

Article

# Unique Information and Secret Key Agreement

Ryan G. James , Jeffrey Emenheiser  and James P. Crutchfield \* 

Complexity Sciences Center and Physics Department, University of California at Davis, One Shields Avenue, Davis, CA 95616, USA; rgjames@ucdavis.edu (R.G.J.); jemenheiser@ucdavis.edu (J.E.)

\* Correspondence: chaos@ucdavis.edu

Received: 1 November 2018; Accepted: 16 December 2018; Published: 24 December 2018

**Abstract:** The partial information decomposition (PID) is a promising framework for decomposing a joint random variable into the amount of influence each source variable  $X_i$  has on a target variable  $Y$ , relative to the other sources. For two sources, influence breaks down into the information that both  $X_0$  and  $X_1$  redundantly share with  $Y$ , what  $X_0$  uniquely shares with  $Y$ , what  $X_1$  uniquely shares with  $Y$ , and finally what  $X_0$  and  $X_1$  synergistically share with  $Y$ . Unfortunately, considerable disagreement has arisen as to how these four components should be quantified. Drawing from cryptography, we consider the secret key agreement rate as an operational method of quantifying unique information. Secret key agreement rate comes in several forms, depending upon which parties are permitted to communicate. We demonstrate that three of these four forms are inconsistent with the PID. The remaining form implies certain interpretations as to the PID's meaning—interpretations not present in PID's definition but that, we argue, need to be explicit. Specifically, the use of a consistent PID quantified using a secret key agreement rate naturally induces a directional interpretation of the PID. We further reveal a surprising connection between third-order connected information, two-way secret key agreement rate, and synergy. We also consider difficulties which arise with a popular PID measure in light of the results here as well as from a maximum entropy viewpoint. We close by reviewing the challenges facing the PID.

**Keywords:** information theory; partial information decomposition; secret key agreement; cryptography

## 1. Introduction

Consider a joint distribution over “source” variables  $X_0$  and  $X_1$  and “target”  $Y$ . Such distributions arise in many settings: sensory integration, logical computing, neural coding, functional network inference, and many others. One promising approach to understanding how the information shared between  $(X_0, X_1)$ , and  $Y$  is organized is the partial information decomposition (PID) [1]. This decomposition seeks to quantify how much of the information shared between  $X_0$ ,  $X_1$ , and  $Y$  is done so redundantly, how much is uniquely attributable to  $X_0$ , how much is uniquely attributable to  $X_1$ , and finally how much arises synergistically by considering both  $X_0$  and  $X_1$  together.

Unfortunately, the lack of a commonly accepted method of quantifying these components has hindered PID's adoption. In point of fact, several proposed axioms are not mutually consistent [2,3]. Furthermore, to date, there is little agreement as to which should hold. Here, we take a step toward understanding these issues by adopting an operational definition for the unique information. This operational definition comes from information-theoretic cryptography and quantifies the rate at which two parties can construct a secret while a third party eavesdrops. Said more simply, for a source and the target to uniquely share information, no other variables can have any portion of that information—their uniquely shared information is a secret that only the source and target have.

There are four varieties of secret key agreement rate depending on which parties are allowed to communicate, each of which defines a different PID. Each variety also relates to a different intuition as to how the PID operates. We discuss several aspects of these different methods and further demonstrate

that three of the four fail to construct an internally consistent decomposition. The surviving method induces a directionality on the PID that has not been explicitly considered before.

Our development proceeds as follows. Section 2 briefly describes the two-source PID. Section 3 reviews the notion of secret key agreement rate and how to quantify it in three contexts: no one communicates, only Alice communicates, and both Alice and Bob communicate. Section 4 discusses the behavior of the PID quantified utilizing secret key agreement rates as unique information and what intuitions are implied by the choice of who is permitted to communicate. Section 5 compares the behavior of the one consistent secret key agreement rate PID with several others proposed in the literature. Section 6 explores two further implications of our primary results, first in a distribution where two-way communication seems to capture synergistic, third-order connected information and second in the behavior of an extant method of quantifying the PID along with maximum entropy methods. Finally, Section 7 summarizes our findings and speculates about PID's future.

## 2. Partial Information Decomposition

Two-source PID seeks to decompose the mutual information  $I[X_0X_1 : Y]$  between “sources”  $X_0$  and  $X_1$  and a “target”  $Y$  into four nonnegative components. The components identify information that is redundant, uniquely associated with  $X_0$ , uniquely associated with  $X_1$ , and synergistic:

$$\begin{aligned} I[X_0X_1 : Y] = & I_{\partial} [X_0 \cdot X_1 \rightarrow Y] && \text{redundant} \\ & + I_{\partial} [X_0 \rightarrow Y \setminus X_1] && \text{unique with } X_0 \\ & + I_{\partial} [X_1 \rightarrow Y \setminus X_0] && \text{unique with } X_1 \\ & + I_{\partial} [X_0X_1 \rightarrow Y] && \text{synergistic.} \end{aligned} \quad (1)$$

Furthermore, the mutual information  $I[X_0 : Y]$  between  $X_0$  and  $Y$  is decomposed into two components:

$$\begin{aligned} I[X_0 : Y] = & I_{\partial} [X_0 \cdot X_1 \rightarrow Y] && \text{redundant} \\ & + I_{\partial} [X_0 \rightarrow Y \setminus X_1] && \text{unique with } X_0 \end{aligned} \quad (2)$$

and, similarly:

$$\begin{aligned} I[X_1 : Y] = & I_{\partial} [X_0 \cdot X_1 \rightarrow Y] && \text{redundant} \\ & + I_{\partial} [X_1 \rightarrow Y \setminus X_0] && \text{unique with } X_1. \end{aligned} \quad (3)$$

In this way, PID relates the four component pieces of information. However, since Equations (1) to (3) provide only three independent constraints for four quantities; it does not uniquely determine how to quantify them in general. That is, this fourth constraint lies outside of the PID.

By the same logic, though, the decomposition is uniquely determined by quantifying exactly one of its constituents. In the case that one wishes to directly quantify the unique information  $I_{\partial} [X_0 \rightarrow Y \setminus X_1]$  and  $I_{\partial} [X_1 \rightarrow Y \setminus X_0]$ , a consistency relation must hold so that they do not overconstrain the decomposition:

$$I_{\partial} [X_0 \rightarrow Y \setminus X_1] + I[X_1 : Y] = I_{\partial} [X_1 \rightarrow Y \setminus X_0] + I[X_0 : Y]. \quad (4)$$

This ensures that using either Equation (2) or Equation (3) results in the same quantification of  $I_{\partial} [X_0 \cdot X_1 \rightarrow Y]$ .

## 3. Secret Key Agreement

Secret key agreement is a fundamental concept within information-theoretic cryptography [4]. Consider three parties—Alice, Bob, and Eve—who each partially observe a source of common randomness, joint probability distribution  $A, B, E \sim p(a, b, e)$ , where Alice has access only to  $a$ , Bob  $b$ ,

and Eve  $e$ . The central challenge is to determine if it is possible for Alice and Bob to agree on a secret key of which Eve has no knowledge. The degree to which they may generate such a secret key immediately depends on the structure of the joint distribution  $A, B, E$ . It also depends on whether Alice and Bob are allowed to publicly communicate.

Concretely, consider Alice, Bob, and Eve each receiving  $n$  independent, identically distributed samples according to  $p(a, b, e)$ —Alice receiving  $A^n$ , Bob  $B^n$ , and Eve  $E^n$ , where  $X^n$  denotes a sequence of random variables  $X_1, X_2, \dots, X_n$ . Note that, although each party's observations  $X_i, X_j$  are independent, the observations of different parties at the same time are correlated according to  $p(a, b, e)$ . A secret key agreement scheme consists of functions  $f$  and  $g$ , as well as a protocol  $h$  for public communication allowing either Alice, Bob, neither, or both to communicate. In the case of a single party being permitted to communicate—say, Alice—she constructs  $C = h(A^n)$  and then broadcasts it to all parties over an authenticated channel. In the case that both parties are permitted communication, they take turns constructing and broadcasting messages of the form  $C_i = h_i(A^n, C_{[0, \dots, i-1]})$  (Alice) and  $C_i = h_i(B^n, C_{[0, \dots, i-1]})$  (Bob) [5]. Said more plainly, Alice's public messages are a function of her observations and any prior public communication from both parties. Bob's public messages are a function of his observations and any prior communication from both parties.

Formally, a secret key agreement scheme is considered  $R$ -achievable if for all  $\epsilon > 0$ :

$$\begin{aligned} K_A &\stackrel{(1)}{=} f(A^n, C), \\ K_B &\stackrel{(2)}{=} g(B^n, C), \\ p(K_A = K_B = K) &\stackrel{(3)}{\geq} 1 - \epsilon, \\ I[K : C, E^n] &\stackrel{(4)}{\leq} \epsilon, \text{ and} \\ \frac{1}{n} H[K] &\stackrel{(5)}{\geq} R - \epsilon, \end{aligned}$$

where (1) and (2) denote the method by which Alice and Bob construct their keys  $K_A$  and  $K_B$ , respectively, (3) states that their keys must agree with arbitrarily high probability, (4) states that the information about the key which Eve—armed with both her private information  $E^n$  as well as the public communication  $C$ —has access is arbitrarily small, and (5) states that the key consists of approximately  $R$  bits per sample.

The greatest rate  $R$  such that an achievable scheme exists is known as the secret key agreement rate. Notational variations indicate which parties are permitted to communicate. In the case that Alice and Bob are not allowed to communicate, their rate of secret key agreement is denoted  $S(A : B || E)$ . When only Alice is allowed to communicate, their secret key agreement rate is  $S(A \rightarrow B || E)$  or, equivalently,  $S(B \leftarrow A || E)$ . When both Alice and Bob are allowed to communicate, their secret key agreement rate is denoted  $S(A \leftrightarrow B || E)$ . In this, we modified the standard notation for secret key agreement rates to emphasize which party or parties communicate.

The secret key agreement rates obey a simple partial order.  $S(A : B || E)$  lower bounds both  $S(A \rightarrow B || E)$  and  $S(A \leftarrow B || E)$ , since no communication is a special case of one party communicating. Similarly, both  $S(A \rightarrow B || E)$  and  $S(A \leftarrow B || E)$  lower bound  $S(A \leftrightarrow B || E)$ , since only one party communicating is a special case of both parties communicating. Other than having some identical lower and upper bounds,  $S(A \rightarrow B || E)$  and  $S(A \leftarrow B || E)$  are themselves generally incomparable.

In the case of no communication,  $S(A : B || E)$  is given by [6]:

$$S(A : B || E) = H[A \wedge B | E], \quad (5)$$

where  $X \wedge Y$  denotes the Gács–Körner common random variable [7]. It is worth noting that the entropy of this variable, the Gács–Körner common information, is not continuous under smooth changes in

the probability distribution. It is also only nonzero for a measure-zero set of distributions within the simplex; specifically, the set of distributions whose joint events form bipartite graphs with multiple connected components [8].

In the case of one-way communication,  $S(A \rightarrow B || E)$  is given by [9]:

$$S(A \rightarrow B || E) = \max \{I[B : K|C] - I[E : K|C]\}, \quad (6)$$

where the maximum is taken over all variables  $C$  and  $K$ , such that the following Markov condition holds:  $C \text{---} K \text{---} A \text{---} B \text{---} E$ . This quantifies the maximum amount of information that Bob can share with the key above the amount that Eve shares with the key, both given the public communication. It suffices, by the Fenchel–Eggleston strengthening of Carathéodory’s theorem [10], to assume  $K$  and  $C$  alphabets are limited:  $|K| \leq |A|$  and  $|C| \leq |A|^2$ .

There is no such closed-form, calculable solution for  $S(A \leftrightarrow B || E)$ ; however, various upper- and lower-bounds are known [5]. One simple lower bound is the supremum of the two one-way secret key agreement rates, as they are both a subset of bidirectional communication. An even simpler upper bound that we will use is the intrinsic mutual information [11]:

$$I[A : B \downarrow E] = \min_{p(\bar{e}|e)} I[A : B|\bar{E}]. \quad (7)$$

This states that the amount of secret information that Alice and Bob share is no greater than their mutual information conditioned on any modification of Eve’s observations. Here,  $\bar{E}$  is an arbitrary stochastic function of  $E$  or alternatively the result of passing  $E$  through a memoryless channel.

The unique PID component  $I_{\partial}[X_0 \rightarrow Y \setminus X_1]$  could be assigned the value of a secret key agreement rate under four different schemes. First, neither  $X_0$  nor  $Y$  may be allowed to communicate. Second, only  $X_0$  can communicate. Third, only  $Y$  is permitted to communicate. Finally, both  $X_0$  and  $Y$  may be allowed to communicate. Note that the eavesdropper  $X_1$  is not allowed to communicate in any secret sharing schemes here.

Secret key agreement rates have been associated with unique information before. One particular upper bound on  $S(A \leftrightarrow B || E)$ —the intrinsic mutual information Equation (7)—is known to not satisfy the consistency condition Equation (4) [12]. Rosas et al. [13] briefly explored the idea of eavesdroppers influence on the PID. More recently, the relationship between a particular method of quantifying unique information and one-way secret key agreement  $S(X_0 \leftarrow Y || X_1)$  has been considered [14]. Furthermore, there are analogous notions of secret key agreement rates within the channel setting, as opposed to the source setting considered here. We leave an analysis of that setting as future work.

#### 4. Cryptographic Partial Information Decompositions

We now address the application of each form of secret key agreement rate as unique information in turn. For each resulting PID, we consider two distributions. The first is that called POINTWISE UNIQUE, chosen here to exemplify the differing intuitions that can be applied to the PID. The second distribution we look at is entitled PROBLEM as it serves as a counterexample demonstrating that three of the four forms of secret key agreement do not result in a consistent decomposition. Both distributions are given in Figure 1. Although we only consider two distributions here, their behaviors are rich enough for us to draw out the two main results of this work. Further examples are given in Section 5.

Interpreting the POINTWISE UNIQUE [15] distribution is relatively straightforward. The target  $Y$  takes on the values “1” and “2” with equal probability. At the same time, exactly one of the two sources (again with equal probability) will be equal to  $Y$ , while the other is “0”. The mutual information  $I[X_0 : Y] = 1/2$  bit and  $I[X_1 : Y] = 1/2$  bit.

The PROBLEM distribution lacks the symmetry of POINTWISE UNIQUE, yet still consists of four equally probable events. The sources are restricted to take on pairs “00”, “01”, “02”, “10”. The target

$Y$  is equal to a “1” if either  $X_0$  or  $X_1$  is “1”, and is “0” otherwise. With this distribution, the mutual information  $I[X_0 : Y] = 0.3113$  bit and  $I[X_1 : Y] = 1/2$  bit.

PNT. UNQ.				PROBLEM			
$X_0$	$X_1$	$Y$	Pr	$X_0$	$X_1$	$Y$	Pr
0	1	1	1/4	0	0	0	1/4
1	0	1	1/4	0	1	1	1/4
0	2	2	1/4	0	2	0	1/4
2	0	2	1/4	1	0	1	1/4

**Figure 1.** Two distributions of interest: The first, POINTWISE UNIQUE, exemplifies the directionality inherent in the one-way secret key agreement rates. The second, PROBLEM, demonstrates that the no-communication, one-way communication with the source communicating (“camel”), and the two-way communication secret key agreement rates result in inconsistent PIDs (partial information decomposition).

#### 4.1. No Public Communication

In the first case, we consider the unique information from  $X_i$  to  $Y$  as the rate at which  $X_i$  and  $Y$  can agree on a secret key while exchanging no public communication:  $I_{\emptyset} [X_i \rightarrow Y \setminus X_j] = S(X_i : Y \parallel X_j)$ . This approach has some appeal: the PID is defined simply by a joint distribution without any express allowance or prohibition on public communication. However, given its quantification in terms of the Gács-Körner common information, the quantity  $S(X_i : Y \parallel X_j)$  does not vary continuously with the distribution of interest. Now, what is the behavior of this measure on our two distributions of interest?

When applied to POINTWISE UNIQUE, each source and the target are unable to construct a secret key. In turn, each unique information is determined to be 0 bits. This results in a redundancy and a synergy each of 1/2 bit.

The PROBLEM distribution demonstrates the inability of  $S(X_i : Y \parallel X_j)$  to construct a consistent PID. In this instance, as in the case of POINTWISE UNIQUE, no secrecy is possible and each unique information is assigned a value of 0 bits. We therefore determine from Equation (2) that the redundancy should be  $I[X_0 : Y] - I_{\emptyset} [X_0 \rightarrow Y \setminus X_1] = 0.3113 \text{ bit} - 0 \text{ bit} = 0.3113 \text{ bit}$ . Equation (3), however, says that the redundancy is  $I[X_1 : Y] - I_{\emptyset} [X_1 \rightarrow Y \setminus X_0] = 1/2 \text{ bit} - 0 \text{ bit} = 1/2 \text{ bit}$ . This contradiction demonstrates that no-communication secret key agreement rate cannot be used as a PID’s unique components.

The resulting partial information decompositions for both distributions are listed in Table 1.

**Table 1.** Partial information decompositions of POINTWISE UNIQUE and PROBLEM when quantified using no-communication secret key agreement rate. POINTWISE UNIQUE decomposes into 0 bits for either unique information and into 1/2 bit for both the redundancy and synergy. PROBLEM’s redundancy and synergy cannot be quantified, since the two secret key agreement rates result in different quantifications. We have denoted this with the symbol  $\times$ .

$S(X_i : Y \parallel X_j)$		
PNT. UNQ.	$I_{\emptyset} [X_0 X_1 \rightarrow Y]$	1/2 bit
	$I_{\emptyset} [X_0 \rightarrow Y \setminus X_1]$	0 bit
	$I_{\emptyset} [X_1 \rightarrow Y \setminus X_0]$	0 bit
	$I_{\emptyset} [X_0 \cdot X_1 \rightarrow Y]$	1/2 bit
PROBLEM	$I_{\emptyset} [X_0 X_1 \rightarrow Y]$	$\times$
	$I_{\emptyset} [X_0 \rightarrow Y \setminus X_1]$	0 bit
	$I_{\emptyset} [X_1 \rightarrow Y \setminus X_0]$	0 bit
	$I_{\emptyset} [X_0 \cdot X_1 \rightarrow Y]$	$\times$

### 4.2. One-Way Public Communication

We next consider the situation when one of the two parties is allowed public communication. This gives us two options: either the source  $X_i$  communicates to target  $Y$  or vice versa. Both situations enshrine a particular directionality in the resulting PID.

The first, where  $X_i$  constructs  $C = h(X_i^n)$  and publicly communicates it, emphasizes the channels  $X_i \rightarrow Y$  and the channel  $X_0X_1 \rightarrow Y$ . This creates a narrative of the sources conspiring to create the target. We call this interpretation the camel intuition, after the aphorism that a camel is a horse designed by committee. The committee member  $X_i$  may announce what design constraints they brought to the table.

The second option, where  $Y$  constructs  $C = h(Y^n)$  and publicly communicates it, emphasizes the channels  $Y \rightarrow X_i$  and  $Y \rightarrow X_0X_1$ . It implies the situation that the sources are imperfect representations of the target. We call this interpretation the elephant intuition, as it recalls the parable of the blind men describing an elephant for the first time. The elephant  $Y$  may announce which of its features is revealed in a particular instance.

#### 4.2.1. Camels

The first option adopts  $I_0 [X_i \rightarrow Y \setminus X_j] = S(X_i \rightarrow Y \parallel X_j)$ , bringing to mind the idea of sources acting as inputs into some scheme by which the target is produced. When viewed this way, one may ask questions such as “How much information in  $X_0$  is uniquely conveyed to  $Y$ ?”. Furthermore, the channels  $X_0 \rightarrow Y$ ,  $X_1 \rightarrow Y$ , and  $X_0X_1 \rightarrow Y$  take center stage.

Through this lens, the POINTWISE UNIQUE distribution has a clear interpretation. Given any realization, exactly one source is perfectly correlated with the target, while the other is impotently “0”. From this vantage, it is clear that the unique information should each be 1/2 bit, and this is borne out with the one-way secret key agreement rate. To see this, consider  $X_0$  broadcasting each time they observed a “1” or a “2”. This corresponds to the joint events “101” and “202”, respectively. It is clear that, when considering just these two events,  $X_0$  and  $Y$  can safely utilize their observations which will agree exactly and with which  $X_1$  has no knowledge. Since these events occur half the time, we conclude that the secret key agreement rate is 1/2 bit. This implies that the redundancy and synergy of this decomposition are both 0 bits.

For the PROBLEM distribution, we find that  $X_1$  can broadcast the times when they observed a “1” or a “2”, which correspond to  $Y$  having observed a “1” or “0”, respectively. In both instances,  $X_0$  observed a “0” and so cannot deduce what the other two have agreed upon. This leads to  $S(X_1 \rightarrow Y \parallel X_0)$  being equal to 1/2 bit. At the same time,  $S(X_0 \rightarrow Y \parallel X_1)$  vanishes. However, PROBLEM’s redundancy and synergy cannot be quantified, since the two secret key agreement schemes imply different redundancies and so are inconsistent with Equation (4).

The resulting PIDs for both are given in Table 2.

**Table 2.** Partial information decompositions of POINTWISE UNIQUE and PROBLEM when quantified using one-way communication secret key agreement rate with the source permitted public communication: POINTWISE UNIQUE decomposes into 1/2 bit for either unique information and into 0 bits for both the redundancy and synergy. PROBLEM’s redundancy and synergy cannot be quantified, indicated by an  $\times$ .

$S(X_i \rightarrow Y \parallel X_j)$		
PNT. UNQ.	$I_0 [X_0X_1 \rightarrow Y]$	0 bits
	$I_0 [X_0 \rightarrow Y \setminus X_1]$	1/2 bit
	$I_0 [X_1 \rightarrow Y \setminus X_0]$	1/2 bit
	$I_0 [X_0 \cdot X_1 \rightarrow Y]$	0 bits
PROBLEM	$I_0 [X_0X_1 \rightarrow Y]$	$\times$
	$I_0 [X_0 \rightarrow Y \setminus X_1]$	0 bits
	$I_0 [X_1 \rightarrow Y \setminus X_0]$	1/2 bit
	$I_0 [X_0 \cdot X_1 \rightarrow Y]$	$\times$

### 4.2.2. Elephants

When the target  $Y$  is the one party permitted communication, one adopts  $I_{\partial} [X_i \rightarrow Y \setminus X_j] = S(X_i \leftarrow Y \parallel X_j)$  and we can interpret the sources as alternate views of the singular target. Consider, for example, journalism where several sources give differing perspectives on the same event. When viewed this way, one might ask a question such as “How much information in  $Y$  is uniquely captured by  $X_0$ ?”. The channels  $X_0 \leftarrow Y$ ,  $X_1 \leftarrow Y$ , and  $X_0X_1 \leftarrow Y$  are paramount with this approach. We denote these in reverse to emphasize that  $Y$  is still the *target* in the PID.

Considered this way, the POINTWISE UNIQUE distribution takes on a different character. The sources each receive identical descriptions of the target—accurate half the time and erased the remainder. The description is identical, however. Nothing is uniquely provided to either source. This is reflected in the secret key agreement rates:  $Y$  can broadcast her observation, restricting events to either “011” and “101” or to “022” and “202”. In either case, these restrictions do not help in the construction of a secret key since  $Y$  cannot further restrict to cases where it is  $X_0$  that agrees with her and not  $X_1$ . This makes each unique information 0 bits, leaving both the redundancy and synergy 1/2 bit.

The PROBLEM distribution’s unique information are  $S(X_0 \leftarrow Y \parallel X_1) = 0$  bit and  $S(X_1 \leftarrow Y \parallel X_0) = 0.1887$  bit. Unlike the prior two decompositions, this unique information satisfies Equation (4). The resulting redundancy is 0.3113 bit while the synergy is 1/2 bit.

Their PIDs are listed in Table 3. Thus, by having  $Y$  publicly communicate and thus invoking a particular directionality, we finally get a consistent PID.

**Table 3.** PID (partial information decomposition) for POINTWISE UNIQUE and PROBLEM when quantified using one-way communication secret key agreement rate with the target permitted public communication: POINTWISE UNIQUE decomposes into 1/2 bit for either unique information, and into 0 bit for both the redundancy and synergy. PROBLEM admits unique information of 0 bit and 0.1887 bit, respectively. This results in a redundancy of 0.3113 bit and a synergy of 1/2 bit, providing a consistent PID.

$S(X_i \leftarrow Y \parallel X_j)$		
PNT. UNQ.	$I_{\partial} [X_0X_1 \rightarrow Y]$	1/2 bit
	$I_{\partial} [X_0 \rightarrow Y \setminus X_1]$	0 bit
	$I_{\partial} [X_1 \rightarrow Y \setminus X_0]$	0 bit
	$I_{\partial} [X_0 \cdot X_1 \rightarrow Y]$	1/2 bit
PROBLEM	$I_{\partial} [X_0X_1 \rightarrow Y]$	1/2 bit
	$I_{\partial} [X_0 \rightarrow Y \setminus X_1]$	0 bit
	$I_{\partial} [X_1 \rightarrow Y \setminus X_0]$	0.1887 bit
	$I_{\partial} [X_0 \cdot X_1 \rightarrow Y]$	0.3113 bit

### 4.3. Two-Way Public Communication

We finally turn to the full two-way secret key agreement rate:  $I_{\partial} [X_i \leftrightarrow Y \setminus X_j] = S(X_i \leftrightarrow Y \parallel X_j)$ . This approach is also appealing, as it does not ascribe any directionality to interpreting the PID. Furthermore, it varies continuously with the distribution, unlike the no-communication case. However, this quantity is generally impossible to compute directly, with only upper and lower bounds known. Fortunately, this only slightly complicates the analyses we wish to make.

In the case of the POINTWISE UNIQUE distribution, it is not possible to extract more secret information than was done in the camel situation. Therefore, the resulting PID is identical: unique information of 1/2 bit and redundancy and synergy of 0 bits.

PROBLEM, however, is again a problem. In this instance, upper and lower bounds on  $S(X_1 \leftrightarrow Y \parallel X_0)$  converge: the larger of the two one-way secret key agreement rates form a lower bound of 1/2 bit, while the upper bound provided by the intrinsic mutual information is also 1/2 bit, and so we know this value exactly. Utilizing the consistency relation Equation (4), we find that the other unique information must be 0.3113 bit in order for the full decomposition to be consistent.

However, the intrinsic mutual information places an upper bound of 0.1887 bit on  $S(X_0 \leftrightarrow Y \parallel X_1)$ . We therefore must conclude that two-way secret key agreement rates cannot be used to directly quantify unique information and a consistent PID cannot be built using them.

The resulting PIDs for both these distributions can be seen in Table 4.

**Table 4.** PID for POINTWISE UNIQUE and PROBLEM when quantified using two-way communication secret key agreement rate: POINTWISE UNIQUE decomposes into 1/2 bit for either unique information, and into 0 bits for both the redundancy and synergy. PROBLEM’s redundancy and synergy cannot be quantified because the two secret key agreement rates result in different quantifications indicated by the symbol  $\times$ .

$S(X_i \leftrightarrow Y \parallel X_j)$		
PNT. UNQ.	$I_{\Delta} [X_0 X_1 \rightarrow Y]$	0 bit
	$I_{\Delta} [X_0 \rightarrow Y \setminus X_1]$	1/2 bit
	$I_{\Delta} [X_1 \rightarrow Y \setminus X_0]$	1/2 bit
	$I_{\Delta} [X_0 \cdot X_1 \rightarrow Y]$	0 bit
PROBLEM	$I_{\Delta} [X_0 X_1 \rightarrow Y]$	$\times$
	$I_{\Delta} [X_0 \rightarrow Y \setminus X_1]$	$\leq 0.1887$ bit
	$I_{\Delta} [X_1 \rightarrow Y \setminus X_0]$	1/2 bit
	$I_{\Delta} [X_0 \cdot X_1 \rightarrow Y]$	$\times$

#### 4.4. Summary

To conclude, then, there is only one secret-key communication scenario— $Y$  publicly communicates—that yields a consistent PID, as in Table 3. The above arguments by counterexample pruned away the unworkable scenarios, narrowing to only one. Naturally, this does not constitute proof that the remaining scenario always leads to a consistent PID. The narrowing, though, allowed us to turn to numerical searches using the `dit` [16] software package. Extensive searches were unable to find a counterexample. Thus, practically, with a high probability, this scenario leads to consistent PIDs.

Though this is the singular viable secret key agreement rate-based PID, we hesitate to fully endorse its use due to the necessary directionality that comes with it. That is, one must invoke a directionality, unspecified by the PID, to have a consistent PID when using secret key agreement as the basis for the PID component of unique information. It is not immediately obvious as to why only  $S(X_i \leftarrow Y \parallel X_j)$  results in a viable partial information decomposition, if this is indeed the case. We leave a proof as to whether  $I_{\Delta} [X_0 \rightarrow Y \setminus X_1] = S(X_0 \leftarrow Y \parallel X_1)$  and  $I_{\Delta} [X_1 \rightarrow Y \setminus X_0] = S(X_1 \leftarrow Y \parallel X_0)$  satisfy Equation (4) for all distributions or not as an open question. Specifically:

$$\max_{K,C} I[X_0 : K|C] - I[X_1 : K|C] + I[X_1 : Y] \stackrel{?}{=} \max_{K,C} I[X_1 : K|C] - I[X_0 : K|C] + I[X_0 : Y], \quad (8)$$

where both optimizations are performed over the space  $X_0 X_1 \leftrightarrow Y \leftrightarrow K \leftrightarrow C$ . We conjecture that  $S(X_i \leftarrow Y \parallel X_j)$  is the only secret key agreement rate resulting in a viable PID due to the fact that the spaces in which the solutions are found are identical.

The reasons why the other three secret key agreement rates fail to form consistent decompositions are likely particular to each scenario. In the case of no communication, the limitations carried over from the Gács–Körner common information play a major role—specifically, that it vanishes even for weakly mixing distributions. In the case of the one-way “camel” secret key agreement rate, it is possible that the failure arises from the optimization spaces of each unique information being different. Finally, for the case of two-way communication, we offer several speculations in Section 6.1.

These measures of unique information can also be applied within the multivariate sources setting [1], though, like other measures of unique information [17,18], these measures cannot fully quantify the general decompositions there.

### 5. Examples

We next demonstrate the behavior of the partial information decomposition quantified using  $I_{\partial} [X_i \rightarrow Y \setminus X_j] = S(X_i \leftarrow Y \parallel X_j)$ , herein denoted  $I_{\leftarrow}$ . On many of the “standard” distributions—RDN, UNQ, XOR— $I_{\leftarrow}$  behaves as intuited by Griffith [19]. Here, we compare its behavior with that of several other proposed measures— $I_{\min}$  [1],  $I_{\text{proj}}$  [20],  $I_{\text{BROJA}}$  [17],  $I_{\text{CCS}}$  [21],  $I_{\text{dep}}$  [18],  $I_{\pm}$  [15], and  $I_{\text{RR}}$  [22]—across five distributions—AND, DIFF, NOT TWO, PNT. UNQ., and TWO BIT COPY. These distributions are given in Figure 2. Note that Reference [14] proved that the secret key agreement rates  $S(X_i \leftarrow Y \parallel X_j)$  lower bound the unique information of  $I_{\text{BROJA}}$ .

AND				DIFF				NOT TWO				PNT. UNQ.				TWO BIT COPY			
$X_0$	$X_1$	$Y$	Pr	$X_0$	$X_1$	$Y$	Pr	$X_0$	$X_1$	$Y$	Pr	$X_0$	$X_1$	$Y$	Pr	$X_0$	$X_1$	$Y$	Pr
0	0	0	1/4	0	0	0	1/4	0	0	0	1/5	0	1	1	1/4	0	0	0	1/4
0	1	0	1/4	0	0	1	1/4	0	1	0	1/5	1	0	1	1/4	0	1	1	1/4
1	1	0	1/4	0	1	0	1/4	1	0	0	1/5	0	2	2	1/4	1	0	2	1/4
1	1	1	1/4	1	0	1	1/4	1	1	1	1/5	2	0	2	1/4	1	1	3	1/4

Figure 2. The five distributions we utilize to compare partial information decompositions.

The AND distribution—the first set of results in Table 5—yields the same decomposition under  $I_{\min}$ ,  $I_{\text{proj}}$ ,  $I_{\text{BROJA}}$ , and  $I_{\leftarrow}$ . In each of these cases, there is no unique information, resulting in 0.311 bit of redundancy and 1/2 bit of synergy.  $I_{\pm}$  produces negative unique values of -1/4 bit with a redundancy of 0.561 bit and a synergy of 3/4 bit.  $I_{\text{CCS}}$ ,  $I_{\text{dep}}$ , and  $I_{\text{RR}}$  all produce positive unique values, indicating that, at least in this case, they interpret unique information as something more than the ability of source and target to agree on a secret when the target can communicate.

The DIFF distribution—the second set of results in Table 5—is named so due to its keen ability to differentiate PID measures. In this example, two pairs of PIDs coincide:  $I_{\min}$  with  $I_{\text{CCS}}$  and  $I_{\text{BROJA}}$  with  $I_{\text{dep}}$ . In magnitude,  $I_{\leftarrow}$  attributes the least to unique information for this distribution, while both  $I_{\text{BROJA}}$  and  $I_{\text{dep}}$  attribute the most.

The NOT TWO distribution—the third set of results in Table 5—is named due to it consisting of all binary events over three variables that do not contain two “1”s. As far as the behavior of the various PIDs goes, the distribution is very similar to the AND distribution.  $I_{\min}$ ,  $I_{\text{proj}}$ ,  $I_{\text{BROJA}}$ , and  $I_{\leftarrow}$  all allot 0 bit to the unique information, while  $I_{\pm}$  finds the unique information to be negative. Here,  $I_{\text{CCS}}$  also finds them to be negative. There is a major difference between the NOT TWO and AND distributions, however: the NOT TWO possesses a great deal of third-order information, while the AND possesses none. This indicates that, although the PIDs of the two distributions are qualitatively similar, their structures are in fact quite distinct.

Next, we consider the PNT. UNQ. distribution—the fourth set of results in Table 5—again. We see that  $I_{\min}$ ,  $I_{\text{proj}}$ ,  $I_{\text{BROJA}}$ , and  $I_{\leftarrow}$  behave as elephants, while  $I_{\text{CCS}}$  and  $I_{\pm}$  behave as camels.  $I_{\text{RR}}$  does not commit to either but leans towards elephant, while  $I_{\text{dep}}$  splits the difference.

Finally, we consider the venerable TWO BIT COPY distribution, and the final set of results in Table 5. Here,  $I_{\min}$  and  $I_{\pm}$  stand out as assigning one bit to redundant information and one bit to synergy, while assigning nothing to unique. All other measures assign one bit to each unique piece of information and nothing to redundancy and synergy. Note that this is not a directionality issue: all four secret key agreement rates agree that the rate at which a secret key can be constructed is 1 bit.

**Table 5.** Comparison of  $I_{\leftarrow}$  with other proposed methods of quantifying the PID.

	$I_{\partial}$	$I_{\min}$	$I_{\text{proj}}$	$I_{\text{BROJA}}$	$I_{\text{ccs}}$	$I_{\text{dep}}$	$I_{\pm}$	$I_{\text{RR}}$	$I_{\leftarrow}$
AND	$X_0 X_1$	0.500 bit	0.500 bit	0.500 bit	0.292 bit	0.270 bit	0.750 bit	0.189 bit	0.500 bit
	$X_0 \setminus X_1$	0.000 bit	0.000 bit	0.000 bit	0.208 bit	0.230 bit	−0.250 bit	0.311 bit	0.000 bit
	$X_1 \setminus X_0$	0.000 bit	0.000 bit	0.000 bit	0.208 bit	0.230 bit	−0.250 bit	0.311 bit	0.000 bit
	$X_0 \cdot X_1$	0.311 bit	0.311 bit	0.311 bit	0.104 bit	0.082 bit	0.561 bit	0.000 bit	0.311 bit
DIFF	$X_0 X_1$	0.085 bit	0.042 bit	0.000 bit	0.085 bit	0.000 bit	0.292 bit	0.029 bit	0.117 bit
	$X_0 \setminus X_1$	0.104 bit	0.146 bit	0.189 bit	0.104 bit	0.189 bit	−0.104 bit	0.160 bit	0.072 bit
	$X_1 \setminus X_0$	0.104 bit	0.146 bit	0.189 bit	0.104 bit	0.189 bit	−0.104 bit	0.160 bit	0.072 bit
	$X_0 \cdot X_1$	0.208 bit	0.165 bit	0.123 bit	0.208 bit	0.123 bit	0.415 bit	0.151 bit	0.240 bit
NOT TWO	$X_0 X_1$	0.551 bit	0.551 bit	0.551 bit	0.553 bit	0.536 bit	0.634 bit	0.531 bit	0.551 bit
	$X_0 \setminus X_1$	0.000 bit	0.000 bit	0.000 bit	−0.002 bit	0.015 bit	−0.083 bit	0.020 bit	0.000 bit
	$X_1 \setminus X_0$	0.000 bit	0.000 bit	0.000 bit	−0.002 bit	0.015 bit	−0.083 bit	0.020 bit	0.000 bit
	$X_0 \cdot X_1$	0.020 bit	0.020 bit	0.020 bit	0.022 bit	0.005 bit	0.103 bit	0.001 bit	0.020 bit
PNT. UNQ.	$X_0 X_1$	0.500 bit	0.500 bit	0.500 bit	0.000 bit	0.250 bit	0.000 bit	0.333 bit	0.500 bit
	$X_0 \setminus X_1$	0.000 bit	0.000 bit	0.000 bit	0.500 bit	0.250 bit	0.500 bit	0.167 bit	0.000 bit
	$X_1 \setminus X_0$	0.000 bit	0.000 bit	0.000 bit	0.500 bit	0.250 bit	0.500 bit	0.167 bit	0.000 bit
	$X_0 \cdot X_1$	0.500 bit	0.500 bit	0.500 bit	0.000 bit	0.250 bit	0.000 bit	0.333 bit	0.500 bit
TWO BIT	$X_0 X_1$	1.000 bit	0.000 bit	0.000 bit	0.000 bit	0.000 bit	1.000 bit	0.000 bit	0.000 bit
	$X_0 \setminus X_1$	0.000 bit	1.000 bit	1.000 bit	1.000 bit	1.000 bit	0.000 bit	1.000 bit	1.000 bit
	$X_1 \setminus X_0$	0.000 bit	1.000 bit	1.000 bit	1.000 bit	1.000 bit	0.000 bit	1.000 bit	1.000 bit
	$X_0 \cdot X_1$	1.000 bit	0.000 bit	0.000 bit	0.000 bit	0.000 bit	1.000 bit	0.000 bit	0.000 bit

## 6. Discussion

We now turn to two follow-on developments arising from the tools developed thus far. First, we define a distribution whose two-way secret key agreement rates behave in a curious manner with very interesting implications regarding the nature of information itself. Second, we take a closer look at an alternative proposal for quantifying unique information and describe its behavior in relationship to the camel-elephant dichotomy defined in Section 4.

### 6.1. When Conversation Is More Powerful Than a Lecture

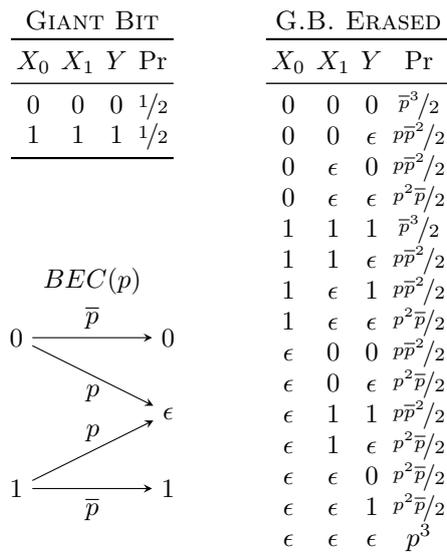
We now explore the PID quantified by two-way secret key agreement further. Though this does not generally form a consistent decomposition, in our exploration of its behavior, we discovered an interesting phenomena independent of its use as a measure of unique information. Consider the GIANT BIT distribution, which exemplifies redundant information. The distribution G.B. ERASED, resulting from passing each variable through an independent binary erasure channel (BEC), exhibits many interesting properties. It is listed in Figure 3. Most notably, the one-way secret key agreement rates between any two variables with the third eavesdropping vanish. However, the two-way secret key agreement rate is equal to  $p\bar{p}^2 = I[X_i : Y|X_j]$  [23]. Furthermore, notice that subtracting Equation (3) from Equation (1) tells us that:

$$\begin{aligned}
 I[X_0 X_1 : Y] - I[X_1 : Y] &= I[X_0 : Y|X_1] \\
 &= I_{\partial} [X_0 \rightarrow Y \setminus X_1] + I_{\partial} [X_0 X_1 \rightarrow Y].
 \end{aligned}
 \tag{9}$$

That is, the conditional mutual information is equal to unique information plus synergistic information.

Evaluating the PID using  $S(X_i \leftrightarrow Y || X_j)$  as unique information results, in this case, in a consistent decomposition. Furthermore, the redundant and synergistic information are zero. This is, however, troublesome: G.B. ERASED possesses nonzero third-order connected information [24], a quantity commonly considered a component of the synergy [18]. Indeed, it is provably attributed to synergy by both the  $I_{\text{dep}}$  [18] and  $I_{\text{BROJA}}$  [17] methods, and likely others as well. No other proposed method of quantifying the PID results in zero redundancy or synergy. The implication here is that, if indeed

the third-order connected information is a form of synergy, the two-way secret key agreement rate overestimates unique information by including some types of synergistic effect. Therefore, we conclude that bidirectional communication between two parties can, in some instances, determine information held solely in trivariate interactions. This may underlie the inability of a two-way secret key agreement rate to form a consistent partial information decomposition: in some instances, the third-order connected information is attributed to both unique information when it should be attributed solely to the synergistic information. It remains to understand (i) how independently and identically transforming a distribution with no third-order connected information can result in its creation and (ii) how only two of the variables can recover it when allowed to communicate.



**Figure 3.** Distribution whose one-way secret key agreement rates are all 0bits, yet has a nonzero two-way secret key agreement rate. It is constructed from the GIANT BIT distribution by passing each variable independently through a binary erasure channel  $BEC(p)$  with erasure probability  $p$ . This distribution has a two-way secret key agreement rate of  $p\bar{p}^2$  between any two variables with the third as an eavesdropper.

### 6.2. $I_{BROJA}$ , the Elephant

The measure  $I_{BROJA}$  of Bertschinger et al. [17] is perhaps the most widely accepted and used method of quantifying the PID. Though popular, it has its detractors [15,21]. Here, we interpret the criticisms leveled and  $I_{BROJA}$  as a product of camel intuitions being applied to an elephantinesque [14] measure. In doing so, we will primarily consider the POINTWISE UNIQUE distribution.

As noted in Section 4.2, if a source is permitted to communicate with the target, then a secret key agreement rate of  $1/2$  bit is achievable, while, if the target communicates with the source, then it is impossible to agree on a secret key. From this camel perspective, it is clear that each source uniquely determines the target, half the time. The elephant perspective, however, allots nothing to unique information as each source is provided with identical information. This would greatly disconcert the camel and may lead one to think that the elephant has “artificially inflated” the redundancy [21]. We next take a closer look at this notion, using  $I_{BROJA}$ .

In the course of computing  $I_{BROJA}$  for the distribution  $p(X_0, X_1, Y)$ , the set of distributions:

$$Q = \{q(X_0, X_1, Y) : \forall i, q(X_i, Y) = p(X_i, Y)\}$$

is considered. The (\*) assumption [17] is then invoked, which states that redundancy and all unique information are constant within this family of distributions. To complete the quantification, the distribution with minimum  $I[X_0X_1 : Y]$  is selected from this family. The resulting distribution

associated with the POINTWISE UNIQUE distribution can be seen in Figure 4. Made explicit, it can now be seen that  $I_{\text{BROJA}}$  does indeed correlate the sources, but, under assumption (\*), this does not affect the redundancy.

One aspect of  $I_{\text{BROJA}}$  and assumption (\*) we believe warrants further investigation is its relationship with maximum entropy philosophy [25]. The latter is, in effect, Occam’s razor applied to probability distributions: given a set of constraints, the most natural distribution to associate with them is that with maximum entropy. As it turns out, this is equivalent to the distribution nearest the unstructured product-of-marginals distribution  $\bar{p}(x, y, z, \dots) = p(x)p(y)p(z) \dots$  [26]:

$$\arg \max_{q \in Q} H[q] = \arg \min_{q \in Q} D_{\text{KL}} [q || \bar{p}] ,$$

where  $D_{\text{KL}} [P || Q]$  is the relative entropy between distributions  $P$  and  $Q$ . Having briefly introduced the ideas behind maximum entropy, we next cast their light on the BROJA optimization employed to calculate  $I_{\text{BROJA}}$ .

Let us first consider the distribution resulting from BROJA optimization. Its entropy is unchanged from the POINTWISE UNIQUE distribution indicating that it has the same amount of structure—they are equally distant from the product distribution. The BROJA distribution has reduced  $I[X_0X_1 : Y]$  mutual information, however, indicating perhaps that the optimization has shifted some of the distribution’s structure away from the sources–target interaction. It is interesting that this optimization could not simply remove the synergy from the distribution altogether, resulting in a larger entropy.

If one takes assumption (\*) and directly applies the maximum entropy philosophy, a different distribution results. This distribution, seen in Figure 4, has a larger entropy than both the POINTWISE UNIQUE and the BROJA intermediate distribution, indicating that it in fact has less structure than either. Under assumption (\*), the MAXENT distribution, also in Figure 4, retains all the redundant and unique information, while under maximum entropy it contains no structure not implied by the source–target marginals—specifically, no synergy. The combination of assumption (\*) and maximum entropy philosophy does not, however, result in a viable partial information decomposition; the maximization of entropy can result in source–target mutual information which exceeds that of the original distribution.

a) BROJA				b) MAXENT			
$X_0$	$X_1$	$Y$	Pr	$X_0$	$X_1$	$Y$	Pr
0	0	1	1/4	0	0	1	1/8
0	0	2	1/4	0	0	2	1/8
1	1	1	1/4	0	1	1	1/8
2	2	2	1/4	0	2	2	1/8
				1	0	1	1/8
				1	1	1	1/8
				2	0	1	1/8
				2	2	2	1/8

**Figure 4.** Two modified forms of the POINTWISE UNIQUE distribution. (a) intermediate distribution resulting from the BROJA optimization. It has the minimum sources–target mutual information consistent with the source–target marginals; (b) maximum entropy distribution consistent with the source–target marginals. It contains no structure beyond that implied by those marginals.

To be clear, this is not to claim that assumption (\*) or BROJA optimization are wrong or incorrect, only that the optimization’s behavior in light of well-established maximum entropy principles is subtle and requires careful investigation. For example, it may be that the source–target marginals do imply some level of triadic interaction and therefore the maximum entropy distribution reflects this lingering synergy. At the same time, BROJA minimization may be capable of maintaining that level of structure implied by the marginals, but somehow shunts it into  $H[Y|X_0X_1]$ .

## 7. Conclusions

At present, a primary barrier for PID's general adoption as a useful and possibly central tool in analyzing how complex systems store and process information is an agreement on a method to quantify its component information. Here, we posited that one reason for disagreement stems from conflicting intuitions regarding the decomposition's operational behavior. To give an operational meaning to unique information and address these intuitions, we equated unique information with the ability of two parties to agree on a secret—a reasonably intuitive operationalization of what it means for two variables to share a piece of information that no others have. This led to numerous observations.

The first is that the PID, as currently defined, is ambivalent to any notion of directionality. There are, however, very clear cases in which the assumption of a directionality—or lack thereof—is critical to the existence of unique information. Consider, for example, the case of the McGurk effect [27] where the visual stimulus of one phoneme and the auditory stimulus of another phoneme gives rise to the perception of a third phoneme. By construction, the stimuli cause the perception, and the channels implicit in a camel intuition are central. If one were to study this interaction using an elephant-like PID, it is unclear that the resulting decomposition would reflect the neurobiological mechanisms by which the perception is produced. Similarly, a camel-like measure would be inappropriate when interpreting simultaneous positron emission tomography (PET) and magnetic resonance imaging (MRI) scans of a tumor.

One can view this as the PID being inherently context-dependent and conclude that quantification requires specifying directionality. In this case, the elephant intuition is apparently more natural, as adopting closely-related notions from cryptography results in a consistent PID. If context demands the camel intuition, though, either a noncryptographic method of quantifying unique information is needed or consistency must be enforced by augmenting the secret key agreement rate. It is additionally possible that associating secret key agreement rates with unique information is fundamentally flawed and that, ultimately, PID entails quantifying unique information as something distinct from the ability to agree on a secret key. Whatever is missing has yet to be identified.

The next observation concerns the third-order connected information. We first demonstrated that such triadic information can be constructed from common information in which each constituent variable is independently and identically modified. Furthermore, it was shown that any two of three parties, when engaging in bidirectional communication, capture the triadic information. Granted, this does not generically occur. For example, if  $X_0$   $X_1$   $Y$  are related by XOR, the distribution contains 1 bit of third-order connected information, but  $S(X_0 \leftrightarrow Y || X_1)$  (or any permutation of the variables) is equal to 0 bits. This suggests that the third-order connected information may not be an atomic quantity, but rather consists of two parts, one accessible to two communicating parties and one not. This idea has been explored in a different context in Reference [28].

Our third observation regards the behavior of the  $I_{\text{BROJA}}$  measure, especially in relation to standard maximum entropy principles. We first demonstrated that  $I_{\text{BROJA}}$  indeed correlates sources, but argued that this behavior only seems inappropriate when adopting a camel intuition. We then discussed how its intermediate distribution is as structured as the initial one and so, if indeed  $I_{\text{BROJA}}$  is operating correctly, it must shunt the dependencies that result in synergy to another aspect of the distribution. Finally, we discussed how the standard maximum entropy approach may remove synergy from a distribution all together. This calls for a more careful investigation as to whether it does (and BROJA optimization is incorrect) or does not and if synergistic information arises from source-target marginals and Occam's razor.

Looking to the future, we trust that this exploration of the relationship between cryptographic secrecy and unique information will provide a basis for future efforts to understand and quantify the partial information decomposition. Furthermore, the explicit recognition of the role that directional intuitions play in the meaning and interpretation of a decomposition should reduce cross-talk and improve understanding as we collectively move forward.

**Author Contributions:** All authors contributed equally to this work.

**Funding:** This material is based upon work supported by, or in part by, Foundational Questions Institute grant FQXi-RFP-1609, the U.S. Army Research Laboratory and the U.S. Army Research Office under contracts W911NF-13-1-0390 and W911NF-13-1-0340 and grant W911NF-18-1-0028, and via Intel Corporation support of the Complexity Sciences Center as an Intel Parallel Computing Center.

**Acknowledgments:** All calculations were performed using the dit Python package [16]. We thank Pradeep Banerjee, Eckehard Olbrich, Samuel Loomis, and Daniel Feldspar for many helpful discussions. As a faculty member, J.P.C. thanks the Santa Fe Institute and the Telluride Science Research Center for their hospitality during visits.

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

## References

- Williams, P.L.; Beer, R.D. Nonnegative decomposition of multivariate information. *arXiv* **2010**, arXiv:1004.2515.
- Rauh, J.; Banerjee, P.; Olbrich, E.; Jost, J.; Bertschinger, N. On extractable shared information. *Entropy* **2017**, *19*, 328. [[CrossRef](#)]
- Rauh, J. Secret sharing and shared information. *Entropy* **2017**, *19*, 601. [[CrossRef](#)]
- Maurer, U.M. Secret key agreement by public discussion from common information. *IEEE Trans. Inf. Theory* **1993**, *39*, 733–742. [[CrossRef](#)]
- Gohari, A.; Günlü, O.; Kramer, G. Coding for positive rate in the source model key agreement problem. *arXiv* **2017**, arXiv:1709.05174.
- Chitambar, E.; Fortescue, B.; Hsieh, M.-H. The conditional common information in classical and quantum secret key distillation. *IEEE Trans. Inf. Theory* **2018**, *64*, 7381–7394. [[CrossRef](#)]
- Gács, P.; Körner, J. Common information is far less than mutual information. *Prob. Control Inf. Theory* **1973**, *2*, 149–162.
- Salamatian, S.; Cohen, A.; Médard, M. Maximum Entropy Functions: Approximate Gacs-Korner for Distributed Compression. *arXiv* **2016**, arXiv:1604.03877.
- Ahlsvede, R.; Csiszár, I. Common randomness in information theory and cryptography. I. Secret sharing. *IEEE Trans. Inf. Theory* **1993**, *39*, 1121–1132 [[CrossRef](#)]
- Wyner, A.; Ziv, J. The rate-distortion function for source coding with side information at the decoder. *IEEE Trans. Inf. Theory* **1976**, *22*, 1–10. [[CrossRef](#)]
- Maurer, U.M.; Wolf, S. Unconditionally secure key agreement and the intrinsic conditional information. *IEEE Trans. Inf. Theory* **1999**, *45*, 499–514 [[CrossRef](#)]
- Bertschinger, N.; Rauh, J.; Olbrich, E.; Jost, J. Shared information—New insights and problems in decomposing information in complex systems. In *Proceedings of the European Conference on Complex Systems 2012*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 251–269.
- Rosas, F.; Ntranos, V.; Ellison, C.J.; Pollin, S.; Verhelst, M. Understanding interdependency through complex information sharing. *Entropy* **2016**, *18*, 38. [[CrossRef](#)]
- Banerjee, P.K.; Olbrich, E.; Jost, J.; Rauh, J. Unique information and deficiencies. *arXiv* **2018**, arXiv:1807.05103.
- Finn, C.; Lizier, J.T. Pointwise partial information decomposition using the specificity and ambiguity lattices. *Entropy* **2018**, *20*, 297. [[CrossRef](#)]
- James, R.G.; Ellison, C.J.; Crutchfield, J.P. Dit: A Python package for discrete information theory. *J. Open Source Softw.* **2018**, *3*, 738. [[CrossRef](#)]
- Bertschinger, N.; Rauh, J.; Olbrich, E.; Jost, J.; Ay, N. Quantifying unique information. *Entropy* **2014**, *16*, 2161–2183. [[CrossRef](#)]
- James, R.G.; Emenheiser, J.; Crutchfield, J.P. Unique information via dependency constraints. *J. Phys. A* **2019**, *52*, 014002. [[CrossRef](#)]
- Griffith, V.; Koch, C. Quantifying synergistic mutual information. In *Guided Self-Organization: Inception*; Springer: Berlin/Heidelberg, Germany, 2014; p. 159.
- Harder, M.; Salge, C.; Polani, D. Bivariate measure of redundant information. *Phys. Rev. E* **2013**, *87*, 012130. [[CrossRef](#)] [[PubMed](#)]
- Ince, R.A.A. Measuring multivariate redundant information with pointwise common change in surprisal. *Entropy* **2017**, *19*, 318. [[CrossRef](#)]

22. Goodwell, A.E.; Kumar, P. Temporal information partitioning: Characterizing synergy, uniqueness, and redundancy in interacting environmental variables. *Water Resour. Res.* **2017**, *53*, 5920–5942. [[CrossRef](#)]
23. Gohari, A.; Anantharam, V. Comments on “information-theoretic key agreement of multiple terminals: Part I”. *IEEE Trans. Inf. Theory* **2017**, *63*, 5440–5442. [[CrossRef](#)]
24. Schneidman, E.; Still, S.; Berry, M.J.; Bialek, W. Network information and connected correlations. *Phys. Rev. Lett.* **2003**, *91*, 238701. [[CrossRef](#)] [[PubMed](#)]
25. Jaynes, E.T. Where do we stand on maximum entropy? In *Essays on Probability, Statistics, and Statistical Physics*; Jaynes, E.T., Ed.; Reidel: London, UK, 1983; p. 210.
26. Amari, S. Information geometry on hierarchy of probability distributions. *IEEE Trans. Inf. Theory* **2001**, *47*, 1701–1711. [[CrossRef](#)]
27. McGurk, H.; MacDonald, J. Hearing lips and seeing voices. *Nature* **1976**, *264*, 746–748. [[CrossRef](#)] [[PubMed](#)]
28. Ince, R.A.A. The partial entropy decomposition: Decomposing multivariate entropy and mutual information via pointwise common surprisal. *arXiv* **2017**, arXiv:1702.01591.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).