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

(Dated: November 6, 2018)

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 informations. 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. These reveal an inconsistency between third-order connected information, two-way secret key agreement rate, and synergy. Similar difficulties arise with a popular PID measure in light the results here as well as from a maximum entropy viewpoint. We close by reviewing the challenges facing the PID.

I. 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]. And, 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.

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.

Our development proceeds as follows. Section II briefly describes the two-source PID. Section III 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 IV discusses the behavior of the PID quantified utilizing secret key agreement rates as unique informations and what intuitions are implied by the choice of who is permitted to communicate. Section V 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 VI summarizes our findings and speculates about PID's future.

II. PARTIAL INFORMATION DECOMPOSITION

Two-source PID seeks to decompose the mutual information $I[X_0 X_1 : Y]$ between "sources" X_0 and X_1 and a "target" Y into four nonnegative components. The com-

---

[*] rgjames@ucdavis.edu
[†] jemenheiser@ucdavis.edu
[‡] chaos@ucdavis.edu

ponents identify information that is redundant, uniquely associated with X_0, uniquely associated with X_1, and synergistic:

$$
\begin{aligned}
\mathrm{I}[X_0 X_1 : Y] = \quad & \mathrm{I}_\partial \left[X_0 \cdot X_1 \to Y\right] && redundant \\
+ & \mathrm{I}_\partial \left[X_0 \to Y \setminus X_1\right] && unique~with~X_0 \\
+ & \mathrm{I}_\partial \left[X_1 \to Y \setminus X_0\right] && unique~with~X_1 \\
+ & \mathrm{I}_\partial \left[X_0 X_1 \to Y\right] . && synergistic
\end{aligned}
\tag{1}
$$

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

$$
\begin{aligned}
\mathrm{I}[X_0 : Y] = \quad & \mathrm{I}_\partial \left[X_0 \cdot X_1 \to Y\right] && redundant \\
+ & \mathrm{I}_\partial \left[X_0 \to Y \setminus X_1\right] . && unique~with~X_0
\end{aligned}
\tag{2}
$$

And, similarly:

$$
\begin{aligned}
\mathrm{I}[X_1 : Y] = \quad & \mathrm{I}_\partial \left[X_0 \cdot X_1 \to Y\right] && redundant \\
+ & \mathrm{I}_\partial \left[X_1 \to Y \setminus X_0\right] . && unique~with~X_1
\end{aligned}
\tag{3}
$$

In this way, PID relates the four component informations. However, since Eqs. (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 informations $\mathrm{I}_\partial \left[X_0 \to Y \setminus X_1\right]$ and $\mathrm{I}_\partial \left[X_1 \to Y \setminus X_0\right]$, a consistency relation must hold so that they do not overconstrain the decomposition:

$$
\begin{aligned}
\mathrm{I}_\partial &\left[X_0 \to Y \setminus X_1\right] + \mathrm{I}[X_1 : Y] \\
&= \mathrm{I}_\partial \left[X_1 \to Y \setminus X_0\right] + \mathrm{I}[X_0 : Y] .
\end{aligned}
\tag{4}
$$

This ensures that using either Eq. (2) or Eq. (3) results in the same quantification of $\mathrm{I}_\partial \left[X_0 \cdot X_1 \to Y\right]$.

III.   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 $ABE \sim p(a, b, e)$, 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 upon a secret key of which Eve has no knowledge. The degree to which they may generate such a secret key immediately depends upon the structure of the joint distribution ABE. It also depends upon whether Alice and Bob are allowed to publicly communicate.

Concretely, consider Alice, Bob, and Eve each receiving n independent, identically distributed samples from ABE—Alice receiving A^n, Bob B^n, and Eve E^n. 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. 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].

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

$$
K_A \overset{(1)}{=} f(A^n, C) ,
$$
$$
K_B \overset{(2)}{=} g(B^n, C) ,
$$
$$
p(K_A = K_B = K) \overset{(3)}{\geq} 1 - \epsilon ,
$$
$$
\mathrm{I}[K : CE^n] \overset{(4)}{\leq} \epsilon , ~\text{and}
$$
$$
\frac{1}{n} \mathrm{H}[K] \overset{(5)}{\geq} R - \epsilon ,
$$

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 to be 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 $\mathrm{S}(A : B \mid\mid E)$. When only Alice is allowed to communicate their secret key agreement rate is $\mathrm{S}(A \to B \mid\mid E)$ or, equivalently, $\mathrm{S}(B \leftarrow A \mid\mid E)$. When both Alice and Bob are allowed to communicate, their secret key agreement rate is denoted $\mathrm{S}(A \leftrightarrow B \mid\mid E)$. In this, we modified the standard notation for secret key agreement rates to emphasize which party or parties communicate.

In the case of no communication, $\mathrm{S}(A : B \mid\mid E)$ is given by [6]:

$$
\mathrm{S}(A : B \mid\mid E) = \mathrm{H}[A \curlywedge B | E] ,
\tag{5}
$$

where $X \curlywedge Y$ denotes the Gács-Körner common random variable [7]. It is worth noting that this quantity does not vary continuously with the distribution and generically vanishes.

In the case of one-way communication, $S(A \rightarrow B \parallel E)$ is given by [8]:

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

where the maximum is taken over all variables C and K, such that the following Markov condition holds: $C \!-\!\!\circ\!-\! K \!-\!\!\circ\!-\! A \!-\!\!\circ\!-\! BE$. It suffices to consider K and C such that $|K| \leq |A|$ and $|C| \leq |A|^2$.

There is no such solution for $S(A \leftrightarrow B \parallel E)$; 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* [9]:

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

This effectively states that any information Eve has access to through any local modification of her observations cannot be secret.

The unique PID component $I_\partial \left[X_0 \setminus X_1 \rightarrow Y\right]$ 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 informations before. One particular upper bound on $S(A \leftrightarrow B \parallel E)$—the intrinsic mutual information Eq. (7)—is known to not satisfy the consistency condition Eq. (4) [10]. More recently, the relationship between a particular method of quantifying unique information and one-way secret key agreement $S(X_0 \leftarrow Y \parallel X_1)$ has been considered [11].

IV. 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 PONTWISE UNIQUE, chosen here to exemplify the differing intuitions that can be applied to the PID.

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	$1/4$

FIG. 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.

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 Fig. 1.

Interpreting the POINTWISE UNIQUE [12] 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 informations $I[X_0 : Y] = 1/2$ 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 informations $I[X_0 : Y] = 0.3113$ bit and $I[X_1 : Y] = 1/2$ bit.

A. 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 upon a secret key while exchanging no public communication: $I_\partial \left[X_i \setminus X_j \rightarrow Y\right] = S(X_i : Y \parallel X_j)$. 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 bit. This results in a redundancy and a synergy each of $1/2$ bit.

$S(X_i : Y \parallel X_j)$	
PNT. UNQ.	
$I_\partial [X_0 X_1 \to Y]$	1/2 bit
$I_\partial [X_0 \to Y \setminus X_1]$	0 bit
$I_\partial [X_1 \to Y \setminus X_0]$	0 bit
$I_\partial [X_0 \cdot X_1 \to Y]$	1/2 bit
PROBLEM	
$I_\partial [X_0 X_1 \to Y]$	✗
$I_\partial [X_0 \to Y \setminus X_1]$	0 bit
$I_\partial [X_1 \to Y \setminus X_0]$	0 bit
$I_\partial [X_0 \cdot X_1 \to Y]$	✗

TABLE I. Partial information decompositions of POINT-WISE UNIQUE and PROBLEM when quantified using no-communication secret key agreement rate. POINTWISE UNIQUE decomposes into 0 bit 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.

$S(X_i \to Y \parallel X_j)$	
PNT. UNQ.	
$I_\partial [X_0 X_1 \to Y]$	0 bit
$I_\partial [X_0 \to Y \setminus X_1]$	1/2 bit
$I_\partial [X_1 \to Y \setminus X_0]$	1/2 bit
$I_\partial [X_0 \cdot X_1 \to Y]$	0 bit
PROBLEM	
$I_\partial [X_0 X_1 \to Y]$	✗
$I_\partial [X_0 \to Y \setminus X_1]$	0 bit
$I_\partial [X_1 \to Y \setminus X_0]$	1/2 bit
$I_\partial [X_0 \cdot X_1 \to Y]$	✗

TABLE II. 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 bit for both the redundancy and synergy. PROBLEM's redundancy and synergy cannot be quantified.

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 bit. We therefore determine from Eq. (2) that the redundancy should be $I[X_0 : Y] - I_\partial [X_0 \setminus X_1 \to Y] = 0.3113\,\text{bit} - 0\,\text{bit} = 0.3113\,\text{bit}$. Equation (3), however, says the redundancy is $I[X_1 : Y] - I_\partial [X_1 \setminus X_0 \to Y] = 1/2\,\text{bit} - 0\,\text{bit} = 1/2\,\text{bit}$. 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 I.

B. 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 $X_i \to Y$ and 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 $Y \to X_i$ and 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.

1. Camels

The first option adopts $I_\partial [X_i \setminus X_j \to Y] = S(X_i \to Y \parallel X_j)$, 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 $X_0 \to Y$ and $X_1 \to Y$ 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 informations should each be 1/2 bit, and this is borne out with the one-way secret key agreement rate. This implies that the redundancy and synergy of this decomposition are both 0 bit.

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 $S(X_1 \to Y \parallel X_0)$ being equal to 1/2 bit. At the same time, $S(X_0 \to Y \parallel X_1)$ 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 Eq. (4).

The resulting PIDs for both are given in Table II.

$\mathrm{S}(X_i \leftarrow Y \mid\mid X_j)$	
$\mathrm{I}_\partial\left[X_0 X_1 \rightarrow Y\right]$	$1/2$ bit
$\mathrm{I}_\partial\left[X_0 \rightarrow Y \setminus X_1\right]$	0 bit
$\mathrm{I}_\partial\left[X_1 \rightarrow Y \setminus X_0\right]$	0 bit
$\mathrm{I}_\partial\left[X_0 \cdot X_1 \rightarrow Y\right]$	$1/2$ bit
$\mathrm{I}_\partial\left[X_0 X_1 \rightarrow Y\right]$	$1/2$ bit
$\mathrm{I}_\partial\left[X_0 \rightarrow Y \setminus X_1\right]$	0 bit
$\mathrm{I}_\partial\left[X_1 \rightarrow Y \setminus X_0\right]$	0.1887 bit
$\mathrm{I}_\partial\left[X_0 \cdot X_1 \rightarrow Y\right]$	0.3113 bit

(First group: PNT. UNQ.; second group: PROBLEM)

TABLE III. PID 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 informations 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.

$\mathrm{S}(X_i \leftrightarrow Y \mid\mid X_j)$	
$\mathrm{I}_\partial\left[X_0 X_1 \rightarrow Y\right]$	0 bit
$\mathrm{I}_\partial\left[X_0 \rightarrow Y \setminus X_1\right]$	$1/2$ bit
$\mathrm{I}_\partial\left[X_1 \rightarrow Y \setminus X_0\right]$	$1/2$ bit
$\mathrm{I}_\partial\left[X_0 \cdot X_1 \rightarrow Y\right]$	0 bit
$\mathrm{I}_\partial\left[X_0 X_1 \rightarrow Y\right]$	✗
$\mathrm{I}_\partial\left[X_0 \rightarrow Y \setminus X_1\right]$	≤ 0.1887 bit
$\mathrm{I}_\partial\left[X_1 \rightarrow Y \setminus X_0\right]$	$1/2$ bit
$\mathrm{I}_\partial\left[X_0 \cdot X_1 \rightarrow Y\right]$	✗

(First group: PNT. UNQ.; second group: PROBLEM)

TABLE IV. 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 bit 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.

2. Elephants

When the target Y is the one party permitted communication, one adopts $\mathrm{I}_\partial\left[X_i \setminus X_j \rightarrow Y\right] = \mathrm{S}(X_i \leftarrow Y \mid\mid X_j)$ 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 $X_0 \leftarrow Y$ and $X_1 \leftarrow Y$ 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, which are 0 bit, leaving both the redundancy and synergy $1/2$ bit.

The PROBLEM distribution's unique informations are $\mathrm{S}(X_0 \leftarrow Y \mid\mid X_1) = 0$ bit and $\mathrm{S}(X_1 \leftarrow Y \mid\mid X_0) = 0.1887$ bit. Unlike the prior two decompositions, these unique informations satisfy Eq. (4). The resulting redundancy is 0.3113 bit while the synergy is $1/2$ bit.

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

C. Two-Way Public Communication

We finally turn to the full two-way secret key agreement rate: $\mathrm{I}_\partial\left[X_i \setminus X_j \rightarrow Y\right] = \mathrm{S}(X_i \leftrightarrow Y \mid\mid X_j)$. 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 informations of $1/2$ bit and redundancy and synergy of 0 bit.

PROBLEM, however, is again a problem. Upper and lower bounds on $\mathrm{S}(X_1 \leftrightarrow Y \mid\mid X_0)$ converge[1] to $1/2$ bit, and so we know this value exactly. Utilizing the consistency relation Eq. (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 $\mathrm{S}(X_0 \leftrightarrow Y \mid\mid X_1)$. 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 IV.

D. Summary

To conclude, then, there is only one secret-key communication scenario—Y publicly communicates—that yields a

---

[1] In this instance, 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.

consistent PID, as in Table III. While we have not proven this, we have been unable to find a counterexample after extensive numerical searches using the dit [13] software package. That is, one must invoke a directionality, unspecified by the PID, to have a consistent PID if using secret key agreement as the basis for the PID component of unique information.

V. 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 IV.

A. When Conversation is More Powerful Than a Lecture

We now explore the PID quantified by two-way secret key agreement further. 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 Fig. 2. 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 $p\bar{p}^2 = \mathrm{I}[X_i : Y | X_j]$ [14]. Furthermore, notice that subtracting Eq. (3) from Eq. (1) tells us that:

$$\mathrm{I}[X_0 X_1 : Y] - \mathrm{I}[X_1 : Y]$$
$$= \mathrm{I}[X_0 : Y | X_1]$$
$$= \mathrm{I}_\partial [X_0 \to Y \setminus X_1] + \mathrm{I}_\partial [X_0 X_1 \to Y] . \quad (8)$$

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

Evaluating the PID using $\mathrm{S}(X_i \leftrightarrow Y \parallel X_j)$ as unique information results, in this case, in a consistent decomposition. Furthermore, the redundant and synergistic informations are zero. This is, however, troublesome: G.B. ERASED possesses nonzero third-order connected information [15], a quantity commonly considered a component of the synergy [16]. Indeed, it is provably attributed to synergy by both the $\mathrm{I}_{\mathrm{dep}}$ [16] and $\mathrm{I}_{\mathrm{BROJA}}$ [17] methods, and likely others as well. No other proposed

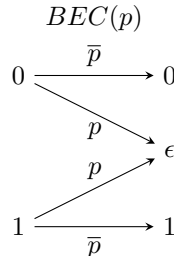| GIANT BIT				G.B. ERASED			
X_0	X_1	Y	Pr	X_0	X_1	Y	Pr
0	0	0	$^1\!/_2$	0	0	0	$\bar{p}^3/2$
1	1	1	$^1\!/_2$	0	0	ϵ	$p\bar{p}^2/2$
				0	ϵ	0	$p\bar{p}^2/2$
				0	ϵ	ϵ	$p^2\bar{p}/2$
				1	1	1	$\bar{p}^3/2$
				1	1	ϵ	$p\bar{p}^2/2$
				1	ϵ	1	$p\bar{p}^2/2$
				1	ϵ	ϵ	$p^2\bar{p}/2$
				ϵ	0	0	$p\bar{p}^2/2$
				ϵ	0	ϵ	$p^2\bar{p}/2$
				ϵ	1	1	$p\bar{p}^2/2$
				ϵ	1	ϵ	$p^2\bar{p}/2$
				ϵ	ϵ	0	$p^2\bar{p}/2$
				ϵ	ϵ	1	$p^2\bar{p}/2$
				ϵ	ϵ	ϵ	p^3

$BEC(p)$

FIG. 2. Distribution whose one-way secret key agreement rates are all 0 bit, yet has 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.

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 component 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. It remains to be understood (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.

B. $\mathrm{I}_{\mathrm{BROJA}}$, the Elephant

The measure of Bertschinger *et al.* [17], here referred to as $\mathrm{I}_{\mathrm{BROJA}}$, is perhaps the most widely accepted and used method of quantifying the PID. Though popular, it has its detractors [12, 18]. Here, we interpret the criticisms leveled and $\mathrm{I}_{\mathrm{BROJA}}$ as a product of camel intuitions being applied to an elephantesque [11] measure. In doing so, we will primarily consider the POINTWISE UNIQUE distribution.

As noted in Section IV B, 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 upon a secret key. From this camel perspective it is clear that each source, half the time, uniquely determines the target. The elephant perspective, however, allots nothing to unique informations 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. 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 informations are constant within this family of distributions. To complete the quantification, the distribution with minimum $I[X_0 X_1 : Y]$ is selected from this family. The resulting distribution associated with the POINTWISE UNIQUE distribution can be seen in Fig. 3. Made explicit, it can now be seen that I_{BROJA} does indeed correlate the sources, but under assumption $(*)$ this does not effect the redundancy.

One aspect of I_{BROJA} and assumption $(*)$ we believe warrants further investigation is its relationship with *maximum entropy* philosophy [19]. 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 $\overline{p}(x, y, z, \ldots) = p(x)p(y)p(z) \ldots$ [20]:

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

where $D_{KL} [P \ || \ P]$ 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_{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 a reduced $I[X_0 X_1 : Y]$ 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.

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	$1/8$

FIG. 3. 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.

If one takes assumption $(*)$ and directly applies the maximum entropy philosophy, a different distribution results. This distribution, seen in Fig. 3, 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 Fig. 3, retains all the redundant and unique informations, while under maximum entropy it contains no structure not implied by the source-target marginals—*e.g.*, no synergy.

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 a 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 $H[Y|X_0 X_1]$.

VI. CONCLUSION

At present, a primary barrier for PID's general adoption as a useful and possibly a central tool in analyzing how complex systems store and process information is an agreement on a method to quantify its component informations. 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 upon a secret. This leads 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 there of—is critical to the existence of unique information. Consider, for example, the case of the McGurk effect [21] 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 PET and 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 upon a secret key. This missing thing 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 from which each variable is then independently and identically modified. Furthermore, it has been shown that two of those three parties, when engaging in bidirectional communication, can capture this triadic information. 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 $S(X_0 \leftrightarrow Y \mid\mid X_1)$ (or any permutation of the variables)

is equal to 0 bit. 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.

Our third observation regards the behavior of the I_{BROJA} measure, especially in relation to standard maximum entropy principles. We first demonstrated that I_{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_{BROJA} is operating correctly, it must shuffle 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 synergistic information is implied under 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.

ACKNOWLEDGMENTS

[1] P. L. Williams and R. D. Beer. Nonnegative decomposition of multivariate information. *arXiv:1004.2515.* 1

[2] J. Rauh, P. Banerjee, E. Olbrich, J. Jost, and N. Bertschinger. On extractable shared information. *Entropy*, 19(7):328, 2017. 1

[3] J. Rauh. Secret sharing and shared information. *Entropy*, 19(11):601, 2017. 1

[4] U. M. Maurer. Secret key agreement by public discussion from common information. *IEEE Trans. Info. Th.*, 39(3):733–742, 1993. 2

[5] A. Gohari, O. Günlü, and G. Kramer. Coding for positive rate in the source model key agreement problem. *arXiv:1709.05174*. 2, 3

[6] E. Chitambar, B. Fortescue, and M.-H. Hsieh. The conditional common information in classical and quantum secret key distillation. *IEEE Trans. Info. Th.*, 2018. 2

[7] P. Gács and J. Körner. Common information is far less than mutual information. *Problems of Control and Information Theory*, 2(2):149–162, 1973. 3

[8] R. Ahlswede and I. Csiszár. Common randomness in information theory and cryptography. I. Secret sharing. *IEEE Trans. Info. Th.*, 39(4):1121–1132, 1993. 3

[9] U. M. Maurer and S. Wolf. Unconditionally secure key agreement and the intrinsic conditional information. *IEEE Trans. Info. Th.*, 45(2):499–514, 1999. 3

[10] N. Bertschinger, J. Rauh, E. Olbrich, and J. Jost. Shared information—new insights and problems in decomposing information in complex systems. In *Proceedings of the European Conference on Complex Systems 2012*, pages 251–269. Springer, 2013. 3

[11] P. K. Banerjee, E. Olbrich, J. Jost, and J. Rauh. Unique informations and deficiencies. *arXiv:1807.05103*. 3, 6

[12] C. Finn and J. T. Lizier. Pointwise partial information decomposition using the specificity and ambiguity lattices. *Entropy*, 20(4):297, 2018. 3, 6

[13] R. G. James, C. J. Ellison, and J. P. Crutchfield. dit: a Python package for discrete information theory. *J. Open Source Software*, 3(25):738, 2018. 6, 8

[14] A. Gohari and V. Anantharam. Comments on "information-theoretic key agreement of multiple terminals: Part I". *IEEE Trans. Info. Th.*, 63(8):5440–5442, 2017. 6

[15] E. Schneidman, S. Still, M. J. Berry, W. Bialek, et al. Network information and connected correlations. *Phys. Rev. Lett.*, 91(23):238701, 2003. 6

[16] R. G. James, J. Emenheiser, and J. P. Crutchfield. Unique information via dependency constraints. *J. Phys. A*, in press, 2018. arXiv:1709.06653. 6

[17] N. Bertschinger, J. Rauh, E. Olbrich, J. Jost, and N. Ay. Quantifying unique information. *Entropy*, 16(4):2161–2183, 2014. 6, 7

[18] R. A.A. Ince. Measuring multivariate redundant information with pointwise common change in surprisal. *Entropy*, 19(7):318, 2017. 6

[19] E. T. Jaynes. Where do we stand on maximum entropy? In E. T. Jaynes, editor, *Essays on Probability, Statistics, and Statistical Physics*, page 210. Reidel, London, 1983. 7

[20] S. Amari. Information geometry on hierarchy of probability distributions. *IEEE Trans. Info. Th.*, 47(5):1701–1711, 2001. 7

[21] H. McGurk and J. MacDonald. Hearing lips and seeing voices. *Nature*, 264(5588):746, 1976. 8