1 shows that in an information decomposition that satisfies a natural generalization of this property, it is possible to prescribe arbitrary nonnegative values to all partial information terms.

These results suggest that perfect secret sharing schemes fit well together with the ideas of Williams and Beer. However, following this intuition too far leads to inconsistencies. As Theorem 4 shows, extending the secret sharing property to pairs of perfect secret sharing schemes leads to negative partial information terms. While other authors have started to build an intuition for negative partial terms and argue that they may be unavoidable in information decompositions, the concluding section collects arguments against such claims and proposes as another possible solutions that the Williams and Beer framework is incomplete and is missing nodes that represent higher order redundancy.

Cryptography, where the goal is not only to transport information (as in coding theory) but also to keep it concealed from unauthorized parties, has initiated many interesting developments in information theory, for example, by introducing new information measures and re-interpreting older ones; see, for example, Maurer and Wolf (1997); Csiszar and Narayan (2004). This manuscript focuses on another contribution of cryptography: probabilistic systems with well-defined distribution of information.

The remainder of this article is organized as follows: Section 2 summarizes definitions and results about secret sharing schemes. Section 3 introduces different secret sharing properties that fix the values that a measure of shared information assigns to perfect secret sharing schemes and combinations thereof. The main result of Section 4 is that the pairwise secret sharing property leads to negative partial information terms. Section 5 discusses the implications of this incompatibility result.

2. Perfect secret sharing schemes

We consider n participants among whom we want to distribute information about a secret in such a way that we can control which subsets of participants together can decrypt the secret.

Definition 1. An *access structure* $\mathcal{A}$ is a family of subsets of $\{1,\dots,n\}$, closed to taking supersets. Elements of $\mathcal{A}$ are called *authorized sets*.

A *secret sharing scheme* with access structure $\mathcal{A}$ is a family of random variables S, $X_1,\dots,X_n$ such that:

- $H(X_A,S) = H(X_A)$, whenever $A \in \mathcal{A}$.

Here, $X_A = (X_i)_{i\in A}$ for all subsets $A \subseteq \{1,\dots,n\}$. A secret sharing scheme is *perfect* if

- $H(X_A,S) = H(X_A) + H(S)$, whenever $A \notin \mathcal{A}$.

The condition for perfection is equivalent to $H(S|X_A) = H(S)$. See Beimel (2011) for a survey on secret sharing.

Theorem 1. *For any access structure $\mathcal{A}$ and any $h > 0$, there exists a perfect secret sharing scheme with access structure $\mathcal{A}$ for which the entropy of the secret S equals $H(S) = h$.*

Proof. Perfect secret sharing schemes for arbitrary access structures were first constructed by Ito *et al.* (1987). In this construction, the entropy of the secret equals 1 bit. Combining n copies of such a secret sharing scheme gives a secret sharing scheme with a secret of n bit. As explained in (Beimel, 2011, Claim 1), the distribution of the secret may be perturbed arbitrarily (as long as the support of the distribution remains the same). In this way it is possible to prescribe the entropy of the secret in a perfect secret sharing scheme. □

Example 1. Let Y_1, Y_2, Y_3, S be independent uniform binary random variables, and let $A = (Y_1, Y_2 \oplus S)$, $B = (Y_2, Y_3 \oplus S)$, $C = (Y_3, Y_1 \oplus S)$, where $\oplus$ denotes addition modulo 2 (or the XOR operation). Then (S, A, B, C) is a perfect secret sharing scheme with access structure

$$\{A, B\}, \{A, C\}, \{B, C\}, \quad \{A, B, C\}.$$

It may be of little surprise that integer addition modulo k is an important building block in many secret sharing schemes.

While existence of perfect secret sharing schemes is solved, there remains the problem of finding efficient secret sharing schemes in the sense that the variables $X_1, \ldots, X_n$ should be as small as possible (in the sense of a small entropy), given a fixed entropy of the secret. For instance, in Example 1, $H(X_i)/H(S) = 2$ for all i. See Beimel (2011) for a survey.

Since an access structure $\mathcal{A}$ is closed to taking supersets, it is uniquely determined by its inclusion-minimal elements

$$\underline{\mathcal{A}} := \big\{A \in \mathcal{A} : \text{if } B \subseteq A \text{ and } B \neq A, \text{ then } B \notin \mathcal{A}\big\}.$$

For instance, in Example 1, the first three elements belong to $\underline{\mathcal{A}}$. The set $\underline{\mathcal{A}}$ has the property that no element of $\underline{\mathcal{A}}$ is a subset of another element of $\underline{\mathcal{A}}$. Such a collection of sets is called an *antichain*. Conversely, any such antichain equals the set of inclusion-minimal elements of a unique access structure.

The antichains have a natural lattice structure, which was used by Williams and Beer to order the different values of shared information and organize them into what they call the *partial information lattice*. The same lattice also has a description in terms of secret sharing.

Definition 2. Let $(A_1, \ldots, A_k)$ and $(B_1, \ldots, B_l)$ be antichains. Then

$$(A_1, \ldots, A_k) \preceq (B_1, \ldots, B_l) \quad :\Longleftrightarrow \quad \text{for any } B_i \text{ there exists } A_j \text{ with } A_j \subseteq B_i.$$

The partial information lattice for the case $n = 3$ is depicted in Figure 1.

Lemma 1. *Let $\mathcal{A}$ be an access structure on $\{1, \ldots, n\}$, and let $(B_1, \ldots, B_l)$ be an antichain. Then $B_1, \ldots, B_l$ are all authorized for $\mathcal{A}$ if and only if $\underline{\mathcal{A}} \preceq (B_1, \ldots, B_l)$.*

Proof. The statement directly follows from the definitions. □

3. Information decompositions of secret sharing schemes

Williams and Beer (2010) proposed to decompose the total mutual information $I(S; X_1, \dots, X_n)$ between a target random variable S and predictor random variables $X_1, \dots, X_n$ according to different ways in which combinations of the variables $X_1, \dots, X_n$ provide unique, shared or synergistic information about S. One of their main ideas is to base such a decomposition on a single *measure of shared information* $I_\cap$, which is a function $I(S; Y_1, \dots, Y_k)$ that takes as arguments a list of random variables, of which the first, S, takes a special role. To arrive at a decomposition of $I(S; X_1, \dots, X_n)$, the variables $Y_1, \dots, Y_k$ are taken to be combinations $X_A = (X_i)_{i \in A}$ of $X_1, \dots, X_n$, corresponding to subsets A of $\{1, \dots, n\}$. For simplicity, $I_\cap(S; X_{A_1}, \dots, X_{A_k})$ is denoted by $I_\cap(S; A_1, \dots, A_k)$ for all $A_1, \dots, A_k \subseteq \{1, \dots, n\}$.

Williams and Beer proposed a list of axioms that such a measure $I_\cap$ should satisfy. It follows from these axioms that it suffices to consider the function $I_\cap(S; A_1, \dots, A_k)$ in the case that $(A_1, \dots, A_k)$ is an antichain. Moreover, $I_\cap(S; \cdot)$ is a monotone function on the partial information lattice (Definition 2). Thus it is natural to write each value $I_\cap(S; A_1, \dots, A_k)$ on the lattice as a sum of local terms I_∂ corresponding to the antichains that lie below $(A_1, \dots, A_k)$ in the lattice:

$$I_\cap(S; A_1, \dots, A_k) = \sum_{(B_1, \dots, B_l) \preceq (A_1, \dots, A_k)} I_\partial(S; B_1, \dots, B_l).$$

The terms I_∂ are called *partial information terms*. This representation always exists, and the partial information terms are uniquely defined (using a Möbius inversion). However, it is not guaranteed that I_∂ is always nonnegative. If I_∂ is nonnegative, then $I_\cap$ is called *locally positive*.

Williams and Beer also defined a function denoted by $I_{\min}$ that satisfies their axioms and that is locally positive. While the framework is intriguing and has attracted a lot of further research (as this special issue illustrates), the function $I_{\min}$ has been critiziced as not measuring the right thing. The difficulty of finding a reasonable measure of shared information that is locally positive (Bertschinger *et al.*, 2013; Rauh *et al.*, 2014) has led some to argue that maybe local positivity is not a necessary requirement for an information decomposition. This issue is discussed further in Section 5.

The goal of this section is to present additional natural properties for a measure of shared information that relate secret sharing with the intuition behind information decompositions. In a perfect secret sharing scheme, any combination of participants knows either nothing or everything about S. This motivates the following definition:

Definition 3. A measure of shared information $I_\cap$ has the *secret sharing property* if and only if for any access structure $\mathcal{A}$ and any perfect secret sharing scheme $(X_1, \dots, X_n, S)$ with access structure $\mathcal{A}$, the following holds:

$$I_\cap(S; A_1, \dots, A_k) = \begin{cases} H(S), & \text{if } A_1, \dots, A_k \text{ are all authorized,} \\ 0, & \text{otherwise,} \end{cases}$$

for any $A_1, \dots, A_k \subseteq \{X_1, \dots, X_n\}$.

Lemma 2. *The secret sharing property is implied by the Williams and Beer axioms.*

Proof. The Williams and Beer axioms imply that

$$I_\cap(S; A_1, \dots, A_k) \le I(S; A_i) = 0$$

whenever A_i is not authorized. On the other hand, when $A_1, \dots, A_k$ are all authorized, then the monotonicity axiom implies

$$I_\cap(S; A_1, \dots, A_k) \ge I_\cap(S; A_1, \dots, A_k, S) = I_\cap(S; S) = H(S). \qquad \square$$

Perfect secret sharing schemes lead to information decompositions with a single nonzero partial information term:

Lemma 3. *If $I_\cap$ has the secret sharing property and if $(X_1, \dots, X_n, S)$ is a perfect secret sharing scheme with access structure $\mathcal{A}$, then*

$$I_\partial(S; A_1, \dots, A_k) = \begin{cases} H(S), & \text{if } \underline{\mathcal{A}} = \{A_1, \dots, A_k\}, \\ 0, & \text{otherwise,} \end{cases} \tag{1}$$

for any $A_1, \dots, A_k \subseteq \{X_1, \dots, X_n\}$.

Proof. Suppose that $\underline{\mathcal{A}} = \{A'_1, \dots, A'_{k'}\}$, and let $J_\partial(S; A_1, \dots, A_k)$ be the right hand side of (1). We need to show that $I_\partial = J_\partial$. Since the Möbius inversion is unique, it suffices to show that $J_\cap = I_\cap$, where

$$J_\cap(S; A_1, \dots, A_k) = \sum_{(B_1,\dots,B_l) \preceq (A_1,\dots,A_k)} J_\partial(S; B_1, \dots, B_l).$$

By Lemma 1,

$$J_\cap(S; A_1, \dots, A_k) = \begin{cases} H(S), & \text{if } A_1, \dots, A_k \text{ are all authorized,} \\ 0, & \text{otherwise,} \end{cases}$$

for any $A_1, \dots, A_k \subseteq \{X_1, \dots, X_n\}$, from which the claim follows. □

What happens when we have several secret sharing schemes involving the same participants? In order to have a clear intuition, assume that the secret sharing schemes satisfy the following definition:

Definition 4. Let $\mathcal{A}_1, \dots, \mathcal{A}_l$ be access structures on $\{1, \dots, n\}$. A *combination of (perfect) secret sharing schemes* with access structures $\mathcal{A}_1, \dots, \mathcal{A}_l$ consists of random variables $S_1, \dots, S_l$, $X_1, \dots, X_n$ such that $(S_i, X_1, \dots, X_n)$ is a (perfect) secret sharing scheme with access structure $\mathcal{A}_i$ for $i = 1, \dots, l$ and such that

$$H(S_i | S_1, \dots, S_{i-1}, S_{i+1}, \dots, S_l, X_A) = H(S_i) \text{ if } A \notin \mathcal{A}_i.$$

This definition ensures that the secrets are independent in the sense that knowing some of the secrets provides no information about the other secrets. Formally, one can see that the secrets are probabilistically independent as follows: For any $A \notin \mathcal{A}_i$ (for example, $A = \emptyset$),

$$H(S_i | S_1, \dots, S_{i-1}, S_{i+1}, \dots, S_l) \geq H(S_i | S_1, \dots, S_{i-1}, S_{i+1}, \dots, S_l, X_A) = H(S_i).$$

In Definition 4, if two access structures $\mathcal{A}_i, \mathcal{A}_j$ are identical, then we can replace S_i and S_j by a single random variable (S_i, S_j) and obtain a smaller combination of (perfect) secret sharing schemes.

In a combination of perfect secret sharing schemes, it is very clear who knows what: Namely, a group of participants knows all secrets for which it is authorized, while it knows nothing about the remaining secrets. This motivates the following definition:

Definition 5. A measure of shared information $I_\cap$ has the *combined secret sharing property* if and only if for any combination of perfect secret sharing schemes with access structures $\mathcal{A}_1, \dots, \mathcal{A}_l$,

$$I_\cap\big((S_1, \dots, S_l); A_1, \dots, A_k\big) = H\big(\{S_i : A_1, \dots, A_k \in \mathcal{A}_i\}\big) \tag{2}$$

(the entropy of those secrets for which $A_1, \dots, A_k$ are all authorized). $I_\cap$ has the *pairwise secret sharing property* if and only if the same holds true in the special case $l = 2$.

The combined secret sharing property implies the pairwise secret sharing property. The pairwise secret sharing property does not follow from the Williams and Beer axioms. For example, $I_{\min}$ satisfies the Williams and Beer axioms, but not the pairwise secret sharing property (as will become apparent in Theorem 2). So one can ask whether the pairwise and combined secret sharing properties are compatible with the Williams and Beer axioms. This question is difficult to answer, since currently there are only two proposed measures of shared information that satisfy the Williams and Beer axioms, namely $I_{\min}$ and the *minimum of mutual informations* (Barrett, 2014)

$$I_{\mathrm{MMI}}(S; A_1, \dots, A_k) := \min_{i=1,\dots,k} I(S; A_i).$$

Both measures do not satisfy the pairwise secret sharing property.

While there has been no further proposal for a function that satisfies the Williams and Beer axioms for arbitrarily many arguments, several measures have been proposed for the "bivariate case" $k = 2$, notably I_{red} of Harder *et al.* (2013) and $\widetilde{SI}$ of Bertschinger *et al.* (2014). The appendix shows that $\widetilde{SI}$ at least satisfies the combined secret sharing property "as far as possible."

Combinations of l perfect secret sharing schemes lead to information decompositions with at most l nonzero partial information terms.

Lemma 4. *Assume that $I_\cap$ has the combined secret sharing property. If $(S_1, \dots, S_l, X_1, \dots, X_n)$ is a combination of perfect secret sharing schemes with pairwise different access structures $\mathcal{A}_1, \dots, \mathcal{A}_l$, then*

$$I_\partial\big((S_1, \dots, S_l); A_1, \dots, A_k\big) = \begin{cases} H(S_i), & \text{if } \underline{\mathcal{A}_i} = \{A_1, \dots, A_k\} \\ & \qquad \text{for some } i \in \{1, \dots, l\}, \\ 0, & \text{otherwise,} \end{cases}$$

for any $A_1, \dots, A_k \subseteq \{X_1, \dots, X_n\}$.

The proof is similar to the proof of Lemma 3 and omitted.

The combined secret sharing property implies that any combination of nonnegative values can be prescribed as partial information values.

Proposition 1. *Suppose that a nonnegative number $h_{\mathcal{A}}$ is given for any antichain $\mathcal{A}$. For any measure of shared information that satisfies the combined secret sharing property, there exist random variables $S, X_1, \dots, X_n$ such that the corresponding partial measure I_∂ satisfies $I_\partial(S; A_1, \dots, A_k) = h_{A_1,\dots,A_k}$ for all antichains $\mathcal{A} = (A_1, \dots, A_k)$.*

Proof. By Theorem 1, for each antichain $\mathcal{A}$ there exists a perfect secret sharing scheme $S_{\mathcal{A}}$, $X_{1,\mathcal{A}}, \dots, X_{n,\mathcal{A}}$ with $H(S_{\mathcal{A}}) = h_{\mathcal{A}}$. Combine independent copies of these perfect secret sharing schemes and let

$$S = (S_{\mathcal{A}})_{\mathcal{A}}, \quad X_1 = (X_{1,\mathcal{A}})_{\mathcal{A}}, \quad \dots, \quad X_n = (X_{n,\mathcal{A}})_{\mathcal{A}},$$

where $\mathcal{A}$ runs over all antichains. Then $S, X_1, \dots, X_n$ is an independent combination of perfect secret sharing schemes, and the statement follows from Lemma 4. □

Unfortunately, not every random variable S can be decomposed in such a way as a combination of secret sharing schemes. However, Proposition 1 suggests that, given a measure $I_\cap$ of shared information that satisfies the combined secret sharing property, $I_\partial(S; \underline{\mathcal{A}})$ can informally be interpreted as a measure that quantifies how much $(X_1, \dots, X_n, S)$ looks like a perfect secret sharing scheme with access structure $\mathcal{A}$.

Lemma 5. *Suppose that $I_\cap$ is a measure of shared information that satisfies the pairwise secret sharing property. If X_1 and X_2 are independent, then*

$$I_\cap\big((X_1, X_2); X_1, X_2\big) = 0.$$

In the language of Ince (2017), the lemma says that the pairwise secret sharing property implies the *independent identity property*.

Proof. Let $S_1 = X_1$, $S_2 = X_2$. Then S_1, S_2, X_1, X_2 is a pair of perfect secret sharing schemes with access structures $\mathcal{A}_1 = \big\{\{1\}\big\}$ and $\mathcal{A}_2 = \big\{\{2\}\big\}$. The statement follows from Definition 5, since X_1 is not authorized for $\mathcal{A}_2$ and X_2 is not authorized for $\mathcal{A}_1$. □

4. Incompatibility with local positivity

Unfortunately, although the combined secret sharing property very much fits the intuition behind the axioms of Williams and Beer, it is incompatible with a nonnegative decomposition according to the partial information lattice:

Theorem 2. *Let $I_\cap$ be a measure of shared information that satisfies the Williams-Beer axioms and has the pairwise secret sharing property. Then I_∂ is not nonnegative.*

Proof. The XOR example, which was already used by Bertschinger *et al.* (2013) and Rauh *et al.* (2014) to prove incompatibility results for properties of information decompositions, can also be used here.

Let X_1, X_2 be independent binary uniform random variables, let $X_3 = X_1 \oplus X_2$, and let $S = (X_1, X_2, X_3)$. Observe that the situation is symmetric in X_1, X_2, X_3. In particular, X_2, X_3 are also independent, and $X_1 = X_2 \oplus X_3$. The following values of $I_\cap$ can be computed from the assumptions:

- $I_\cap\big(S; X_1, (X_2X_3)\big) = I_\cap\big(S; X_1, (X_1X_2X_3)\big) = I_\cap(S; X_1) = 1\,\text{bit}$, since X_1 is a function of (X_2, X_3) and by the monotonicity axiom.
- $I_\cap(S; X_1, X_2) = I_\cap\big((X_1X_2X_3); X_1, X_2\big) = I_\cap\big((X_1X_2); X_1, X_2\big) = 0$ by Lemma 5.

By monotonicity, $I_\cap(S; X_1, X_2, X_3) = 0$. Moreover,

$$I_\cap\big(S; (X_1X_2), (X_1X_3), (X_2X_3)\big) \leq 2\,\text{bit},$$

since 2 bit is the total entropy in the system. But then

$$\begin{aligned} I_\partial\big(S; (X_1X_2), (X_1X_3), (X_2X_3)\big) &= I_\cap\big(S; (X_1X_2), (X_1X_3), (X_2X_3)\big) \\ &\quad - I_\cap\big(S; X_1, (X_2X_3)\big) - I_\cap\big(S; X_2, (X_1X_3)\big) - I_\cap\big(S; X_3, (X_1X_2)\big) \pm 0 \\ &\leq 2\,\text{bit} - 3\,\text{bit} = -1\,\text{bit}, \end{aligned}$$

where ± 0 denotes values of $I_\cap$ that vanish. Thus, I_∂ is not nonnegative. □

Note that the random variables $(S = (X_1, X_2, X_3), X_1, X_2, X_3)$ from the proof of Theorem 2 form three perfect secret sharing schemes that do not satisfy the definition of a combination of perfect secret sharing schemes. The three secrets X_1, X_2, X_3 are not independent, but they are pair-wise independent (and so Lemma 4 does not apply).

Remark 1. The XOR example from the proof of Theorem 2 (which was already used by Bertschinger *et al.* (2013) and Rauh *et al.* (2014)) was criticized by Chicharro and Panzeri (2017) on the grounds that it involves random variables that stand in a deterministic functional relation (in the sense that $X_3 = X_1 \oplus X_2$). Chicharro and Panzeri argue that in such a case it is not appropriate to use the full partial information lattice. Instead, the functional relationship should be used to eliminate (or identify) nodes from the lattice. Thus, while the monotonicity axiom of Williams and Beer implies $I_\cap(S; X_3, (X_2, X_3)) = I_\cap(S; X_3)$ (and so $\{3; 23\}$ is not part of the partial information lattice), the same axiom also implies that $I_\cap(S; X_3, (X_1, X_2)) = I_\cap(S; X_3)$ in the XOR example, and so $\{3; 12\}$ should similarly be excluded from the lattice when analyzing this particular example. But note that the first argument

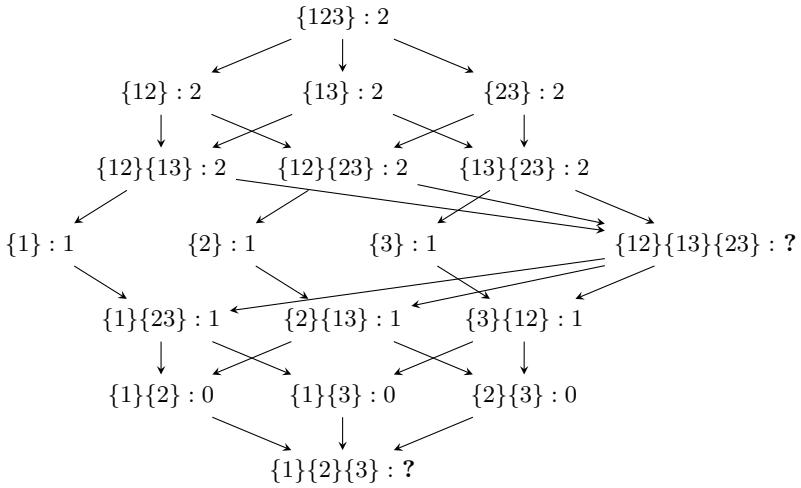

FIGURE 1. The partial information lattice for $n = 3$. Each node is indexed by an antichain. The values (in bit) of the shared information in the XOR example from the proof of Theorem 2 according to the pairwise secret sharing property are given after the colon.

is a formal argument that is valid for all joint distributions of S, X_1, X_2, X_3, while the second argument takes into account the particular underlying distribution.

It is easy to work around this objection. The deterministic relationship disappears when an arbitrarily small stochastic noise is added to the joint distribution. To be precise, let X_1, X_2 be independent binary random variables, and let X_3 be binary with

$$P(X_3 = x_3 | X_1 = x_1, X_2 = x_2) = \begin{cases} 1 - \epsilon, & \text{if } x_3 = x_1 \oplus x_2, \\ \epsilon, & \text{otherwise,} \end{cases}$$

for $0 \leq \epsilon \leq 1$. For $\epsilon = 0$, the example from the proof is recovered. Assuming that the partial information terms depend continuously on this joint distribution, the partial information term $I_\partial\big(S; (X_1X_2), (X_1X_3), (X_2X_3)\big)$ will still be negative for small $\epsilon > 0$. Thus, assuming continuity, the conclusion of Theorem 2 still holds true when the information decomposition according to the full partial information lattice is only considered for random variables that do not satisfy any functional deterministic constraint.

Remark 2. Analyzing the proof of Theorem 2, one sees that the independent identity axiom (Lemma 5) is the main ingredient to arrive at the contradiction. The same property also arises in the other uses of the XOR example (Bertschinger *et al.*, 2013; Rauh *et al.*, 2014).

5. Discussion

Perfect secret sharing schemes correspond to systems of random variables in which it is very clearly specified "who knows what." In such a system, it is easy to assign intuitive values to the shared information nodes in the partial information lattice, and one may conjecture that the intuition behind this assignment is the same intuition that underlies the Williams and Beer axioms, which define the partial information lattice. Moreover, following the same intuition, independent combinations of perfect secret sharing schemes can be used as a tool to construct systems of random variables with prescribable (nonnegative) values of partial information.

Unfortunately, this extension to independent combinations of perfect secret sharing schemes is not without problems: By Theorem 2, it leads to decompositions with negative partial information terms. But what does it mean that the examples derived from the same intuition as the Williams and Beer axioms contradict the same axioms in this way? Is this an indication that the whole idea of information decomposition does not work (and that the question posed in the first paragraph of the introduction cannot be answered affirmatively)?

There are several ways out of this dilemma. The first solution is to assign different values to combinations of perfect secret sharing schemes. This solution will not be pursued further in this text, as it would change the interpretation of the information decomposition as measuring "who knows what."

The second solution is to accept negative partial values in the information decomposition. It has been argued that negative values of information can be given an intuitive interpretation in terms of confusing or misleading information. For event-wise (also called "local") information quantities, such as the event-wise mutual information $i(s;x) = \log(p(s)/p(s|x))$, this interpretation goes back to the early days of information theory Fano (1961). Sometimes, this phenomenon is called "misinformation" (Ince, 2017; Wibral *et al.*, 2015). However, in the usual language, misinformation refers to "false or incorrect information, especially when it is intended to trick someone" Macmillan Publishers Limited (retrieved on 2017/10/05), which is not the effect that is modelled here. Thus, the word misinformation should be avoided, in order not to mislead the reader into the wrong intuition.

While negative event-wise information quantities are well-understood, the situation is more problematic for average quantities. When an agent receives side-information in the form of the value x of a relevant random variable X, she changes her strategy. While the prior strategy should be based on the prior distribution $p(S)$, the new strategy should be based on the posterior $p(S|X = x)$. Clearly, in a probabilistic setting, any change of strategy can lead to a better or worse result in a single instance. On average, though, side-information never hurts (and it is never advantageous on average to ignore side-information), which is why the mutual information is never negative. Similarly, it is natural to expect non-negativity of other information quantities. It is difficult to imagine how correct side-information (or an aspect thereof) can be misleading on average. The situation is different for incorrect information, where the interpretation of a negative value is much easier.

More conceptually, I would suspect that an (averaged) information quantity that may change its sign actually conflates different aspects of information[1], just as the interaction information (or co-information) conflates synergy and redundancy Williams and Beer (2010).

In any case, allowing negative partial values alters the interpretation of an information decomposition to a point where it is questionable whether the word "decomposition" is still appropriate. When decomposing an object into parts, the parts should in some reasonable way be sub-objects. For example, in a Fourier decomposition of a function, the Fourier components are never larger than the function (in the sense of the L^2-norm), and the sum of the squared L^2-norms of the Fourier coefficients equals the squared L^2-norm of the original function. As another example, given a (positive) amount of money and two investment options, it may indeed be possible to invest a negative share of the total amount into one of the two options in order to increase the funds that can be invested in the second option.

[1]One can argue whether the same should be true for event-wise quantities. Recently, Ince (2017) suggested to also write the event-wise mutual information as a difference of non-negative quantities.

However, such short selling is regulated in many countries with much stronger rules than ordinary trading.

I do not claim that an information decomposition with negative partial information terms cannot possibly make sense. However, it has to be made clear precisely how to interpret negative terms, and it is important to distinguish between correct information that leads to a suboptimal decision due to unlikely events happening ("bad luck") and incorrect information that leads to decisions being based on the wrong posterior probabilities (as opposed to the "correct" conditional probabilities).

A third solution is to change the underlying lattice structure of the decomposition. A first step in this direction was done by Chicharro and Panzeri (2017) who propose to decompose mutual information according to subsets of the partial information lattice. However, it is also conceivable that the lattice has to be enlarged.

Williams and Beer derived the partial information lattice from their axioms together with the assumption that everything can be expressed in terms of shared information (that is, according to "who knows what"). Shared information is sometimes equivalently called *redundant information*, but it may be necessary to distinguish the two. Information that is shared by several random variables is information that is accessible to each single random variable, but redundancy can also arise at higher orders. An example is the infamous XOR example from the proof of Theorem 2: In this example, each pair X_i, X_j is independent and contains of two bits, but the total system X_1, X_2, X_3 has only two bits. Therefore, there is one bit of redundancy. However, this redundancy bit is not located anywhere specifically: It is not contained in either of X_1, X_2, X_3, and thus it is not shared information. Since the redundant bit is not part of X_1, it is not "shared" by X_1 in this sense. This phenomenon corresponds to the fact that random variables can be pairwise independent without being independent.

This kind of higher order redundancy does not have a place in the partial information lattice, so it may be that nodes corresponding to higher order redundancy have to be added. When the lattice is enlarged in this way, the structure of the Möbius inversion is changed, and it is possible that the resulting lattice leads to nonnegative partial information terms, without changing those cumulative information values that are already present in the original lattice. If this approach succeeds, the answer to the question from the introduction will be negative: Simply classifying information according to "who knows what" (i.e. shared information) does not work, since it does not capture higher order redundancy. The analysis of extensions of the partial information lattice is scope for future work.

Acknowledgments. I thank Fero Matúš for teaching me about secret sharing schemes. I am grateful to Guido Montúfar and Pradeep Kr. Banerjee for their remarks about the manuscript, and to Nils Bertschinger, Jürgen Jost and Eckehard Olbrich for many inspiring discussions on the topic. I thank the reviewers for many comments, in particular concerning the discussion. I thank the organizers and participants of the PID workshop in December 2016 in Frankfurt, where the material was first presented.

Appendix A. Combined secret sharing properties for small k

This section discusses the defining equation (2) of the combined secret sharing property for $k = 1$ and $k = 2$. The case $k = 1$ is incorporated in the definition of a combination of perfect secret sharing schemes: The following lemma implies that any measure of shared information that satisfies self-redundancy satisfies (2) for $k = 1$. Recall that Williams and Beer's self-redundancy axiom implies that $I_\cap(S; X_A) = I(S; X_A)$.

Lemma 6. *Let $(S_1, \ldots, S_l, X_1, \ldots, X_n)$ be a combination of perfect secret sharing schemes with access structures $\mathcal{A}_1, \ldots, \mathcal{A}_l$. Then*

$$I\big((S_1, \ldots, S_l); X_A\big) = H\big(\{S_i : A \in \mathcal{A}_i\}\big).$$

Proof. Suppose that the secret for which A is authorized are $S_1, \ldots, S_m$. Then

$$\begin{aligned} H(S_1, \ldots, S_l|X_A) &= H(S_1, \ldots, S_m|X_A) + H(S_{m+1}, \ldots, S_l|S_1, \ldots, S_m, X_A) \\ &= H(S_{m+1}, \ldots, S_l|S_1, \ldots, S_m, X_A) \le H(S_{m+1}, \ldots, S_l) \le \sum_{i=m+1}^{l} H(S_i). \end{aligned}$$

On the other hand,

$$\begin{aligned} H(S_{m+1}, \ldots, S_l|S_1, \ldots, S_m, X_A) &= \sum_{i=m+1}^{l} H(S_i|S_1, \ldots, S_{i-1}, X_A) \\ &\ge \sum_{i=m+1}^{l} H(S_i|S_1, \ldots, S_{i-1}, S_{i+1}, \ldots, S_l, X_A) = \sum_{i=m+1}^{l} H(S_i). \end{aligned}$$

By independence (remark after Definition 4), $\sum_{i=m+1}^{l} H(S_i) = H(S_{m+1}, \ldots, S_l)$ and $\sum_{i=1}^{m} H(S_i) = H(S_1, \ldots, S_m)$. Thus,

$$I\big((S_1, \ldots, S_l); X_A\big) = H(S_1, \ldots, S_l) - H(S_1, \ldots, S_l|X_A) = H(S_1, \ldots, S_m). \quad \square$$

The next result shows that the bivariate measure of shared information $\widetilde{SI}(S; X, Y)$ proposed by Bertschinger *et al.* (2014) satisfies Eq. (2) for $k \le 2$. The reader is referred to *loc. cit.* for definitions and elementary properties of $\widetilde{SI}$.

Proposition 2. *Let $(S_1, \ldots, S_l, X_1, \ldots, X_n)$ be a combination of perfect secret sharing schemes with access structures $\mathcal{A}_1, \ldots, \mathcal{A}_l$, Then*

$$\widetilde{SI}\big((S_1, \ldots, S_l); X_{A_1}, X_{A_2}\big) = H\big(\{S_i : A \in \mathcal{A}_1 \cap \mathcal{A}_2\}\big).$$

Proof. For given A_1, A_2, suppose that $S_1, \ldots, S_m$ are the secrets for which at least one of A_1 or A_2 is authorized and that $S_{m+1}, \ldots, S_l$ are the secrets for which neither A_1 nor A_2 is authorized alone.

Let P be the joint distribution of $S_1, \ldots, S_l, X_{A_1}, X_{A_2}$. Let Δ_P be the set of alternative joint distributions for $S_1, \ldots, S_l, X_{A_1}, X_{A_2}$ that have the same marginal distributions as P on the subsets $(S_1, \ldots, S_l, X_{A_1})$ and $(S_1, \ldots, S_l, X_{A_2})$. According to the definition of $\widetilde{SI}$, we need to compare P with the elements of Δ_P and find the maximum of $H_Q\big((S_1, \ldots, S_l)\big|X_{A_1}, X_{A_2}\big)$ over $Q \in \Delta_P$, where the subscript to H indicates with respect to which of these joint distributions the conditional entropy is evaluated.

Define a distribution Q^* for $S_1, \ldots, S_l, X_{A_1}, X_{A_2}$ by

$$Q^*(s_1, \ldots, s_l, x_1, x_2) = P(s_1, \ldots, s_l)P(x_{A_1} = x_1|s_1, \ldots, s_l)P(x_{A_2} = x_2|s_1, \ldots, s_l).$$

Then $Q^* \in \Delta_P$. Under P, the secrets $S_{m+1}, \ldots, S_l$ are independent of X_{A_1} (marginally) and independent of X_{A_2}, and so $S_{m+1}, \ldots, S_l$ are independent of

the pair (X_{A_1}, X_{A_2}) under Q^*. On the other hand, $S_1, \ldots, S_m$ are a function of either X_{A_1} or X_{A_2} under P, and so $S_1, \ldots, S_m$ is a function of (X_{A_1}, X_{A_2}) under Q^*. Thus,

$$H_{Q^*}(S_1, \ldots, S_l | X_{A_1}, X_{A_2}) = H_{Q^*}(S_{m+1}, \ldots, S_l) = H_P(S_{m+1}, \ldots, S_l).$$

On the other hand, under any joint distribution $Q \in \Delta_P$, the secrets $S_1, \ldots, S_m$ are functions of X_{A_1}, X_{A_2}, whence

$$H_Q(S_1, \ldots, S_l | X_{A_1}, X_{A_2}) \leq H_Q(S_{m+1}, \ldots, S_l) = H_P(S_{m+1}, \ldots, S_l).$$

It follows that Q^* solves the optimization problem in the definition of $\widetilde{SI}$.

Suppose that the secrets for which X_{A_1} is authorized are $S_1, \ldots, S_r$ and that the secrets for which X_{A_2} is authorized are $S_s, \ldots, S_m$ (with $1 \leq r, s \leq m$). One computes

$$I_{Q^*}\big((S_1, \ldots, S_l); X_{A_1} \big| X_{A_2}\big) = H(S_1, \ldots, S_{s-1}) = \sum_{i=1}^{s-1} H(S_i) \quad \text{and}$$

$$I_{Q^*}\big((S_1, \ldots, S_l); X_{A_1}\big) = H(S_1, \ldots, S_r) = \sum_{i=1}^{r} H(S_i),$$

whence

$$\begin{aligned}\widetilde{SI}\big((S_1, \ldots, S_l); X_{A_1}, X_{A_2}\big) \\ &= I_{Q^*}\big((S_1, \ldots, S_l); X_{A_1}\big) - I_{Q^*}\big((S_1, \ldots, S_l); X_{A_1} \big| X_{A_2}\big) \\ &= \sum_{i=s}^{r} H(S_i) = H(S_s, \ldots, S_r). \quad \square\end{aligned}$$

References

Williams, P.; Beer, R. Nonnegative Decomposition of Multivariate Information. *arXiv:1004.2515v1* **2010**.

Beimel, A. Secret-sharing Schemes: A Survey. Proceedings of the Third International Conference on Coding and Cryptology; Springer-Verlag: Berlin, Heidelberg, 2011; pp. 11–46.

Maurer, U.; Wolf, S. The intrinsic conditional mutual information and perfect secrecy. Proc. IEEE ISIT, 1997.

Csiszar, I.; Narayan, P. Secrecy capacities for multiple terminals. *IEEE Transactions on Information Theory* **2004**, *50*, 3047–3061.

Ito, M.; Saito, A.; Nishizeki, T. Secret sharing scheme realizing general access structure. Proceedings of the IEEE Global Telecommunication Conf., 1987, pp. 99–102.

Bertschinger, N.; Rauh, J.; Olbrich, E.; Jost, J. Shared Information — New Insights and Problems in Decomposing Information in Complex Systems. In *Proc. ECCS 2012*; Springer, 2013; pp. 251–269.

Rauh, J.; Bertschinger, N.; Olbrich, E.; Jost, J. Reconsidering unique information: Towards a multivariate information decomposition. Proc. IEEE ISIT, 2014, pp. 2232–2236.

Barrett, A.B. An exploration of synergistic and redundant information sharing in static and dynamical Gaussian systems. *CoRR* **2014**, *abs/1411.2832*.

Harder, M.; Salge, C.; Polani, D. A Bivariate measure of redundant information. *Phys. Rev. E* **2013**, *87*, 012130.

Bertschinger, N.; Rauh, J.; Olbrich, E.; Jost, J.; Ay, N. Quantifying unique information. *Entropy* **2014**, *16*, 2161–2183.

Ince, R. Measuring multivariate redundant information with pointwise common change in surprisal. *Entropy* **2017**, *19*, 318.

Chicharro, D.; Panzeri, S. Synergy and Redundancy in Dual Decompositions of Mutual Information Gain and Information Loss. *Entropy* **2017**, *19*.

Fano, R.M. *Transmission of Information*; MIT Press: Cambridge, MA, 1961.

Wibral, M.; Lizier, J.T.; Priesemann, V. Bits from Brains for Biologically Inspired Computing. *Frontiers in Robotics and AI* **2015**, *2*, 5.

Macmillan Publishers Limited. Macmillan Dictionary. Available at `http://www.macmillandictionary.com/`, retrieved on 2017/10/05.

Ince, R. The Partial Entropy Decomposition: Decomposing multivariate entropy and mutual information via pointwise common surprisal. *arXiv:1702.01591* **2017**.

E-mail address: `jrauh@mis.mpg.de`

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