entropy

MDPI

Review

Unique Information Through the Lens of Channel Ordering: An Introduction and Review

Pradeep Kr. Banerjee

Institute for Data Science Foundations, Blohmstraße 15, 21079 Hamburg, Germany; pradeep.banerjee@tuhh.de

Abstract: The problem of constructing information measures with a well-defined interpretation is of fundamental significance in information theory. A good definition of an information measure entails certain desirable properties while also providing answers to operational problems. In this work, we investigate the properties of the unique information, an information measure that quantifies a deviation from the Blackwell order. Beyond providing an accessible introduction to the topic from a channel ordering perspective, we present a novel resource-theoretic characterization of unique information in a cryptographic task related to secret key agreement. Our operational view of unique information entails rich physical intuition that leads to new insights into secret key agreement in the context of non-negative decompositions of the mutual information into redundant and synergistic contributions. Through this lens, we illuminate new directions for research in partial information decompositions and information-theoretic cryptography.

Keywords: comparison of channels; unique information; Blackwell order; information-theoretic cryptography; secret key rate; secrecy monotones; synergy; redundancy; Le Cam deficiency; resource theories

1. Introduction

Shannon's pioneering work [1] characterized the capacity of a physical channel by way of maximum mutual information. Since then, information theory has had a special relation to communication engineering, even though ideas and tools from information theory have been successfully applied in many other research fields, such as cryptography, statistics, machine learning, complex systems, and biology, to name a few.

Despite significant progress in information theory, many fundamental questions remain regarding the nature of information. One of the primary challenges is that information is not a conserved quantity, making it difficult to track and describe its distribution across composite systems. A composite system consists of multiple interacting subsystems, each of which may hold *unique* (or exclusive) information, or share *redundant* (or shared) information. Additionally, there are cases where some information is not directly accessible to any individual subsystem but can only be determined by considering the entire system. For example, a checksum for a set of digits can only be computed when all the digits are known. Such *synergistic* effects are especially relevant in cryptography, where the objective is for the encrypted message to reveal no information about the original message without the corresponding key.

How should the amount of unique, shared, and synergistic information be measured? This question can be approached from two different points of view, namely, the *axiomatic* and the *operational* [2]. In an *axiomatic* approach, one posits certain desirable properties that a measure of information should satisfy. This point of view goes back to Shannon [1],

check for updates

Academic Editor: Daniel Chicharro

Received: 11 October 2024
Revised: 18 December 2024
Accepted: 26 December 2024
Published: 1 January 2025

Citation: Banerjee, P.K. Unique Information Through the Lens of Channel Ordering: An Introduction and Review. *Entropy* **2025**, *27*, 29. https://doi.org/10.3390/e27010029

who showed that his definition of entropy is the only one that satisfies certain intuitively appealing properties. Shannon notes that such an axiomatic characterization is "*in no way necessary for the theory*" but "*lends a certain plausibility*" to the definitions and that the "*real justification of these definitions, however, will reside in their implications*" [1]. Thus, in Shannon's view, the ultimate criterion for accepting some quantity as a measure of information is whether it provides answers to interesting problems. This is an *operational* or *pragmatic* view of information. For example, Shannon's coding theorems endow the entropy and mutual information with concrete meaning in operational tasks related to data compression and transmission. Rényi [2,3] and Csiszár [4,5] comment that for problems that lay outside the scope of these theorems, both the axiomatic and the operational points of view deserve attention and can, in fact, be used to "control" or inform the other when constructing new measures of information. Understanding the properties of these measures helps clarify the fundamental limits of operational problems. Dually, analyzing such problems motivates the quest for new information measures.

This review investigates the properties of the unique information (UI), an information measure introduced by Bertschinger et al. [6], which quantifies deviations from the Blackwell order. We adopt Shannon's pragmatic stance, focusing on an operational view of unique information from a channel ordering perspective. This builds on the original definition of UI in [6], which is motivated by the idea that unique information should be "useful". Bertschinger et al. formalized this idea in terms of decision problems: Whenever Bob has unique information about something Alice knows (which is not accessible to Eve), there is a decision problem in which Bob has an advantage over Eve. By leveraging tools from resource theories [7–11], we provide a concrete formalization of this conceptual framework.

Resource theories provide an abstract operational framework for studying what physical transformations between a given set of objects are possible under restrictions that follow from the nature of the system under investigation. Within this framework, resources are measured by *monotones*, quantities that do not increase under allowed operations. We present a novel resource-theoretic characterization of UI in a fundamental cryptographic task related to secret key agreement, showing that UI functions similarly to classical *secret key rates* [12–15], and is in fact a monotone that quantifies the "resourcefulness" or secrecy content of a source distribution. This operational characterization not only extends the applicability of UI in cryptographic settings but also opens up new avenues for its study within resource theory.

Our resource-theoretic approach represents a significant departure from existing frameworks on bivariate partial information decompositions, offering a fresh perspective on how UI can be leveraged in the broader context of information-theoretic cryptography. While most existing approaches have focused on the shared information within these decompositions using an axiomatic framework, we shift the focus to an operational view of UI, grounded in the context of channel preorders. Shannon emphasized that the value of an information measure should be judged by its implications in practical tasks, rather than its adherence to abstract properties alone. Following this viewpoint, we argue that UI's significance emerges most clearly when applied in concrete settings like decision-making or cryptographic problems, where operational utility takes precedence over purely axiomatic considerations.

This work serves as both a review and a formalization of existing research on UI and related measures based on channel orderings, with a particular focus on interpreting these insights through the lens of resource theories. The focus of this review is primarily on measures akin to UI, particularly those rooted in channel orderings, and does not extend to other types of information measures that fall outside this framework. We draw extensively from previously published works [16–21] and integrate key insights from unpublished

portions of the author's PhD thesis [22], which are not available in the public domain. By synthesizing these contributions, this review not only provides a comprehensive overview of prior research on UI and channel preorders, but also introduces novel resource-theoretic perspectives, offering a fresh and compelling advancement in the study of bivariate partial information decompositions and information-theoretic cryptography.

Outline. The paper is organized as follows: Section 2 provides a brief review of prior work on non-negative bivariate information decompositions and presents a formal description of the problem, with a focus on the properties of the function UI as introduced by Bertschinger et al. [6]. Section 3 offers a self-contained exposition on channel orderings in information theory. In Section 4, we review Le Cam deficiencies and their generalizations [23–25], which exhibit properties analogous to the UI. These deficiencies quantify the cost of approximating one channel by another through randomization, capturing deviations from output- and input-degraded channel orderings. This provides insight into the distinctions between the bivariate decompositions of Bertschinger et al. [6] and Harder et al. [26]. In Section 5, we review the operational significance of UI in a cryptographic task related to secret key agreement [16,17]. Finally, Section 6 presents a novel resource-theoretic characterization of the main results from Section 5, demonstrating that the UI serves as a resource "monotone" quantifying the secrecy content of a given distribution under a specific class of allowed operations.

2. Bivariate Partial Information Decompositions

Notation and conventions. We shall use notation that is commonly used in information theory [27,28]. We assume that random variables S, Y, Z, etc., are finite, as are all other random variables in this work. The set of all probability measures on a finite set $\mathcal{S}$ is denoted by $\mathbb{P}_{\mathcal{S}}$. A *channel* μ from $\mathcal{S}$ to $\mathcal{Z}$ is a family $\mu = \{\mu_s\}_{s\in\mathcal{S}}$ of probability distributions on $\mathcal{Z}$, one for each possible input $s \in \mathcal{S}$. We write $\mathsf{M}(\mathcal{S};\mathcal{Z})$ to denote the space of all channels from $\mathcal{S}$ to $\mathcal{Z}$. Given two channels, $\mu \in \mathsf{M}(\mathcal{S};\mathcal{Z})$ and $\rho \in \mathsf{M}(\mathcal{Z};\mathcal{Y})$, the composition $\rho \circ \mu \in \mathsf{M}(\mathcal{S};\mathcal{Y})$ of μ with ρ is defined as follows: $\rho \circ \mu_s(y) = \sum_{z\in\mathcal{Z}} \rho_z(y)\mu_s(z)$ for all $s \in \mathcal{S}$, $z \in \mathcal{Z}$. A *binary symmetric channel* with parameter p, denoted as $\mathsf{BSC}(p)$, is a channel from $\mathcal{S} = \{0,1\}$ to $\mathcal{Y} = \{0,1\}$ that flips each bit independently with some error probability $p \in [0, \frac{1}{2}]$. A *binary erasure channel* on $\mathcal{S} = \{0,1\}$ with erasure probability $\epsilon \in [0,1]$, denoted as $\mathsf{BEC}(\epsilon)$, is a channel from $\mathcal{S}$ to $\mathcal{Y} = \mathcal{S} \cup \{\mathsf{e}\}$ such that $Y = S$ with probability $1-\epsilon$ and $Y = \mathsf{e}$ with probability ϵ. Given two distributions P and Q, the Kullback–Leibler (KL) divergence from P to Q is denoted as $D(P\|Q)$. $H(S)$ denotes the Shannon entropy of random variable S. $h(\cdot)$ is the binary entropy function, $h(p) = -p\log p - (1-p)\log(1-p)$ for $p \in (0,1)$ and $h(0) = h(1) = 0$.

The mutual information of two random variables S and Y is defined as

$$I(S;Y) = H(S) + H(Y) - H(SY). \tag{1}$$

I measures the total amount of correlation between S and Y and possesses the following key properties [28,29]:

$$I(S;Y) = I(Y;S) \qquad \textit{(symmetry)}, \tag{2a}$$
$$I(S;Y) \geq 0;\ I(S;Y) = 0 \iff S \textit{ is independent of } Y \qquad \textit{(non-negativity)}, \tag{2b}$$
$$I(S;YZ) \geq I(S;Y) \qquad \textit{(strong subadditivity)}. \tag{2c}$$

Strong subadditivity is equivalent to the non-negativity of the conditional mutual information:

$$I(S;Z|Y) \geq 0;\ I(S;Z|Y) = 0 \iff S - Y - Z,$$

where $S - Y - Z$ denotes that S, Y, and Z form a Markov chain in that order. Strong subadditivity also implies the following key property of the mutual information, namely, its monotonicity with respect to data processing

$$S - Z - Y \implies I(S;Z) \geq I(S;Y), \text{ with equality if and only if } S - Y - Z \quad (\textit{data processing inequality}).$$

Another integral property of the mutual information is the following equality, which is called the chain rule:

$$I(S;YZ) = I(S;Y) + I(S;Z|Y) \qquad \textit{(chain rule).} \tag{3}$$

In general, conditioning on an additional random variable can either increase or decrease the mutual information. We consider three canonical distributions to illustrate this point. Each of these distributions capture a fundamentally different kind of interaction between three jointly distributed random variables.

Example 1. *The* RDN, XOR, *and the* COPY *distributions.*

RDN*: If S, Y, and Z are uniformly distributed binary random variables with $S = Y = Z$, then conditioning on Z decreases the mutual information between S and Y. This is an instance of a purely* redundant *interaction where Y and Z convey the same information about S.*

XOR*: If S and Y are independent binary random variables, and $Z = S \oplus Y$ (where $\oplus$ denotes the binary* XOR *operation), then conditioning on Z increases the mutual information between S and Y. This is an instance of a purely* synergistic *interaction where neither Y nor Z individually conveys any information about S, but jointly, they fully determine S.*

COPY*: If Y and Z are independent uniformly distributed binary random variables, and $S = (Y, Z)$ then $I(S;Y) = I(S;Y|Z) = H(Y) = 1$ bit, and $I(S;Z) = I(S;Z|Y) = H(Z) = 1$ bit. This is an instance of an interaction that is neither redundant nor synergistic, but purely* unique*, for now, Y and Z each uniquely conveys 1 bit of information about S.*

In general, all three forms of interaction—unique, redundant, and synergistic—can coexist simultaneously. Our goal is to disentangle the individual contributions to the mutual information between S and (Y, Z) arising from these interactions. Specifically, we distinguish S as the *target* variable of interest, with Y and Z serving as *predictor* variables.

Let $\widetilde{UI}$, $\widetilde{SI}$, and $\widetilde{CI}$ be non-negative functions that depend continuously on the joint distribution of (S, Y, Z). The mutual information between S and Y can be decomposed into two components: information Y has about S that is *unknown* to Z (referred to as the *unique* or *exclusive* information of Y with respect to Z), and information Y has about S that is *known* to Z (referred to as the *shared* or *redundant* information). This decomposition is given by the following:

$$I(S;Y) = \underbrace{\widetilde{UI}(S;Y\backslash Z)}_{\text{unique } Y \text{ with respect to } Z} + \underbrace{\widetilde{SI}(S;Y,Z)}_{\text{shared (redundant)}}. \tag{4}$$

Conditioning on Z eliminates the shared information but introduces *complementary* (or *synergistic*) information arising from the interaction between Y and Z. This is expressed as follows:

$$I(S;Y|Z) = \underbrace{\widetilde{UI}(S;Y\backslash Z)}_{\text{unique } Y \text{ with respect to } Z} + \underbrace{\widetilde{CI}(S;Y,Z)}_{\text{complementary (synergistic)}}. \tag{5}$$

The unique information can be interpreted either as the conditional mutual information without synergy or as the mutual information without redundancy. Applying the chain rule for mutual information, the total mutual information between S and (Y, Z) can be decomposed into four distinct terms, as illustrated in Figure 1:

$$I(S;YZ) = \widetilde{UI}(S;Y\backslash Z) + \widetilde{SI}(S;Y,Z) + \widetilde{UI}(S;Z\backslash Y) + \widetilde{CI}(S;Y,Z). \tag{6}$$

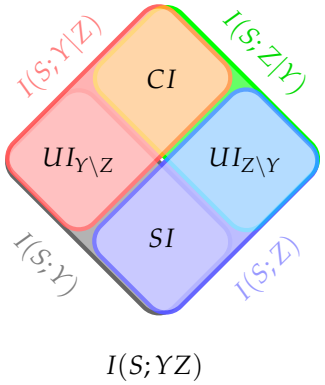

Figure 1. An illustration of the information decomposition in Equations (4)–(6).

Equations (4)–(6) leave only a single degree of freedom; i.e., it suffices to specify either a measure for $\widetilde{SI}$, for $\widetilde{CI}$, or for $\widetilde{UI}$. Any definition of the measure $\widetilde{UI}$ fixes two of the terms in (6), which, in turn, also determines the other terms by (4) and (5). This gives rise to the following *consistency condition*:

$$I(S;Y) + \widetilde{UI}(S;Z\backslash Y) = I(S;Z) + \widetilde{UI}(S;Y\backslash Z). \tag{7}$$

The *coinformation* [30] is defined as the difference between the shared and synergistic information. It serves as a symmetric measure of correlation among three random variables:

$$CoI(S;Y;Z) = \widetilde{SI}(S;Y,Z) - \widetilde{CI}(S;Y,Z) = I(S;Y) - I(S;Y|Z). \tag{8}$$

Coinformation is called *interaction information* (with a change of sign) in [31] and *multiple mutual information* in [32]. The XOR distribution in Example 1 shows that CoI can be negative. Coinformations and entropies are related by a Möbius inversion [30]. Equation (8) can equivalently be written as a linear combinations of entropies:

$$CoI(S;Y;Z) = H(S) + H(Y) + H(Z) - H(SY) - H(SZ) - H(YZ) + H(SYZ). \tag{9}$$

Yeung [33] discusses properties of the CoI as a signed measure using analogies between sets and random variables. Te Sun [34] studies the more general question of what linear combinations of entropies are always non-negative.

Yet, another way to express the CoI is in terms of mutual informations:

$$CoI(S;Y;Z) = I(S;Y) + I(S;Z) - I(S;YZ). \tag{10}$$

Equation (10) shows that CoI can be interpreted as a measure of the "extensivity" of mutual information, i.e., how the mutual information increases as we combine Y and Z [35]: If $CoI = 0$, then the mutual information is exactly extensive in the sense that $I(S;YZ)$ is the sum of the mutual informations $I(S;Y)$ and $I(S;Z)$. If $CoI > 0$, then the mutual information is subextensive and the shared component dominates the synergistic

component. Conversely, if $CoI < 0$, then the mutual information is superextensive and the synergistic component dominates the shared component.

Coinformation is a widely utilized measure in neuroscience and related fields, with positive values interpreted as redundancy and negative values as synergy [36–45]. However, it cannot detect interactions where redundancy and synergy are perfectly balanced [46].

The *correlational importance*, a non-negative measure for evaluating the role of correlations in neural coding [47–49] (see also [50]), aligns conceptually with complementary information. Notably, it can sometimes exceed the total mutual information, as demonstrated in specific examples [51].

Non-negative decompositions of the form (4)–(6) that seek to disentangle the synergistic and redundant contributions to the total information that a pair of predictors convey about the target S were first considered by Williams and Beer [46]. Some notable follow-up works include [6,26,52–59]. For the general case of k finite predictor variables, Williams and Beer proposed the *partial information lattice* framework to decompose the mutual information between the target and predictors into a sum of non-negative terms corresponding to the different ways in which combinations of the predictor variables convey shared, unique, or complementary information about S. The lattice is a consequence of certain natural properties of the shared information, sometimes called the *Williams–Beer axioms*. The underlying idea is that any information about S can be classified according to "who knows what", i.e., which information about S is shared by which subsets of the predictors [59]. Specializing to the bivariate case $(k = 2)$, the Williams–Beer axioms only put crude bounds on the values of the functions $\widetilde{SI}$, $\widetilde{UI}$, and $\widetilde{CI}$ in (4)–(6). Additional axioms have been proposed in [26,60]. See Appendix A for a brief review of these axioms. Unfortunately, some of these axioms contradict each other, and the question for the right axiomatic characterization of shared information is still open.

Bertschinger et al. [6] proposed a pragmatic approach to decompositions of the form (4)–(6) based on the idea that if Y has unique information about S with respect to Z, then there must be a situation or task where such unique information is useful. This idea is formalized in terms of decision problems. We recall the definitions in [6].

Definition 1 ([6]). *For some finite state spaces $\mathcal{Y}$, $\mathcal{Z}$, and $\mathcal{S}$, let $\mathbb{P}_{\mathcal{S}\times\mathcal{Y}\times\mathcal{Z}}$ be the set of all joint distributions of (S, Y, Z). Given $P \in \mathbb{P}_{\mathcal{S}\times\mathcal{Y}\times\mathcal{Z}}$, let*

$$\Delta_P := \{Q \in \mathbb{P}_{\mathcal{S}\times\mathcal{Y}\times\mathcal{Z}} : Q_{SY}(s,y) = P_{SY}(s,y) \text{ and } Q_{SZ}(s,z) = P_{SZ}(s,z)\} \tag{11}$$

denote the set of all joint distributions of (S, Y, Z) that have the same marginals on (S, Y) and (S, Z) as P. The unique information that Y conveys about S with respect to Z is defined as

$$UI(S; Y\backslash Z) = \min_{Q\in\Delta_P} I_Q(S; Y|Z), \tag{12a}$$

where the subscript Q in I_Q denotes the joint distribution on which the function is computed. Specifying (12a) *fixes the other three functions in* (6)*, which are then*

$$UI(S; Z\backslash Y) = \min_{Q\in\Delta_P} I_Q(S; Z|Y), \tag{12b}$$

$$SI(S; Y, Z) = \max_{Q\in\Delta_P} CoI_Q(S; Y; Z), \tag{12c}$$

$$CI(S; Y, Z) = I(S; Y|Z) - UI(S; Y\backslash Z). \tag{12d}$$

The functions UI, SI, and CI are non-negative and satisfy (4)–(6) (and hence (7)). Furthermore, the function SI satisfies the bivariate Williams–Beer axioms [6] (see Appendix A):

$$\begin{aligned} &SI(S;Y,Z) = SI(S;Z,Y) && \text{(symmetry)},\\ &SI(S;Y) = I(S;Y) && \text{(self-redundancy)},\\ &SI(S;Y,Z) \leq SI(S;Y) \text{ with equality if } Z \text{ is a function of } Y && \text{(bivariate monotonicity)}. \end{aligned} \tag{13}$$

The definition of the function UI is rooted in a notion of channel domination due to Blackwell [61]. Intuitively, one channel dominates another if the latter can be "simulated" by the former by some stochastic degradation. UI satisfies the following key property which we call the *Blackwell property* (see Definition 10):

Lemma 1 (Vanishing UI [6], Lemma 6)**.** *For a given joint distribution P_{SYZ}, $UI(S;Y\backslash Z)$ vanishes if and only if there exists a random variable Y' such that $S - Z - Y'$ is a Markov chain and $P_{SY'} = P_{SY}$.*

Blackwell's theorem [61,62] establishes that $UI(S;Y\backslash Z) = 0$ is equivalent to the assertion that, for any decision problem involving the prediction of S, having access to Z provides the same predictive capability as having access to Y (see Theorem 1).

Given $(S,Y,Z) \sim P$, let

$$Q^0(s,y,z) = \begin{cases} \dfrac{P(s,y)P(s,z)}{P(s)}, & \text{if } P(s) > 0,\\ 0, & \text{else.} \end{cases} \tag{14}$$

Observe that $Q^0 \in \Delta_P$. Moreover, Q^0 defines a Markov chain $Y - S - Z$. The following lemma gives conditions under which the function SI vanishes:

Lemma 2 (Vanishing SI [6], Lemma 9)**.** *SI vanishes if and only if $I_{Q^0}(Y;Z) = 0$.*

Lemma 3 characterizes the quantities UI, SI, and CI among alternative definitions of information decompositions.

Lemma 3 ([6], Lemma 3)**.** *Let $\widetilde{UI}(S;Y\backslash Z)$, $\widetilde{UI}(S;Z\backslash Y)$, $\widetilde{SI}(S;Y,Z)$, and $\widetilde{CI}(S;Y,Z)$ be non-negative functions on $\mathbb{P}_{\mathcal{S}\times\mathcal{Y}\times\mathcal{Z}}$ satisfying equations (4)–(6), and assume that the following holds:*

$(*)$ *$\widetilde{UI}$ depends only on the marginal distributions of the pairs (S,Y) and (S,Z).*

Then, $\widetilde{UI} \leq UI$, $\widetilde{SI} \geq SI$, and $\widetilde{CI} \geq CI$ with equality if and only if there exists $Q \in \Delta_P$ such that $\widetilde{CI}_Q(S;Y,Z) = 0$.

By Lemma 3, (12a–12d) is the *only* information decomposition that satisfies $(*)$ and the following property:

$(**)$ For each $P \in \Delta$, there is $Q \in \Delta_P$ with $CI_Q(S;Y,Z) = 0$.

Assumption $(*)$ in Lemma 3 is motivated by the Blackwell property, which also depends only on the marginal distributions of the pairs (S,Y) and (S,Z).

Given $(S,Y,Z) \sim P$, let

$$Q^* \in \operatorname*{arg\,min}_{Q\in\Delta_P} I_Q(S;Y|Z). \tag{15}$$

By definition, $I_{Q^*}(S;Y|Z) = UI(S;Y\backslash Z)$. The distribution Q^* is called a *minimum synergy distribution* as

$$CI_P(S;Y,Z) = 0 \text{ if and only if } P \in \operatorname*{arg\,min}_{Q\in\Delta_P} I_Q(S;Y|Z). \tag{16}$$

Although theoretically promising, the operational significance of the UI is not immediately evident, except in cases where it vanishes, reflecting the Blackwell property. One of our main aims is to address this gap by examining UI's operational relevance in practical model systems through the following key observations:

- The Blackwell relation induces a partial order on channels with the same input alphabet. Most channels are incomparable, meaning one cannot always simulate another by degradation. In such cases, UI quantifies the degree of deviation from simulating one channel by another.
- Weaker notions of channel comparison, such as the "less noisy" property [63], have operational significance through vanishing C_S, where C_S is the secrecy capacity of the wiretap channel [64,65]. Similar to how C_S measures deviation from the less noisy order, a nonvanishing UI quantifies a deviation from the Blackwell order and bounds operational quantities in secret key agreement tasks. In particular, UI acts as a *secrecy monotone*, never increasing under local operations in one-way secret key agreement protocols, making it an upper bound on the *one-way secret key rate* $S_{\rightarrow}$ [66]. This endows the UI with operational significance.
- Finally, the best-known upper bounds on the *two-way secret key rate* $S_{\leftrightarrow}$ involve a secret key decomposition [67]. We show that UI satisfies a similar property, ensuring UI is never greater than the best-known computable upper bound on $S_{\leftrightarrow}$. We conjecture that UI serves as a lower bound on $S_{\leftrightarrow}$ and identify a class of distributions where they coincide.

3. Comparison of Channels

Given two channels that convey information about the same random variable, a natural question is "which channel is better?". Depending on the task at hand, some orderings are more natural or mathematically more appealing than others. For example, ordering channels according to their capacity is often too coarse to be useful in practice. In a seminal paper [61], David Blackwell introduced an ordering of channels in terms of risks of statistical decision rules. Blackwell showed that such an ordering can be equivalently characterized in terms of a purely probabilistic relation between the channels. Blackwell formulated his result in terms of a decision problem, where a decision maker or agent reacts to the outcome of a statistical experiment. In information-theoretic parlance, a statistical experiment is just a noisy channel [4,25]. Shannon [68] independently introduced a criterion for ordering communication channels from a random coding perspective, which is weaker than Blackwell's criterion.

We provide a self-contained introduction to channel orderings in information theory. Such orderings are a well-studied subject in network information theory [69]. For instance, the capacity region of broadcast channels (without feedback) depends only on the component channels and is known for a number of special cases when one of the components is "better" than the other in some well-defined sense (see, e.g., [70,71]).

The Blackwell order. The Blackwell order evaluates channels with a common input alphabet by comparing the minimal expected loss a rational agent incurs when making decisions based on their outputs. This concept is formalized through decision problems under uncertainty (see [24] for an in-depth discussion).

Consider a *decision problem* $(\pi_S, \mathcal{A}, \ell)$, where $\mathcal{A}$ is the set of possible actions, $\ell(s, a)$ represents the bounded loss incurred when the agent chooses action $a \in \mathcal{A}$ in state $s \in \mathcal{S}$, and π_S is the prior distribution over the state space $\mathcal{S}$.

The agent observes a random variable Z via a channel $\mu : \mathcal{S} \to \mathcal{Z}$ before choosing an action. A rational agent selects a strategy $\rho \in \mathsf{M}(\mathcal{Z}; \mathcal{A})$ to minimize the *expected loss* (or *risk*), defined as follows:

$$R(\pi_S, \mu, \rho, \ell) := \sum_{s \in \mathcal{S}} \pi_S(s) \sum_{a \in \mathcal{A}} \rho \circ \mu_s(a) \ell(s, a). \tag{17}$$

The *optimal risk* for channel μ is as follows:

$$R(\pi_S, \mu, \ell) := \min_{\sigma \in \mathcal{A}_\mu} \sum_{s \in \mathcal{S}} \pi_S(s) \sum_{a \in \mathcal{A}} \sigma_s(a) \ell(s, a). \tag{18}$$

where $\mathcal{A}_\mu = \{\rho \circ \mu : \rho \in \mathsf{M}(\mathcal{Z}; \mathcal{A})\}$. Optimal strategies can always be chosen deterministically, so it suffices to consider deterministic strategies.

Now, suppose the agent has access to another random variable Y via a second channel $\kappa \in \mathsf{M}(\mathcal{S}; \mathcal{Y})$ with the same input alphabet $\mathcal{S}$. The agent will *always* prefer Z to Y if, for any decision problem, the optimal risk using Z is no greater than that using Y. This leads to the following definition.

Definition 2. *Given $\mu \in \mathsf{M}(\mathcal{S}; \mathcal{Z})$, $\kappa \in \mathsf{M}(\mathcal{S}; \mathcal{Y})$, and a probability distribution π_S on $\mathcal{S}$ such that $P_{SZ}(s, z) = \pi_S(s)\mu_s(z)$ and $P_{SY}(s, y) = \pi_S(s)\kappa_s(y)$, we say that Z is* always more informative *about S than Y and write $Z \sqsupseteq_S Y$ if $R(\pi_S, \kappa, \ell) \geq R(\pi_S, \mu, \ell)$ for any decision problem (with π_S fixed as above).*

The variables can also be ranked probabilistically: Z is *always* preferred over Y if, given access to Z, a single use of Y can be simulated by sampling $y' \in \mathcal{Y}$ after each observation $z \in \mathcal{Z}$. This implies that Y provides no additional utility beyond what Z already offers.

Definition 3. *Write $Z \sqsupseteq'_S Y$ if there exists a random variable Y' such that the pairs (S, Y) and (S, Y') are statistically indistinguishable, and $S - Z - Y'$ is a Markov chain.*

Intuitively, Z knows everything that Y knows about S in both these situations. Blackwell showed the equivalence of these two relations [61]. The following is a statement of Blackwell's theorem for random variables [62]:

Theorem 1 (Blackwell's theorem for random variables)**.** $Z \sqsupseteq_S Y \iff Z \sqsupseteq'_S Y$.

The original statement of Blackwell's theorem [61] allows us to directly compare the channels κ and μ and the input distribution on $\mathcal{S}$ can be arbitrary.

Definition 4. *We say that μ is* always more informative *than κ and write $\mu \sqsupseteq_{\mathcal{S}} \kappa$ if $R(\pi_S, \kappa, \ell) \geq R(\pi_S, \mu, \ell)$ for any $(\pi_S, \mathcal{A}, \ell)$.*

Definition 5. *We say that κ* is output-degraded *(or* post-garbled*) from μ and write $\mu \sqsupseteq_{\mathcal{S}}^{\mathsf{odeg}} \kappa$ if $\kappa = \lambda \circ \mu$ for some $\lambda \in \mathsf{M}(\mathcal{Z}; \mathcal{Y})$.*

The relation $\sqsupseteq_{\mathcal{S}}^{\mathsf{odeg}}$ is also called the *degradation order* (see, e.g., [72]).

Theorem 2 (Blackwell's Theorem (1953) [61])**.** $\mu \sqsupseteq_{\mathcal{S}} \kappa \iff \mu \sqsupseteq_{\mathcal{S}}^{\mathsf{odeg}} \kappa$.

See [73] for a simple proof of Blackwell's theorem.

If π_S has full support, then $\mu \sqsupseteq_{\mathcal{S}} \kappa \iff Z \sqsupseteq_S Y$ (Theorem 4 in [62]) and it suffices to look only at different loss functions. In the sequel, we assume that π_S has full support, and we call $\sqsupseteq_{\mathcal{S}}$ and $\sqsupseteq_S$ the *Blackwell orders*.

Strictly speaking, the Blackwell order is only a preorder rather than a partial order as there exist channels $\kappa \neq \mu$ that satisfy $\kappa \sqsupseteq_{\mathcal{S}} \mu \sqsupseteq_{\mathcal{S}} \kappa$ (when κ arises from μ by permuting the output alphabet). However, for our purposes, such channels can be treated as equivalent.

We write $\mu \sqsupset_{\mathcal{S}} \kappa$ if $\mu \sqsupseteq_{\mathcal{S}} \kappa$ and $\kappa \not\sqsupseteq_{\mathcal{S}} \mu$. By Blackwell's theorem, this indicates that μ performs at least as well as κ in any decision problem and that there exist decision problems in which μ outperforms κ.

A related order is the *zonotope order*, which is weaker than the Blackwell order [62,74]. For the special case of binary-valued channel inputs, i.e., $|\mathcal{S}| = 2$, the Blackwell order defines a lattice and is identical to the zonotope order [62,74] and its generalization, the k-decision order [61].

The Shannon order. Shannon proposed a criterion for simulating one channel from another based on a random coding argument [68]. Shannon's criterion allows for randomization at *both* the input and the output of the simulating channel as well as for shared randomness between its input and output.

Definition 6 ([68]). *Given two channels $\kappa \in \mathsf{M}(\mathcal{S}';\mathcal{Y})$ and $\mu \in \mathsf{M}(\mathcal{S};\mathcal{Z})$, we say that μ* includes *κ and write $\mu \sqsupseteq^{inc} \kappa$ if for some $k \in \mathbb{N}$, there exists a probability distribution $g \in \mathbb{P}_{[k]}$ and k pairs of pre- and post-channels $(\alpha_i, \beta_i) \in \mathsf{M}(\mathcal{S}';\mathcal{S}) \times \mathsf{M}(\mathcal{Z};\mathcal{Y})$, $1 \le i \le k$, such that $\kappa = \sum_{i=1}^{k} g(i)(\beta_i \circ \mu \circ \alpha_i)$.*

Shannon showed that if $\mu \sqsupseteq^{inc} \kappa$, then the existence of a good coding scheme for κ implies the existence of a good coding scheme for μ, where "goodness" is measured in the sense of low probability of error. Let Σ be the set of all convex combinations of products of the channels in $\mathsf{M}(\mathcal{S}';\mathcal{S})$ with those in $\mathsf{M}(\mathcal{Z};\mathcal{Y})$, i.e.,

$$\Sigma = \operatorname{conv}\left(\alpha \otimes \beta \in \mathsf{M}(\mathcal{S}' \times \mathcal{Z}; \mathcal{S} \times \mathcal{Y}) : \alpha \in \mathsf{M}(\mathcal{S}';\mathcal{S}),\ \beta \in \mathsf{M}(\mathcal{Z};\mathcal{Y})\right), \tag{19}$$

where $\operatorname{conv}(C)$ denotes the convex hull of C, and $(\alpha \otimes \beta)_{s',z}(s,y) = \alpha_{s'}(s)\beta_z(y)$ for each $s \in \mathcal{S}$, $s' \in \mathcal{S}'$, $z \in \mathcal{Z}$, and $y \in \mathcal{Y}$. By Carathéodory's theorem [75], any channel $\chi \in \Sigma$ can be represented as a convex combination of at most $|\mathcal{S}' \times \mathcal{Z} \times \mathcal{S} \times \mathcal{Y}| + 1$ product channels. Given $\mu \in \mathsf{M}(\mathcal{S};\mathcal{Z})$ and $\chi \in \Sigma$, define the *skew-composition* $\chi \circ_s \mu \in \mathsf{M}(\mathcal{S}';\mathcal{Y})$ of μ with χ as follows: $\chi \circ_s \mu(y|s') = \sum_{s \in \mathcal{S}, z \in \mathcal{Z}} \chi_{s',z}(s,y)\mu_s(z)$ for all $s' \in \mathcal{S}'$, $y \in \mathcal{Y}$. We then have the following equivalent characterization of the Shannon order:

Proposition 1 ([76]). *$\mu \sqsupseteq^{inc} \kappa$ if and only if there exists $\chi \in \Sigma$ such that $\kappa = \chi \circ_s \mu$.*

Nasser [76] gave a characterization of the Shannon order that is similar to Blackwell's theorem.

In Definition 6, the input and output alphabets of both κ and μ may be different. If the channels share a common input alphabet, i.e., $\mathcal{S}' = \mathcal{S}$, then $\mu \sqsupseteq_{\mathcal{S}}^{\mathsf{odeg}} \kappa \implies \mu \sqsupseteq^{inc} \kappa$. The converse implication is not true in general and the Shannon order is weaker than the Blackwell order [25].

The input-degraded order. Given two channels that share a common output alphabet, Nasser [77] introduced the following ordering:

Definition 7 ([77]). *Let $\bar{\kappa} \in \mathsf{M}(\mathcal{Y};\mathcal{S})$ and $\bar{\mu} \in \mathsf{M}(\mathcal{Z};\mathcal{S})$ be two channels with a common output alphabet. We say that* $\bar{\kappa}$ is input-degraded from *$\bar{\mu}$ and write $\bar{\mu} \sqsupseteq_{\mathcal{S}}^{ideg} \bar{\kappa}$ if $\bar{\kappa} = \bar{\mu} \circ \bar{\lambda}$ for some $\bar{\lambda} \in \mathsf{M}(\mathcal{Y};\mathcal{Z})$.*

Proposition 2 ([77]).

$$\bar{\mu} \sqsupseteq_{\mathcal{S}}^{ideg} \bar{\kappa} \iff \operatorname{conv}\left(\{\bar{\kappa}_y\}_{y \in \mathcal{Y}}\right) \subset \operatorname{conv}\left(\{\bar{\mu}_z\}_{z \in \mathcal{Z}}\right)$$

where $\operatorname{conv}(C)$ *denotes the convex hull of* C.

Nasser [77] gave a characterization of the input-degraded order that is similar to Blackwell's theorem.

The more capable and less noisy orders. Given two channels $\kappa \in \mathsf{M}(\mathcal{S};\mathcal{Y})$ and $\mu \in \mathsf{M}(\mathcal{S};\mathcal{Z})$ with a common input alphabet, Körner and Marton introduced the following two orderings [63]:

Definition 8. *μ is said to be* more capable *than κ, denoted $\mu \sqsupseteq^{mc} \kappa$, if $I(S;Z) \geq I(S;Y)$ for every probability distribution $P_S \in \mathbb{P}_{\mathcal{S}}$.*

Definition 9. *μ is said to be* less noisy *than κ, denoted $\mu \sqsupseteq^{ln} \kappa$, if $I(U;Z) \geq I(U;Y)$ for every P_{US} such that $U - S - YZ$ is a Markov chain.*

An equivalent characterization of the less noisy relation is the following [78]: $\mu \sqsupseteq^{\mathsf{ln}} \kappa$ if and only if $I(S;Z) - I(S;Y)$ is a concave function of the input probability distribution P_S.

We note the following relationship between the Blackwell, less noisy and the more capable preorders:

Proposition 3 ([63])**.**

$$\mu \sqsupseteq_{\mathcal{S}}^{odeg} \kappa \implies \mu \sqsupseteq^{ln} \kappa \implies \mu \sqsupseteq^{mc} \kappa. \tag{20}$$

As the following examples show, the converse of neither implication is true in general [63].

Example 2 (Broadcast channel consisting of a BSC and a BEC [69,79])**.** *A memoryless broadcast channel model $(\mathcal{S}, \xi_s(y,z), \mathcal{Y} \times \mathcal{Z})$ consists of three sets $\mathcal{S}$, $\mathcal{Y}$, and $\mathcal{Z}$, and a channel $\xi \in \mathsf{M}(\mathcal{S};\mathcal{Y} \times \mathcal{Z})$. Let $\kappa_s(y) := \sum_{z\in\mathcal{Z}} \xi_s(y,z)$ and $\mu_s(z) := \sum_{y\in\mathcal{Y}} \xi_s(y,z)$ be the two components of ξ.*

Consider a broadcast channel with $\kappa = \mathsf{BSC}(p)$ with crossover probability $p \in (0, 1/2)$, and $\mu = \mathsf{BEC}(\epsilon)$ with erasure probability $\epsilon \in (0,1)$. Then, the following hold:

1. *For $0 < \epsilon \leq 2p$, Y is output-degraded from Z.*
2. *For $2p < \epsilon \leq 4p(1-p)$, Z is less noisy than Y, but Y is not output-degraded from Z.*
3. *For $4p(1-p) < \epsilon \leq h(p)$, Z is more capable than Y, but not less noisy.*
4. *For $h(p) < \epsilon < 1$, ξ does not belong to any of the three classes.*

Example 3 (Doubly symmetric binary erasure (DSBE) source [12,80])**.** *A DSBE source with parameters (p,ϵ) is defined as follows: $P_{SYZ}(s,y,z) = P_{SY}(s,y)p_{Z|SY}(z|s,y)$ where $P_{SY}(0,0) = P_{SY}(1,1) = p/2$, $P_{SY}(0,1) = P_{SY}(1,0) = (1-p)/2$, and $P_{Z|SY}(z|s,y)$ is an erasure channel, i.e., $Z = SY$ with probability $1-\epsilon$ and $Z = \mathsf{e}$ with probability ϵ. Without loss of generality, we may assume $p > \frac{1}{2}$. Then, the following hold:*

1. *For $0 < \epsilon \leq 2(1-p)$, Y is output-degraded from Z.*
2. *For $2(1-p) < \epsilon \leq 4p(1-p)$, Z is less noisy than Y, but Y is not output-degraded from Z.*
3. *For $4p(1-p) < \epsilon \leq h(p)$, Z is more capable than Y, but not less noisy.*
4. *For $h(p) < \epsilon < 1$, a $\mathsf{DSBE}(p,\epsilon)$ source does not belong to any of the three classes.*

4. Unique Information and Channel Deficiencies

How can we determine whether Y possesses unique information about S that is not available to Z? Consider the channels κ and μ with a *common input* alphabet $\mathcal{S}$, as illustrated in Figure 2a. If μ can be reduced to κ by appending a post-channel λ at its *output*, then μ can be said to *include* κ. Similarly, for the channels $\bar{\kappa}$ and $\bar{\mu}$ with a *common output* alphabet $\mathcal{S}$, as shown in Figure 2b, $\bar{\mu}$ can be considered to include $\bar{\kappa}$ if it reduces to $\bar{\kappa}$ by adding a pre-channel $\bar{\lambda}$ at its *input*.

In both cases, one would expect Y to provide no unique information about S relative to Z. A nonzero unique information would then serve as a measure of the extent to which one channel deviates from being an inclusion or randomization of the other.

Figure 2. (**a**) Simulation of the channel κ through a randomization at the output of μ, where κ and μ share a *common input* alphabet $\mathcal{S}$. (**b**) Simulation of the channel $\bar{\kappa}$ through a randomization at the input of $\bar{\mu}$, where $\bar{\kappa}$ and $\bar{\mu}$ share a *common output* alphabet $\mathcal{S}$.

The function UI in Definition 1 is based on the idea of approximating one channel by randomizing its *output* (see Figure 2a). In contrast, Harder et al. [26] defined a measure of shared information through a difference in two KL divergence terms, where one term involves randomization at the *input* (see Figure 2b). In both cases, the resulting decompositions of the total mutual information are non-negative.

Banerjee et al. [16] introduce two quantities that generalize Le Cam's notion of *weighted deficiency* [23–25] between channels. Weighted deficiencies quantify the cost of approximating one channel from another via randomizations and are closely related to the function UI. Depending on whether the randomization occurs at the output or input, two different forms of weighted deficiency arise: the *weighted output KL deficiency* and the *weighted input KL deficiency*. Both of these induce non-negative bivariate decompositions [16]. Interestingly, the decomposition corresponding to the weighted input deficiency coincides with the one introduced by Harder et al. [26] (see Proposition 8).

4.1. Generalized Le Cam Deficiencies

The Blackwell order provides a natural criterion to determine if a variable Y has unique information about S with respect to Z or not; see Definitions 2 and 3.

Definition 10 (Blackwell property)**.** *Y has no unique information about S with respect to $Z :\Longleftrightarrow Z \sqsupseteq'_S Y$.*

The function UI satisfies the Blackwell property (see Lemma 1). When $UI(S;Y\backslash Z)$ vanishes, we say that Z is *Blackwell-sufficient* for Y with respect to S.

Theorem 1 states that if the relation $Z \sqsupseteq_S Y$ (resp. $Y \sqsupseteq_S Z$) does not hold, then there exist a loss function and a set of actions that render Y (resp. Z) more useful. This statement motivates the following definition [6]:

Definition 11. *Y has unique information about S with respect to Z if there exists a set of actions $\mathcal{A}$ and a loss function $\ell(s,a) \in \mathbb{R}^{\mathcal{S}\times\mathcal{A}}$ such that $R(\pi_S,\kappa,\ell) < R(\pi_S,\mu,\ell)$.*

The relation $\sqsupseteq_{\mathcal{S}}^{\mathsf{odeg}}$ is a preorder on the family of all channels with the same input alphabet $\mathcal{S}$ (see Definition 5). In general, we cannot always simulate one channel by a randomization of the other. To be able to compare any two channels, Lucien Le Cam introduced the notion of channel *deficiencies* [23,24]:

Definition 12. *Given $\mu \in \mathsf{M}(\mathcal{S};\mathcal{Z})$ and $\kappa \in \mathsf{M}(\mathcal{S};\mathcal{Y})$, the* Le Cam deficiency of μ with respect to κ *is*

$$\delta(\mu,\kappa) := \inf_{\lambda\in\mathsf{M}(\mathcal{Z};\mathcal{Y})} \sup_{s\in\mathcal{S}} \|\lambda\circ\mu_s - \kappa_s\|_{\mathsf{TV}}. \tag{21}$$

where $\|\cdot\|_{\mathsf{TV}}$ *denotes the total variation distance.*

Note that $\delta(\mu,\kappa)=0$ if and only if $\mu \sqsupseteq_{\mathcal{S}}^{\mathsf{odeg}} \kappa$.

Definition 13. *Given* $\mu \in \mathsf{M}(\mathcal{S};\mathcal{Z})$, $\kappa \in \mathsf{M}(\mathcal{S};\mathcal{Y})$ *and a probability distribution* π_S *on* $\mathcal{S}$, *the* weighted Le Cam deficiency of μ with respect to κ *is*

$$\delta^{\pi}(\mu,\kappa) := \inf_{\lambda \in \mathsf{M}(\mathcal{Z};\mathcal{Y})} \mathbb{E}_{s\sim\pi_S} \|\lambda \circ \mu_s - \kappa_s\|_{\mathsf{TV}}. \tag{22}$$

The *Le Cam randomization criterion* [23] establishes that deficiencies quantify the maximal gap in optimal risks between decision problems when using the channel μ instead of κ.

Theorem 3 ([23]). *Fix* $\mu \in \mathsf{M}(\mathcal{S};\mathcal{Z})$, $\kappa \in \mathsf{M}(\mathcal{S};\mathcal{Y})$, *and a probability distribution* π_S *on* $\mathcal{S}$, *and write* $\|\ell\|_\infty = \max_{s,a} \ell(s,a)$. *For every* $\epsilon > 0$, $\delta^{\pi}(\mu,\kappa) \leq \epsilon$ *if and only if* $R(\pi_S,\mu,\ell) - R(\pi_S,\kappa,\ell) \leq \epsilon\|\ell\|_\infty$ *for any set of actions* $\mathcal{A}$ *and any bounded loss function* ℓ.

Raginsky [25] introduced a broad class of deficiency-like quantities based on a "generalized" divergence between probability distributions that maintains a monotonicity property with respect to data processing. Specializing this to the KL divergence, we have the following definition:

Definition 14. *The* output KL deficiency of μ with respect to κ *is*

$$\delta_o(\mu,\kappa) := \inf_{\lambda \in \mathsf{M}(\mathcal{Z};\mathcal{Y})} \sup_{s\in\mathcal{S}} D(\kappa_s \| \lambda \circ \mu_s), \tag{23}$$

where the subscript o *in* δ_o *emphasizes the fact that the randomization is at the* output *of the channel* μ.

In a spirit similar to [25] and Section 6.2 in [24], one can define a weighted output KL deficiency [16]:

Definition 15. *The* weighted output KL deficiency of μ with respect to κ *is*

$$\delta_o^{\pi}(\mu,\kappa) := \min_{\lambda \in \mathsf{M}(\mathcal{Z};\mathcal{Y})} D(\kappa \| \lambda \circ \mu | \pi_S). \tag{24}$$

The weighted output KL deficiency quantifies the cost of approximating one observed variable from the other (and vice versa) through Markov kernels. Notably, $\delta_o^{\pi}(\mu,\kappa)=0$ if and only if $Z \sqsupseteq'_S Y$, capturing the intuition that a small value of $\delta_o^{\pi}(\mu,\kappa)$ implies that Z is *approximately Blackwell-sufficient* for Y with respect to S. Using Pinsker's inequality, we obtain the following:

$$\delta^{\pi}(\mu,\kappa) \leq \sqrt{\tfrac{\ln(2)}{2}\delta_o^{\pi}(\mu,\kappa)}. \tag{25}$$

Bounding the weighted output KL deficiency is sufficient to guarantee that the differences in optimal risks remain bounded for any decision problem of interest [16]:

Proposition 4. *Fix* $\mu \in \mathsf{M}(\mathcal{S};\mathcal{Z})$, $\kappa \in \mathsf{M}(\mathcal{S};\mathcal{Y})$, *and a prior probability distribution* π_S *on* $\mathcal{S}$, *and write* $\|\ell\|_\infty = \max_{s,a} \ell(s,a)$. *For every* $\epsilon > 0$, *if* $\delta_o^{\pi}(\mu,\kappa) \leq \epsilon$, *then* $R(\pi_S,\mu,\ell) - R(\pi_S,\kappa,\ell) \leq \sqrt{\epsilon\frac{\ln(2)}{2}}\|\ell\|_\infty$ *for any set of actions* $\mathcal{A}$ *and any bounded loss function* ℓ.

Recall the data processing inequality for the mutual information:

$$Z - Y - W \implies I(Z;W) \leq \min\{I(Z;Y), I(Y;W)\}. \tag{26}$$

Lemma 4 shows that the weighted output KL deficiency satisfies a similar inequality:

Lemma 4. *Let $\mu \in \mathsf{M}(\mathcal{S};\mathcal{Z})$, $\kappa \in \mathsf{M}(\mathcal{S};\mathcal{Y})$, and $\nu \in \mathsf{M}(\mathcal{S};\mathcal{W})$ be three channels with a common input alphabet and let π_S be a given distribution on $\mathcal{S}$. Then,*

$$Z \sqsupseteq_S Y \sqsupseteq_S W \implies \delta_o^\pi(\mu,\nu) \leq \min\{\delta_o^\pi(\mu,\kappa), \delta_o^\pi(\kappa,\nu)\}.$$

See Appendix B for a proof. One can also define a weighted deficiency for the input-degraded order in Definition 7 [16].

Definition 16. *The* weighted input KL deficiency of $\bar{\mu}$ with respect to $\bar{\kappa}$ *is*

$$\delta_i^\pi(\bar{\mu},\bar{\kappa}) := \min_{\bar{\lambda}\in\mathsf{M}(\mathcal{Y};\mathcal{Z})} D(\bar{\kappa}\|\bar{\mu}\circ\bar{\lambda}|\pi_Y), \tag{27}$$

where the subscript i in δ_i emphasizes the fact that the randomization is at the input *of the channel $\bar{\mu}$.*

The weighted input KL deficiency satisfies the following monotonicity property:

Lemma 5. *Let $\bar{\mu} \in \mathsf{M}(\mathcal{Z};\mathcal{S})$, $\bar{\kappa} \in \mathsf{M}(\mathcal{Y};\mathcal{S})$, and $\bar{\nu} \in \mathsf{M}(\mathcal{W},\mathcal{S})$ be three channels with a common output alphabet, and let π_W be a given distribution on $\mathcal{W}$. Then,*

$$\bar{\mu} \sqsupseteq_{\mathcal{S}}^{ideg} \bar{\kappa} \implies \delta_i^\pi(\bar{\mu},\bar{\nu}) \leq \delta_i^\pi(\bar{\kappa},\bar{\nu}).$$

The proof is similar to the first part of the proof of Lemma 4 and is omitted.

4.2. Non-Negative Mutual Information Decompositions

Given an information measure that captures some aspect of unique information but does not satisfy the consistency condition (7), we can construct the corresponding bivariate information decomposition as follows:

Lemma 6 ([81], Proposition 9)**.** *Let $\delta : \mathbb{P}_{\mathcal{S}\times\mathcal{Y}\times\mathcal{Z}} \to \mathbb{R}$ be a non-negative function that satisfies*

$$\delta(S; Y \setminus Z) \leq \min\{I(S;Y), I(S;Y|Z)\}.$$

Then, a bivariate information decomposition is given by

$$\begin{aligned}
UI_\delta(S;Y\setminus Z) &= \max\left\{\delta(S;Y\setminus Z), \delta(S;Z\setminus Y) + I(S;Y) - I(S;Z)\right\},\\
UI_\delta(S;Z\setminus Y) &= \max\left\{\delta(S;Z\setminus Y), \delta(S;Y\setminus Z) + I(S;Z) - I(S;Y)\right\},\\
SI_\delta(S;Z,Y) &= \min\left\{I(S;Y) - \delta(S;Y\setminus Z), I(S;Z) - \delta(S;Z\setminus Y)\right\},\\
CI_\delta(S;Z,Y) &= \min\left\{I(S;Y|Z) - \delta(S;Y\setminus Z), I(S;Z|Y) - \delta(S;Z\setminus Y)\right\}.
\end{aligned}$$

We refer to the construction in Lemma 6 as the *UI construction*. The unique information UI_δ generated by this construction is the smallest UI function among all bivariate information decompositions with $UI \geq \delta$.

This construction can be used to derive new non-negative bivariate decompositions.

4.2.1. Decomposition Based on the Weighted Output KL Deficiency

Proposition 5 ([16]). *Let $(S, Y, Z) \sim P$, and let π_S be the marginal distribution of S. Let $\kappa \in \mathsf{M}(\mathcal{S}; \mathcal{Y})$ resp. $\mu \in \mathsf{M}(\mathcal{S}; \mathcal{Z})$ be two channels describing the conditional distribution of Y resp. Z, given S. Define*

$$\delta(S; Y \setminus Z) = \delta_o^{\pi}(\mu, \kappa), \tag{28}$$

where δ_o is the weighted output KL deficiency (24). Then, the functions UI_δ, SI_δ, and CI_δ in Lemma 6 define a non-negative bivariate decomposition.

Lemma 7 ([16]). *Define*

$$UI_o(S; Y \backslash Z) = \max\{\delta_o^{\pi}(\mu, \kappa), \delta_o^{\pi}(\kappa, \mu) + I(S; Y) - I(S; Z)\}. \tag{29}$$

Then, $UI_o(S; Y \backslash Z)$ vanishes if and only if Y has no unique information about S with respect to Z (according to Definition 10).

From Lemma 3, we have the following relationship between the different quantities:

Lemma 8.

$$\delta_o^{\pi}(\mu, \kappa) \leq UI_o(S; Y \backslash Z) \leq UI(S; Y \backslash Z),$$

The next proposition follows from Lemmas 1 and 7, and Definition 15.

Proposition 6.

$$\delta_o^{\pi}(\mu, \kappa) = 0 \iff UI_o(S; Y \backslash Z) = 0 \iff UI(S; Y \backslash Z) = 0.$$

4.2.2. Decomposition Based on the Weighted Input KL Deficiency

Proposition 7 ([16]). *Let $(S, Y, Z) \sim P$, and let π_Y resp. π_Z be the induced marginal distributions of Y resp. Z, both assumed to have full support. Let $\bar{\kappa} \in \mathsf{M}(\mathcal{Y}; \mathcal{S})$ and $\bar{\mu} \in \mathsf{M}(\mathcal{Z}; \mathcal{S})$ be two channels such that $\bar{\kappa} = P_{S|Y}$ and $\bar{\mu} = P_{S|Z}$. Define*

$$\delta(S; Y \setminus Z) = \delta_i^{\pi}(\bar{\mu}, \bar{\kappa}), \tag{30}$$

where δ_i is the weighted input KL deficiency (27). Then, the functions UI_δ, SI_δ, and CI_δ in Lemma 6 define a non-negative bivariate decomposition.

Harder et al. [26] introduced a measure of *shared information* based on reverse information (rI) projections [82] onto a convex set of probability measures.

Definition 17. *For $C \subset \mathbb{P}_{\mathcal{S}}$, let* conv$(C)$ *denote the convex hull of C. Let*

$$Q_{y \searrow Z}(S) \in \operatorname*{arg\,min}_{Q \in \mathrm{conv}(\{\bar{\mu}_z\}_{z \in \mathcal{Z}}) \subset \mathbb{P}_{\mathcal{S}}} D(\bar{\kappa}_y \| Q)$$

*be the rI-*projection *of $\bar{\kappa}_y$ onto the convex hull of the points $\{\bar{\mu}_z\}_{z \in \mathcal{Z}} \in \mathbb{P}_{\mathcal{S}}$. Define the* projected information of Y onto Z with respect to S *as*

$$I_S(Y \searrow Z) := \mathbb{E}_{(s,y) \sim \bar{\kappa} \times \pi_Y} \log \frac{Q_{y \searrow Z}(s)}{\bar{\kappa} \circ \pi_Y(s)}, \tag{31}$$

and the shared information as

$$SI_{red}(S;Y,Z) := \min\{I_S(Y \searrow Z), I_S(Z \searrow Y)\}. \tag{32}$$

Proposition 8 states that implicit in the above construction is the weighted input KL deficiency $\delta_i^\pi(\bar{\mu},\bar{\kappa})$.

Proposition 8 ([16]). $I_S(Y \searrow Z) = I(S;Y) - \delta_i^\pi(\bar{\mu},\bar{\kappa})$.

An immediate consequence of Proposition 8 is that the decomposition proposed by Harder et al. [26] and that in Proposition 7 are equivalent.

Remark 1 (SI_{red} is not continuous). *$I_S(Y \searrow Z)$ and $I_S(Z \searrow Y)$ are defined in terms of conditional probability $\bar{\kappa}_y = P_{S|Y=y}$ and $\bar{\mu}_z = P_{S|Z=z}$, which are only defined for those y, z with $\pi_Y(y) > 0$ and $\pi_Z(z) > 0$. Therefore, $I_S(Y \searrow Z)$ and $I_S(Z \searrow Y)$ are discontinuous when probabilities tend to zero. For a concrete example, see Example 3 in* [18].

Remark 2 (Vanishing sets of *UI* and deficiencies). *The Blackwell order compares two channels with a* common input *alphabet. This order has found applications in network information theory* [69,79]. *In wiretap channel models* [64,65] *(see Section 5.2), one considers a memoryless broadcast channel $\xi : \mathcal{S} \to \mathcal{Y} \times \mathcal{Z}$ where Alice selects the inputs to ξ, while Bob and Eve observe, resp., the Y-outputs and the Z-outputs. Bob's component channel is defined as $\kappa_s(y) = \sum_{z\in\mathcal{Z}} \xi_s(y,z)$ and Eve's as $\mu_s(z) = \sum_{y\in\mathcal{Y}} \xi_s(y,z)$. The secrecy capacity of the wiretap channel, C_S, quantifies a deviation from the less noisy order and depends on ξ only through the component channels κ and μ (see Proposition 9). Likewise, when the distribution of the input to ξ is fixed, the UI and weighted output deficiency δ_o^π quantify a deviation from the Blackwell order and depend on ξ only through κ and μ. Proposition 6 shows that the sets on which UI and δ_o^π vanish are the same.*

On the other hand, the weighted input deficiency δ_i^π quantifies a deviation from the input-degraded order, which compares two channels with a common output *alphabet. This ordering appears more natural in some settings, e.g., when learning a classifier (see, e.g.,* [81]*). We can again define a channel model $\bar{\xi} : \mathcal{Y} \times \mathcal{Z} \to \mathcal{S}$. The associated component channels $\bar{\kappa}(s|y)$ and $\bar{\mu}(s|z)$ are, however, not uniquely determined by $\bar{\xi}$ (also see Remark 1). In Theorem 22 of* [6], *it was claimed that the vanishing sets of δ_i^π and UI coincide. However, Banerjee et al. [83] showed that this assertion is incorrect (see Example 28b in [83]).*

Remark 3 (Decompositions based on known bounds on the secret key rates). *In Section 5, we show that the function UI shares conceptual similarities with secret key rates* [12,84]. *The UI construction can be used to obtain bivariate information decompositions from the one-way ($S_\rightarrow$,* (42)*) and two-way secret key rates ($S_\leftrightarrow$), as well as from related information functions defined as bounds on these rates. These functions include the secrecy capacity of the wiretap channel C_S* (36), *the* intrinsic information $I_\downarrow$ (47), *the* reduced intrinsic information $I_{\downarrow\downarrow}$ (50), *and the* minimum intrinsic information B_1 (52). *Each of these bounds can be expressed as optimization problems over Markov kernels of bounded size. For the complete chain of inequalities, see* (57). *Like UI, both $S_\rightarrow$ and C_S depend solely on the marginal distributions of the pairs (S,Y) and (S,Z). However, unlike UI, none of these functions satisfy the consistency condition* (7). *Nevertheless, since these bounds are upper-bounded by $\min\{I(S;Y), I(S;Y|Z)\}$, we can utilize the UI construction outlined in Lemma 6 to derive new non-negative decompositions. An analysis of the properties of these decompositions is reserved for future study.*

5. Unique Information and Secrecy Monotones

The contents of this section have a distinct cryptographic flavor. Our main goal is to establish the operational significance of the UI in Definition 1. In order to keep the exposition reasonably self-contained, we collect all relevant definitions and models in Section 5.2. Theorem 8, the triangle inequality for the UI (Property P.7), and Theorem 9 are the main results in this section. Theorem 8 was first derived in [16], while the contents of Sections 5.4 and 5.5 expands on the work in [16,17].

5.1. Motivation and Synopsis

Consider the *source model* for secret key agreement between Alice and Bob, who are distant from each other and must communicate over a noiseless but insecure (public) channel in the presence of an adversary, Eve [12,85]. Alice, Bob, and Eve observe i.i.d. copies of random variables S, Y, and Z, respectively, where $(S, Y, Z) \sim P$. Alice and Bob aim to agree on a secret key by exchanging messages over the public channel according to a predefined protocol. Eve is aware of the protocol and can intercept and read all the messages exchanged. The maximum rate at which Alice and Bob can compute a key such that Eve's total information (from both Z and the entire communication) about the key is negligibly small is referred to as the *two-way secret key rate*, $S_{\leftrightarrow}$. If Alice is allowed to send only one message and Bob sends none, the corresponding key rate is called the *one-way secret key rate*, $S_{\rightarrow}$.

The secret key rates are conceptually similar to the function UI. While $UI(S; Y \backslash Z)$ is interpreted as the *information about S known to Y, but not to Z*, $S_{\leftrightarrow}(S; Y \| Z)$ can be interpreted as the *information common to S and Y, which is unique with respect to Z*.

For example, consider the RDN distribution from Example 1, where Alice, Bob, and Eve each share one uniformly random bit. In this case, since Eve knows the exact values of S and Y, Alice and Bob cannot share a secret. This is reflected in the values of $UI(S; Y \backslash Z)$ and $UI(Y; S \backslash Z)$, both of which are zero.

As another example, consider the XOR distribution in Example 1, where the values of any two variables in (S, Y, Z) determine the third. Clearly, if Alice can only observe S and Bob can only observe Y, they cannot generate a secret key. This is also apparent from the values of $UI(S; Y \backslash Z)$ and $UI(Y; S \backslash Z)$, both of which are zero. However, if Alice is also able to observe Z, she can compute Y, which can then be used as a key that is perfectly secret from Eve, since Eve's variable Z is independent of the key Y.

Intuitively, when Alice and Bob share some common information that is unique with respect to Eve, they can *exploit* this information to generate a secret key. A distribution combining elements of the XOR and RDN models exemplifies the potential advantage of such a setup:

Example 4 (The XORRDN distribution [12,84]). *Consider the following distribution:* $P_{SYZ}(0,0,0) = P_{SYZ}(0,1,1) = P_{SYZ}(1,0,1) = P_{SYZ}(1,1,0) = \frac{1}{8}$, *and* $P_{SYZ}(2,2,2) = P_{SYZ}(3,3,3) = \frac{1}{4}$. *The table below shows the distribution (with Z's value in parentheses):*

	S			
Y (Z)	0	1	2	3
0	1/8 (0)	1/8 (1)	.	.
1	1/8 (1)	1/8 (0)	.	.
2	.	.	1/4 (2)	.
3	.	.	.	1/4 (3)

If Eve observes 2 *or* 3*, she can determine the exact values of S and Y. When she observes* 0 *or* 1*, she can infer that Alice and Bob's values lie within* $\{0,1\}$*, but within this range, their observations are independent. Consequently, no secret key agreement is possible in this case. This is reflected in the values of* $UI(S;Y\backslash Z)$ *and* $UI(Y;S\backslash Z)$*, both of which are zero.*

Consider now the modified distribution: $P_{SYZ}(0,0,0) = P_{SYZ}(0,1,1) = P_{SYZ}(1,0,1) = P_{SYZ}(1,1,0) = \frac{1}{8}$*, and* $P_{SYZ}(2,2,0) = P_{SYZ}(3,3,1) = \frac{1}{4}$*, where Eve's variable Z can only assume binary values.*

	S			
Y (Z)	0	1	2	3
0	1/8 (0)	1/8 (1)	.	.
1	1/8 (1)	1/8 (0)	.	.
2	.	.	1/4 (0)	.
3	.	.	.	1/4 (1)

Now Bob (resp. Alice) has 1 bit of unique information about Alice's (resp. Bob's) values with respect to Eve (namely, the ability to distinguish whether Alice sees values in the XOR *or the* RDN *quadrant) which can be used to agree on 1 bit of secret.*

A computable characterization of the one-way secret key rate is known [66] (see Theorem 6). In contrast, determining the two-way key rate for a given distribution, or even the condition when it is positive, seems difficult, and its value is known only for a handful of distributions [66,67,86,87]. For protocols with unbounded communication, computing the two-way key rate for a general distribution is a fundamental and open area of inquiry in information-theoretic cryptography.

A standard technique for deriving upper bounds on the two-way key rate is to consider functions of joint distributions called *secrecy monotones* or simply *monotones*, which satisfy the following property: In any secret key agreement protocol, a monotone can *never increase* if Alice and Bob are only allowed to perform a well-defined class of physical operations called local operations (LOs) and public communication (PC) [14,67,88]. Theorem 8 shows that the UI is an upper bound on the one-way secret key rate. This is a consequence of the fact that the function UI is a monotone when the class of allowed operations is local operations and one-way public communication.

The state-of-the-art upper bounds on the two-way key rate are based on the following key property (see Theorem 4 in [67]): For any tuple (S,Y,Z,Z'),

$$S_{\leftrightarrow}(S;Y|Z) \leq S_{\leftrightarrow}(S;Y|Z') + S_{\rightarrow}(SY;Z'|Z). \tag{33}$$

In [67,89], a heuristic interpretation of this decomposition is provided: Let $s = S_{\leftrightarrow}(S;Y|Z)$. Consider a fourth party, Charlie, who receives i.i.d. copies of Z' but does not have access to the public channel. If we decompose s into two parts: s_1, which Charlie does not know, and $s_2 = s - s_1$, which Charlie knows about the shared secret key between S and Y with respect to Z, then s_1 is at most $S_{\leftrightarrow}(S;Y|Z')$, while s_2 is at most $S_{\rightarrow}(SY;Z'|Z)$.

Gohari et al. [80] gave an alternative interpretation of (33): For any $(S,Y,Z,Z')\sim P$, if the induced channel $P_{Z|SY}$ dominates $P_{Z'|SY}$ in the *less noisy* sense (see Definition 9), then the second term $S_{\rightarrow}(SY;Z'|Z)$ vanishes. Thus, $S_{\rightarrow}(SY;Z'|Z)$ represents the "penalty" for deviating from the less noisy condition when substituting $P_{Z|SY}$ with $P_{Z'|SY}$.

The function UI satisfies a triangle inequality, which implies the following property that resembles (33): For any (S,Y,Z,Z'),

$$UI(S;Y\backslash Z) \leq UI(S;Y\backslash Z') + UI(SY;Z'\backslash Z). \tag{34}$$

From (34), we conclude that the UI is never greater than the best-known computable upper bound on $S_{\leftrightarrow}$. We also give an example where the UI is not lower than the best-known lower bound on the two-way rate. We conjecture that the UI lower-bounds the two-way key rate and discuss implications of the conjecture.

5.2. Information-Theoretic Secrecy Models

We begin by reviewing some fundamental models in information-theoretic cryptography. Some excellent references include [13–15] and Section 17.3 in [27].

Suppose that Alice wishes to transmit a message to Bob over a *noiseless* channel such that an adversary, Eve, who has access to the channel, obtains no information about the message. The channel is assumed to be *authenticated* in the sense that Eve has only read access to the channel and cannot modify or insert messages without being detected. Authentication can be guaranteed, for instance, if Alice and Bob initially share a short secret key [90]. The assumption that the channel is noiseless entails no loss of practicality if we assume that powerful error correction schemes exist, so that the message can be recovered with an arbitrarily small probability of error. This assumption is convenient because it allows us to focus solely on secrecy without having to worry about communication efficiency. We will call such a noiseless and authenticated channel the *public channel*. The terminology is, of course, suggestive of the fact that the channel is insecure. While it is often impractical to assume that a secure channel (e.g., a trusted courier) is always available whenever such a need arises, without loss of generality, we will assume that insecure public channels (e.g., telephone lines) are always available.

The Shannon model. Shannon introduced a simple model of a cryptosystem [91] as follows. Let random variables $M \in \mathcal{M}$ and $C \in \mathcal{C}$ model, resp., the message and the codeword or ciphertext. Alice and Bob share a common secret key modeled by a random variable $K \in \mathcal{K}$. We assume that K is independent of M. Let $e : \mathcal{M} \times \mathcal{K} \to \mathcal{C}$ and $d : \mathcal{C} \times \mathcal{K} \to \mathcal{M}$ denote, resp., Alice's encoding and Bob's decoding function. The pair (e, d) is called a *coding scheme*. We assume that Eve has no knowledge of the key but knows the coding scheme and that Bob can decode messages without error, i.e., $M = d(C, K)$ if $C = e(M, K)$. Alice encodes M into a ciphertext C using the secret key before sending it over the public channel. Since the channel is public, Eve receives an identical copy of C as Bob. A coding scheme is said to achieve *perfect secrecy* if Eve's equivocation about the message given the ciphertext as measured by the conditional entropy $H(M|C)$ equals her a priori uncertainty about the message, i.e., $H(M|C) = H(M)$, or, equivalently $I(M;C) = 0$. Shannon gave a necessary condition for communication in perfect secrecy.

Theorem 4 ([91]). *If a coding scheme achieves perfect secrecy, then* $H(K) \geq H(M)$.

To see this, note that by assumption, $H(M) = H(M|C)$. Since Bob can decode messages without error, we have $H(M|CK) = 0$. The claim follows from $H(M) = H(M|C) = H(M|C) - H(M|CK) = I(K;M|C) \leq H(K|C) \leq H(K)$.

From an algorithmic perspective, perfect secrecy can be realized using a public channel and a secret key by means of a simple coding scheme called the *one-time pad* (OTP) [92]:

Example 5 (OTP). *The message M is a l-bit string and the key K is a uniformly distributed l-bit string which is independent of the message. Alice computes $C = M \oplus K$ and Bob computes $C \oplus K$, where $\oplus$ denotes a bit-wise* XOR *operation. Alice's encoding guarantees that $H(C) = l$. Also, $H(C|M) = H(K|M) = H(K) = l$, since there is a one-to-one mapping between C and K given M, and K is independent of M. We thus have $I(M;C) = H(C) - H(C|M) = 0$, which shows that the OTP achieves perfect secrecy.*

The OTP guarantees that Eve can do no better than randomly guess M and that there exists *no* algorithm that could extract any information about M from C. The OTP is *unconditionally secure* in the sense that this is true even when Eve has unlimited computing power. The OTP is also provably secure in the sense that very precise statements can be made about the information that is leaked to Eve under some well-defined notion of statistical independence (or near-independence) of the message from Eve's observations.

Contrast this with *computationally secure* cryptosystems which are based on computational complexity theory [93,94]. The security of these systems is based on the following assumptions: (a) Eve's computational resources, specified by some model of computation, are bounded and, (b) certain one-way functions exist that are computationally "hard" to invert (see Chapter 2 in [95]). The existence of such one-way functions is an open conjecture [96]. Candidates for one-way functions are the discrete-logarithm and the integer factorization problem which form, resp., the basis of the Diffie–Hellman key exchange [93] and the RSA public-key cryptosystem [94]. Efficient randomized algorithms are known for the discrete-logarithm and the integer factorization problem on quantum computers [97]. Hence, public-key cryptosystems are not only provably insecure in theory, but also potentially in practice.

The OTP implements unconditional secrecy with low complexity. However, its applicability is limited in practice since Alice and Bob must share a secret key in advance. Furthermore, the key must at least be as long as the message and can be used only once. Theorem 4, however, shows that the OTP is optimal with respect to key length. Hence, any unconditionally secure cryptosystem is necessarily as impractical as the OTP.

On the other hand, the assumption that the Eve has precisely the *same* information as Bob (except for the secret key) is unrealistic in general. This is, for instance, the case in computational security schemes, which assume that Eve's channel is noiseless, but her computational resources are bounded. Physical communication channels are noisy, and in real systems, Eve has some minimal uncertainty about the signal received by Bob. The following example shows that if Eve's observation is in some sense "noisier" than Bob's, then information-theoretically secure communication is possible even when Alice and Bob do *not* share a secret key in advance.

Example 6 (The binary erasure wiretap channel [13,64])**.** *Consider the following simplistic scenario: Alice wishes to send one bit of information to Bob over a binary public channel. Eve's channel is not as perfect as Bob's: she observes a corrupted version of the bit at the output of a* $\mathsf{BEC}(\epsilon)$*. Hence, Eve knows the bit with probability* $1-\epsilon$*, and her equivocation equals* ϵ*.*

Let us assume that Alice has access to a source of private randomness which is independent of the message and the channel. To augment Eve's equivocation, Alice chooses a message M *uniformly at random from the set* $\{0,1\}$ *and employs the following coding scheme: She takes the set of all* n*-bit sequences* $\{0,1\}^n$ *and splits them into two* bins*,* b_0 *and* b_1*, which comprise all* n*-bit sequences with odd, resp., even parity. To send a message* $m \in \{0,1\}$*, Alice transmits a codeword* S^n *chosen uniformly at random in* b_m*. The rate of the code is* $\frac{1}{n}$ *bits per transmitted channel symbol.*

Clearly, Bob can recover the correct message by determining the parity of the received codeword. Eve, however, observes a sequence $Z^n \in \{0,1,\mathsf{e}\}^n$ *that has* $n\epsilon$ *erasures on average. Define a binary random variable* E *such that* $E=0$ *if* Z^n *contains no erasures and* $E=1$ *otherwise. If* $E=0$*, Eve can decode the message correctly. However, if* $E=1$*, the parity of the erased bits is equally likely to be odd or even. We can lower bound Eve's equivocation as follows:*

$$\begin{aligned} H(M|Z^n) \geq H(M|Z^n,E) &\overset{(a)}{=} H(M|Z^n,E=1)(1-(1-\epsilon)^n) \\ &= H(M)(1-(1-\epsilon)^n) = 1-(1-\epsilon)^n, \end{aligned}$$

where equality (a) follows from the fact that $Pr[E = 1] = 1 - (1 - \epsilon)^n$ *and* $H(M|Z^n, E = 0) = 0$. *Hence,* $I(M; Z^n) \leq (1 - \epsilon)^n$, *which vanishes exponentially fast in* n. *By repeating this process, Alice and Bob can agree on a secret key of arbitrary length.*

The coding scheme in Example 6 is secure in an asymptotic sense since it requires that the *total* amount of information leaked to Eve vanishes as n goes to infinity, i.e., $\lim_{n\to\infty} I(M; Z^n) = 0$. This is less stringent than requiring an exact statistical independence of M and Z^n and is often mathematically more tractable [13]. We call this the *strong secrecy* condition. Alternatively, one can require that the *rate* at which information is leaked to Eve vanishes as n goes to infinity, i.e., $\lim_{n\to\infty} \frac{1}{n} I(M; Z^n) = 0$. We call this the *weak secrecy* condition. This requirement is weaker than the strong secrecy condition since it is satisfied as long as $I(M; Z^n)$ grows at most sublinearly in n.

The wiretap channel model. Example 6 shows that one can use a noisy channel as a "cryptographic resource". We now consider a more general case first considered by Wyner [64] and subsequently generalized by Csiszár and Körner [65], where the main channel from Alice to Bob is no longer noiseless. Given a broadcast channel $\xi \in \mathsf{M}(\mathcal{S}; \mathcal{Y} \times \mathcal{Z})$, let $\kappa_s(y) := \sum_{z\in\mathcal{Z}} \xi_s(y, z)$ and $\mu_s(z) := \sum_{y\in\mathcal{Y}} \xi_s(y, z)$ be the two components of ξ. Alice chooses the input to ξ, and we refer to κ as the "main channel" and μ as "Eve's channel".

Alice uses a stochastic encoder to map the message M into an input S^n to the channels κ and μ. Bob and Eve observe, resp., the corresponding outputs Y^n and Z^n. Bob wishes to decode the message with a small probability of error such that Eve's information about the message is arbitrarily small. The largest achievable rate at which Alice can send a message to Bob is called the *secrecy capacity* $C_S(S; Y|Z)$. We give a formal definition.

Definition 18 ([65]). *The secrecy capacity of the wiretap channel is the largest rate* R *such that for every* $\epsilon > 0$, $\delta > 0$, *and sufficiently large* n, *there exist random variables* M, S^n, Y^n, *and* Z^n *satisfying* $M - S^n - Y^n Z^n$, *where* Y^n *and* Z^n *are connected with* S^n *via the channels* κ *and* μ, *resp., and* M *is distributed on a set* $\mathcal{M}$ *with* $\frac{1}{n} \log |\mathcal{M}| > R - \delta$ *and with a suitable (deterministic) decoder* $d : \mathcal{Y}^n \to \mathcal{M}$,

$$Pr[d(Y^n) \neq M] < \epsilon \qquad \textit{(reliability)}, \tag{35a}$$

$$H(M|Z^n) > \log |\mathcal{M}| - \epsilon \qquad \textit{(strong secrecy)}. \tag{35b}$$

Equation (35a) ensures that the Bob's probability of error is arbitrarily small while (35b) ensures that Eve has negligible information about the message.

The secrecy capacity of the wiretap channel admits the following characterization.

Theorem 5 ([65], Corollary 2). *The secrecy capacity* $C_S(S; Y|Z)$ *of the wiretap channel is*

$$C_S(S; Y|Z) = \max[I(U; Y) - I(U; Z)] \tag{36}$$

for random variables (U, S, Y, Z) *such that* $U - S - YZ$ *is a Markov chain and* $P_{Y|S} = \kappa$, $P_{Z|S} = \mu$. *The auxiliary variable* U *may be assumed to have a range of size at most* $|\mathcal{S}|$.

C_S depends on ξ only through its marginals κ and μ [65]. When the distribution of the input to ξ is fixed, C_S depends only on the marginal distributions of the pairs (S, Y) and (S, Z). Hence, we can analyze if secure communication is possible or not by restricting our attention to Δ_P (see Definition 1). Proposition 9 shows that one can interpret the quantity C_S as quantifying a deviation from the less noisy order.

Proposition 9 ([65], Corollary 3)**.**

$$\mu \sqsupseteq^{ln} \kappa \iff C_S(S;Y|Z) = 0. \tag{37}$$

The setting originally considered by Wyner [64] is a special case of the wiretap channel model where Eve's channel is *physically degraded* from the main channel in the sense that $\xi = \kappa \times \lambda$ for some $\lambda \in \mathsf{M}(\mathcal{Y};\mathcal{Z})$. We call this the *degraded wiretap channel* model. The binary erasure wiretap channel in Example 6 is an instance of this model where κ is a noiseless channel and $\mu = \mathsf{BEC}(\epsilon)$. The coding scheme in this example is apparently not that useful since the transmission rate goes to zero as n goes to infinity, albeit more slowly than does $I(M;Z^n)$. Nevertheless, the example suggests that when Alice is allowed to use a *stochastic* encoder, she can map a given message to a bin of codewords, and then select one of them at random to "confuse" Eve and achieve some secrecy guarantee. This intuition is brought to bear by Wyner, who showed that it is possible to transmit at a rate bounded away from zero and still achieve some secrecy guarantee by using a random binning scheme.

The secrecy capacity of the degraded wiretap channel is

$$C_S^w(S;Y|Z) = \max_{P_S}[I(S;Y) - I(S;Z)] = \max_{P_S} I(S;Y|Z), \tag{38}$$

where the second equality follows from the fact $S - Y - Z$ is a Markov chain by assumption. Note that $C_S(S;Y|Z) \geq C_S^w(S;Y|Z)$ since $U = S$ is a valid choice in (36). Also note that if Eve obtains the same information as Bob; i.e., if $Z = Y$, then $C_S^w = C_S = 0$. This is consistent with our analysis of Shannon's model and the general idea that it is *impossible* to realize unconditional security "from scratch", i.e., if only public channels are available.

For jointly distributed random variables $(S,Y,Z) \sim P$, $I(S;Y|Z)$ is a concave function of P_S for fixed $P_{YZ|S}$ (see Lemma 3.3 in [13]). Thus, the optimization problem in (38) is a convex program. We can also relate C_S^w to the main channel capacity $C_\kappa := \max_{P_S} I(S;Y)$ and to Eve's channel capacity $C_\mu := \max_{P_S} I(S;Z)$ as follows:

$$C_S^w(S;Y|Z) = \max_{P_S}[I(S;Y) - I(S;Z)] \geq \max_{P_S} I(S;Y) - \max_{P_S} I(S;Z) = C_\kappa - C_\mu.$$

The secrecy capacity of the degraded wiretap channel is hence at least as large as the difference between the main channel capacity and Eve's channel capacity. Note that if μ is physically degraded from κ, then $\kappa \sqsupseteq_S^{\mathsf{odeg}} \mu$, but not conversely. However, since C_S and C_S^w depend on ξ only through its marginals κ and μ, there is no real difference between output-degraded channels and physically degraded channels from the point of view of secure communication.

For the models discussed so far, a necessary condition for Alice and Bob to be able to communicate in secrecy is that they have an explicit *physical advantage* over Eve. In Shannon's model, for instance, Alice and Bob need to share a secret key in advance, while in Wyner's model, the main channel must be less noisier than Eve's. An obvious weakness of these models is that in a practical application, it may not often be possible to guarantee such an advantage. A key question is whether Alice and Bob can exchange messages in secrecy when they do *not* have a physical advantage to start with. Consider the following example:

Example 7 (Binary broadcast channel with independent BSCs, Lemma 1 in [85])**.** *Let* $\xi = \kappa \times \mu$ *where* $\kappa = \mathsf{BSC}(\epsilon)$ *and* $\mu = \mathsf{BSC}(\delta)$ *and* $\epsilon \leq \frac{1}{2}, \delta \leq \frac{1}{2}$*. The secrecy capacity of* ξ *is*

$$C_S(S;Y|Z) = \begin{cases} h(\delta) - h(\epsilon), & \text{if } \delta > \epsilon \\ 0, & \text{otherwise} \end{cases} \tag{39}$$

C_S vanishes whenever Bob's channel is noisier than Eve's in the sense that $\delta \leq \epsilon$. Here, $h(\cdot)$ is the binary entropy function.

Consider now a variation in the scenario in Example 7, where Bob can also send messages to Alice over an insecure public channel. We are interested in whether secrecy guarantees are possible in the range $0 < \delta \leq \epsilon < \frac{1}{2}$ for this augmented scenario. The following example, due to Maurer [85], shows an ingenious trick to achieve this.

Example 8 (Public feedback from Bob to Alice increases secrecy capacity [85])**.** *Alice inputs a random bit S to the "real" channel ξ where $S \sim \mathsf{Bernoulli}\left(\frac{1}{2}\right)$. Let $E \sim \mathsf{Bernoulli}(\epsilon)$ and $D \sim \mathsf{Bernoulli}(\delta)$ be, resp., the independent error bits of the main channel and Eve's channel. Bob observes $Y = S \oplus E$ and Eve observes $Z = S \oplus D$. We assume that the main channel is noisier than Eve's in the sense that $\delta \leq \epsilon$.*

To send a message bit C, Bob computes $W = C \oplus Y = C \oplus S \oplus E$ and sends it over the public channel. Since Alice knows S, she computes $W \oplus S = C \oplus E$. Eve, on the other hand, only knows Z, and she computes $W \oplus Z = C \oplus E \oplus D$. In effect, this procedure simulates a "conceptual" broadcast channel from Bob to Alice and Eve, where the conceptual main channel (to Alice) is equivalent to the real main channel and Eve's conceptual channel is a composition of the real main channel and Eve's real channel. This corresponds exactly to Wyner's degraded wiretap channel scenario, where Eve's conceptual channel is physically degraded from the main channel, thus allowing for some positive secrecy rate. Maurer showed that a suitably modified notion of secrecy capacity (called the secrecy capacity with public discussion*) for this augmented scenario is equal to $h(\epsilon + \delta - 2\epsilon\delta) - h(\epsilon)$, which is strictly positive unless $\epsilon = \frac{1}{2}$, $\delta = 0$ or $\delta = 1$, (see Proposition 1 in* [85]*).*

Example 8 highlights the important fact that noiseless feedback can increase the secrecy capacity. This is true even when the feedback is known to Eve and she has a physical advantage over Bob. Crucially, the latter finding suggests that the necessity of the condition that Bob has a physical advantage over Eve to achieve a positive secrecy capacity in Example 7 stems from a restriction imposed by rate-limited one-way communication. These observations motivate the study of more general models of secret key agreement using two-way or interactive public communication.

The source model for secret key agreement using public discussion. Maurer introduced the *source model* for secret key agreement [12,85]. In this model, Alice, Bob, and Eve observe n i.i.d. copies of random variables S, Y, and Z, respectively, where (S, Y, Z) follows a joint distribution known to all parties, referred to as the *source*. Alice and Bob aim to agree on a common secret key by communicating interactively over a public channel that is observable by Eve.

The *two-way* public communication protocol proceeds in rounds, with Alice and Bob alternately exchanging messages. Alice sends messages in the odd-numbered rounds, and Bob sends messages in the even-numbered rounds. Each message is a function of the sender's observation and all previously exchanged messages. At the conclusion of the protocol, Alice (resp. Bob) computes a key K (resp. K') as a function of S^n (resp. Y^n) and C, the set of all exchanged messages.

Definition 19 ([85])**.** *The* two-way secret key rate *for the source model, denoted as $S_{\leftrightarrow}(S; Y|Z)$, is the maximum rate R such that for every $\epsilon > 0$ and sufficiently large n, there exists a two-way*

public communication protocol that outputs keys K and K' (ranging over some common set $\mathcal{K}$) satisfying

$$Pr[K = K'] \geq 1 - \epsilon \qquad \textit{(reliability)}, \tag{40a}$$

$$\tfrac{1}{n} I(K; C, Z^n) \leq \epsilon \qquad \textit{(weak secrecy)}, \tag{40b}$$

$$\tfrac{1}{n} H(K) > \tfrac{1}{n} \log |\mathcal{K}| - \epsilon \qquad \textit{(uniformity)}, \tag{40c}$$

and achieving $\frac{1}{n} H(K) \geq R - \epsilon$, where C is the amount of public communication consumed in the protocol.

Equations (40a) and (40c) ensure, resp., that the keys are equal to each other with high probability and that they are almost uniformly distributed. Equation (40b) ensures that the *rate* at which Eve learns information about the keys is negligibly small. A still stronger definition requires that Eve's *total* information about the key is negligibly small, i.e.,

$$I(K; C, Z^n) \leq \epsilon; \qquad \textit{(strong secrecy)}. \tag{41}$$

Both these definitions give the same secret key rate [98]. Moreover, this rate is achievable without using private randomness at either Alice's or Bob's end. This is unlike the wiretap channel model, where coding schemes for the strong secrecy and the weak secrecy condition are very different [13] and randomness in the encoding process plays a crucial role in enabling secure communication.

Note that Definition 19 of the two-way rate says nothing about the amount of public communication (i.e., the number of rounds) required to agree on a secret key, which can be arbitrarily large. However, models imposing some restriction on the possible communication are also of interest. We say that the protocol is *one-way* if Alice is allowed to send only one message and Bob none. The corresponding key rate is called the *one-way secret key rate* $S_{\rightarrow}(S; Y|Z)$. The one-way key rate is a lower bound on the two-way key rate. $S_{\rightarrow}$ admits the following characterization.

Theorem 6 ([66], Theorem 1)**.** *The* one-way secret key rate $S_{\rightarrow}(S; Y|Z)$ *for the source model is the solution of the following optimization problem:*

$$S_{\rightarrow}(S; Y|Z) = \max_{P_{UV|SYZ}: V-U-S-YZ} I(U; Y|V) - I(U; Z|V). \tag{42}$$

In this optimization problem, it suffices to restrict the range of the random variables U and V to sizes $|\mathcal{S}|^2$ and $|\mathcal{S}|$, respectively.

The bounds on the cardinalities imply that the optimization domain is a set of stochastic matrices of finite size, which makes it possible to turn this theorem into an algorithm to compute $S_{\rightarrow}$.

The following trivial bounds on the two-way rate are known [85]:

Proposition 10.

$$\max\{I(S; Y) - I(S; Z), I(Y; S) - I(Y; Z)\} \leq S_{\leftrightarrow}(S; Y|Z) \leq \min\{I(S; Y), I(S; Y|Z)\}.$$

For some sources, the lower bound in Proposition 10 can be negative (see [99] for an operational interpretation of the lower bound when such is the case). If neither $I(S; Y) > I(S; Z)$ nor $I(Y; S) > I(Y; Z)$ holds, then Alice and Bob can exploit the authenticity of

the public channel to "distill" observations for which Alice and Bob have an advantage over Eve.

Maurer [85] and Maurer and Wolf [100] considered a scenario where a satellite broadcasts random bits at a low signal power and earthlings Alice, Bob, and Eve receive these bits over independent binary channels. Secret key agreement is *always* possible in this scenario unless Eve's channel is noiseless or either Alice or Bob receives no information at all about these bits. The following example describes an advantage distillation strategy called the "repeat-code protocol" for this scenario.

Example 9 (The "satellite" source with independent BSCs [85,100])**.** *Let* $R \sim \textsf{Bernoulli}\left(\frac{1}{2}\right)$. *We pass* R *through three independent binary symmetric channels with parameters* α, β, *and* ϵ, *resp., to obtain* S, Y, *and* Z. *We assume that* $0 \leq \alpha, \beta < \frac{1}{2}$, *and* $0 < \epsilon < \min\{\alpha, \beta\}$. *Thus, Eve has an initial advantage over Alice and Bob in the sense that* $I(S;Z) > I(S;Y)$ *and* $I(Y;Z) > I(Y;S)$.

Given n *realizations of the source, Alice and Bob exploit the authenticity of the public channel to reverse Eve's advantage as follows: Alice generates a bit* $C \sim \textsf{Bernoulli}\left(\frac{1}{2}\right)$ *and sends* $S^n \oplus C^n$ *over the public channel, where* $\oplus$ *denotes a bit-wise* XOR *operation and* C^n *is a vector consisting of* n *repetitions of the bit* C. *Bob computes* $(S^n \oplus C^n) \oplus Y^n$ *and publicly "accepts" if and only if his output is equal to either* $(0, 0, \ldots, 0)$ *or* $(1, 1, \ldots, 1)$, *when Alice retains* C; *or else, Alice discards* C. *In other words, Alice and Bob make use of a code comprising two* n*-bit codewords* $(0, 0, \ldots, 0)$ *and* $(1, 1, \ldots, 1)$ *and retain a bit only if their observations are either highly correlated or highly anti-correlated. Eve computes* $(S^n \oplus C^n) \oplus Z^n$ *and her optimal guess for* C *is* 0 *if at least half of the bits in her string is* 0, *and* 1 *otherwise. As* n *goes to infinity, Bob's average error probability when guessing the bit* C *sent by Alice decreases asymptotically faster than Eve's and that the secret key rate is strictly positive in this scenario. This protocol can be used over multiple rounds to further reduce Eve's information.*

The design of practical secret key agreement protocols turns out to be a simpler problem than the construction of wiretap channel codes [13]. A wiretap code needs to *simultaneously* guarantee reliable communication of a message to Bob (35a) and secrecy against Eve (35b). On the other hand, keys are random strings that are not meant to convey any information by themselves and do not need to be known in advance. Alice and Bob can freely shuffle, combine, or discard their observations. This allows for the design of *sequential* key distillation strategies that handle the reliability constraint (40a) and secrecy constraint (40b) *independently*. Since one can always post-process weakly secret keys, strong secrecy comes "for free," i.e., a rate achievable under the weak secrecy condition (40b) is also achievable under the strong secrecy condition (41) (see Theorem 1 in [98]).

A typical key agreement protocol operates in sequential phases [13]: First, Alice, Bob, and Eve observe n realizations of a source. Second, if neither Alice nor Bob has an initial advantage over Eve, they use an *advantage distillation* strategy to reverse Eve's advantage. Third, Alice and Bob exchange messages over the public channel and apply error correction techniques to process their observations and agree on a common bit string. This phase is called *information reconciliation*. Since the error correction information is public, the common bits are only partially secret from Eve. Fourth, Alice and Bob use a suitable hash function to distill a (shorter) highly secret string about which Eve has virtually no information. This phase is called *privacy amplification by public discussion* [101]. Finally, they use the key as an OTP for secure encryption.

The channel model for secret key agreement using public discussion. A *channel model* for secret key agreement generalizes the source model [66,85]. The model involves a channel $\xi \in \mathsf{M}(\mathcal{S}; \mathcal{Y} \times \mathcal{Z})$. Alice selects the inputs to ξ, while Bob and Eve observe, resp., the corresponding Y-outputs and the Z-outputs of ξ. Alice and Bob also have access to a

public channel. The definitions of the two-way and one-way secret key rates are similar to those for the source model (see Section 17.3 in [27]). Given a channel model, Alice can emulate the associated source model by choosing i.i.d. copies of a random variable S as inputs to ξ. The corresponding channel outputs are i.i.d. copies of Y and Z. Hence, any key rate achieved by a source model with generic variables (S, Y, Z) subject to $P_{YZ|S} = \xi$ is also achieved by the associated channel model [66].

The wiretap channel model may be regarded as a channel model where no public communication is allowed. Clearly, the secret key rate for the channel model defined by ξ is at least as large as the secrecy capacity C_S of the associated wiretap channel defined by the components κ and μ of ξ (see Theorem 5). Ahlswede and Csiszár [66] showed that the one-way secret key rate for the channel model is equal to the secrecy capacity of the associated wiretap channel model. Thus, the one-way rate depends on ξ only through κ and μ. Note, however, that the same is not true for the two-way rate [27].

The secret key rate for the channel model is sometimes called the secrecy capacity with public discussion [85] (e.g., see Example 8). This denomination is slightly misleading because the former characterizes a secret key rate, not a secure communication rate [13].

In the sequel, we shall concern ourselves primarily with the source model.

5.3. Known Bounds on the Two-Way Secret Key Rate

5.3.1. Lower Bounds

The best-known lower bound on $S_{\leftrightarrow}$ uses two-way public communication [67,80]. Given random variables $U_1, U_2, \cdots, U_k$ satisfying the Markov chain conditions

$$U_i - SU_{1:i-1} - YZ, \text{ for odd } i \tag{43}$$

$$U_i - YU_{1:i-1} - SZ, \text{ for even } i \tag{44}$$

and for any integer ζ such that $1 \leq \zeta \leq k$, we have $S_{\leftrightarrow}(S; Y|Z) \geq L(S; Y|Z)$ where

$$L(S;Y|Z) = \sum_{\substack{i \geq \zeta \\ \text{odd } i}} I(U_i;Y|U_{1:i-1}) - I(U_i;Z|U_{1:i-1}) + \sum_{\substack{i \geq \zeta \\ \text{even } i}} I(U_i;S|U_{1:i-1}) - I(U_i;Z|U_{1:i-1}), \tag{45}$$

and the cardinality bounds on $U_1, U_2, \ldots, U_k$ satisfy

$$|\mathcal{U}_i| \leq \begin{cases} |\mathcal{S}| \prod_{l=1}^{i-1} |\mathcal{U}_l| & \text{for } i \text{ odd}, \\ |\mathcal{Y}| \prod_{l=1}^{i-1} |\mathcal{U}_l| & \text{for } i \text{ even}. \end{cases} \tag{46}$$

The bound (45) is difficult to evaluate but is quite intuitive: depending on whether i is odd or even, the individual terms can be understood from the form of the one-way secret key rate in Theorem 6 when either Alice or Bob sends a public message.

5.3.2. Upper Bounds

As noted in Proposition 10, a trivial upper bound on $S_{\leftrightarrow}(S; Y|Z)$ is $\min\{I(S;Y), I(S;Y|Z)\}$ [85].

The two-way rate equals the conditional mutual information when Eve helps Alice and Bob by announcing her variable, i.e., $S_{\leftrightarrow}(SZ; YZ|Z) = I(S;Y|Z)$. This ascribes an operational meaning to $I(S;Y|Z)$ as the key rate obtained when Alice and Bob have an explicit advantage over Eve.

If Eve sends Z through a channel $P_{Z'|Z}$, then the key rate cannot decrease. Thus, we have $S_{\leftrightarrow}(S;Y|Z) \leq S_{\leftrightarrow}(S;Y|Z') \leq I(S;Y|Z')$ for any $P_{Z'|Z}$ [12]. This observation motivates an improved bound by way of the *intrinsic information*, $I_{\downarrow}$:

$$S_{\leftrightarrow}(S;Y|Z) \leq I(S;Y\downarrow Z) := \min_{P_{Z'|Z:\, SY-Z-Z'}} I(S;Y|Z'). \tag{47}$$

where Z' may be assumed to have a range of size at most $|\mathcal{Z}|$ [102]. Unlike the UI, which depends only on the marginal distributions of the pairs (S,Y) and (S,Z), $I_{\downarrow}$ depends on the full joint distribution, and also does not satisfy the consistency condition (7). Proposition 11 shows that $I_{\downarrow}$ is never lower than the UI.

Proposition 11. $UI(S;Y\backslash Z) \leq I(S;Y\downarrow Z)$.

See Appendix B for a proof.

Renner and Wolf [84] noted that the intrinsic information exhibits a property called "locking"; i.e., it can drop by an arbitrarily large amount on giving away a bit of information to Eve. In contrast, the two-way rate satisfies

$$S_{\leftrightarrow}(S;Y|ZU) \geq S_{\leftrightarrow}(S;Y|Z) - H(U) \tag{48}$$

for jointly distributed random variables (S,Y,Z,U) (see Theorem 3 in [84]), and the conditional mutual information satisfies an analogous property:

$$I(S;Y|ZU) \geq I(S;Y|Z) - H(U). \tag{49}$$

Renner and Wolf [84] proposed an improved upper bound called the *reduced intrinsic information* $I_{\downarrow\downarrow}$, which does not exhibit locking:

$$\begin{aligned} I(S;Y\downarrow\downarrow Z) &:= \inf_{P_{U|SYZ}} I(S;Y\downarrow ZU) + H(U) \\ &\geq \inf_{P_{U|SYZ}} S_{\leftrightarrow}(S;Y|ZU) + H(U) = S_{\leftrightarrow}(S;Y|Z). \end{aligned} \tag{50}$$

Choosing U to be a constant, one immediately obtains $I(S;Y\downarrow\downarrow Z) \leq I(S;Y\downarrow Z)$. $I(S;Y\downarrow\downarrow Z)$ does not lock since

$$\begin{aligned} I(S;Y\downarrow\downarrow Z) = \inf_{P_{V|SYZ}} I(S;Y\downarrow ZV) + H(V) &\leq \inf_{P_{U'|SY(Z,U)}} I(S;Y\downarrow ZUU') + H(UU') \\ &\leq I(S;Y\downarrow\downarrow ZU) + H(U), \end{aligned}$$

where the inequality in the second step follows from restricting the infimum to random variables $V = UU'$.

The tightest known upper bound on the two-way rate is [67]

$$B_2(S;Y|Z) := \inf_{P_{Z'|SYZ}} I(S;Y|Z') + S_{\rightarrow}(SY;Z'|Z). \tag{51}$$

Unfortunately, B_2 cannot be computed explicitly, as no bound on the size of Z' is known.

A slightly weaker but computable upper bound is given by the *minimum intrinsic information* [67].

$$B_1(S;Y|Z) := \min_{P_{Z'|SYZ}} I(S;Y|Z') + I(SY;Z'|Z), \tag{52}$$

where $|\mathcal{Z}'| \leq |\mathcal{S}||\mathcal{Y}||\mathcal{Z}|$.

Summarizing, we have the following chain of bounds on the two-way rate.

$$\begin{aligned} C_S(S;Y|Z) \leq S_{\rightarrow}(S;Y|Z) \leq L(S;Y|Z) &\leq S_{\leftrightarrow}(S;Y|Z) \\ &\leq B_2(S;Y|Z) \leq B_1(S;Y|Z) \\ &\leq I(S;Y{\downarrow\downarrow}Z) \leq I(S;Y{\downarrow}Z) \leq I(S;Y|Z). \end{aligned} \tag{53}$$

5.4. Properties of the UI

In this section, we show that the function UI shares some fundamental properties of the secret key rate.

We first recall the trivial bounds on the UI [6]:

$$I(S;Y) - I(S;Z) \leq UI(S;Y\backslash Z) \leq \min\{I(S;Y), I(S;Y|Z)\}. \tag{54}$$

These bounds match the trivial bounds on the two-way secret key rate in Proposition 10 (note that $S_{\leftrightarrow}(S;Y|Z)$ is symmetric under permutations of S and Y, while $UI(S;Y\backslash Z)$ is not). In the adversarial setting in Example 4, if either Eve has less information about S than Bob or, by symmetry, less information about Y than Alice, then Alice and Bob can exploit this difference to extract a secret key.

Property P.1 states that the UI does not exhibit locking.

P.1 (*UI does not lock*). For jointly distributed random variables (S, Y, Z, U),

$$UI(S;Y\backslash ZU) \geq UI(S;Y\backslash Z) - H(U). \tag{55}$$

This property is useful as it ensures that the unique information that Y has about S with respect to an adversary Z cannot "unlock", i.e., drop by an arbitrarily large amount on giving away some information to Z.

Property P.1 and Proposition 11 together imply that $UI(S;Y\backslash Z) \leq I(S;Y{\downarrow\downarrow}Z)$, a fact that will be generalized later in Theorem 9.

Property P.2 states that UI can never increase under local operations of Alice and Bob. The counterpart of this property for the secret key rate is Lemma 4 in [12]. On a related note, in Section 6.3, we discuss a construction that enforces monotonicity under local operations for an arbitrary information measure.

P.2 (*Monotonicity under local operations (LOs) of Alice and Bob*). For all (S, S', Y, Z) such that YZ–S–S' is a Markov chain, $UI(S;Y\backslash Z) \geq UI(S';Y\backslash Z)$. Likewise, for all (S, Y, Y', Z) such that SZ–Y–Y' is a Markov chain, $UI(S;Y\backslash Z) \geq UI(S;Y'\backslash Z)$.

Suppose Alice publicly announces the value of a random variable. Then, Property P.3 states that UI can never increase.

P.3 (*Monotonicity under public communication (PC) by Alice*). For all (S, Y, Z) and functions f over the support of S, $UI((S, f(S)); (Y, f(S))\backslash(Z, f(S))) \leq UI(S;Y\backslash Z)$.

The basic unit of secrecy is the "secret bit" Φ. This is any distribution defined on the sets $\{0,1\} \times \{0,1\} \times \mathcal{Z}$ such that

$$\Phi(s, y, z) := \tfrac{1}{2}\delta_{s,y} \times Q_Z(z), \tag{56}$$

where Q_Z is an arbitrary distribution.

For the secret bit, UI satisfies an intuitive normalization property:

P.4 (*Normalization*). $UI_\Phi(S;Y\backslash Z) = UI_\Phi(Y;S\backslash Z) = 1$.

Given many independent copies of $(S, Y, Z)\sim P$, the goal of a secret key agreement protocol is to distill as many copies of Φ as possible. The following two properties, additiv-

ity and asymptotic continuity, are important since we are concerned with the asymptotic rate of secret key distillation.

P.5 (*Additivity on tensor products*). Let random variables (S, S', Y, Y', Z, Z') be such that (S, Y, Z) is independent of (S', Y', Z'). Then, $UI(SS'; YY' \backslash ZZ') = UI(S; Y \backslash Z) + UI(S'; Y' \backslash Z')$.

Property P.5 is shown in Lemma 19 of [6].

Asymptotic continuity is a stronger form of continuity that takes into account convergence in relation to the dimension of the underlying state space [11,18,103–105]. Specifically, a function f is said to be asymptotically continuous if

$$|f(P) - f(P')| \leq C\epsilon \log |\mathcal{S}| + \zeta(\epsilon)$$

for all joint distributions $P, P' \in \mathbb{P}_{\mathcal{S}}$, where C is a constant, $\epsilon = \frac{1}{2}\|P - P'\|_1$, and $\zeta : [0,1] \to \mathbb{R}_+$ is a continuous function that converges to zero as $\epsilon \to 0$ [11].

As an example, entropy is asymptotically continuous (see, e.g., Lemma 2.7 in [27]): for any $P, P' \in \mathbb{P}_{\mathcal{S}}$, if $\frac{1}{2}\|P - P'\|_1 \leq \epsilon$, then

$$|H_P(S) - H_{P'}(S)| \leq \epsilon \log |\mathcal{S}| + h(\epsilon),$$

where $h(\cdot)$ is the binary entropy function. Likewise, the conditional mutual information satisfies asymptotic continuity in the following sense [84,106]: for any $P, P' \in \mathbb{P}_{\mathcal{S}\times\mathcal{Y}\times\mathcal{Z}}$, if $\frac{1}{2}\|P - P'\|_1 \leq \epsilon$, then

$$|I_P(S; Y|Z) - I_{P'}(S; Y|Z)| \leq \epsilon \log \min\{|\mathcal{S}|, |\mathcal{Y}|\} + 2h(\epsilon).$$

Note that the right-hand side of the above inequality does not depend explicitly on the cardinality of Z.

The function UI is as asymptotically continuous:

P.6 (*Asymptotic continuity*). For any $P, P' \in \mathbb{P}_{\mathcal{S}\times\mathcal{Y}\times\mathcal{Z}}$ and $\epsilon \in [0,1]$, if $\|P - P'\|_1 = \epsilon$, then $UI_{P'}(S; Y \backslash Z) - UI_P(S; Y \backslash Z) \leq \zeta(\epsilon) + \frac{5}{2}\epsilon \log \min\{|\mathcal{S}|, |\mathcal{Y}|\}$ for some bounded, continuous function $\zeta : [0,1] \to \mathbb{R}_+$ such that $\zeta(0) = 0$.

The function UI satisfies a triangle inequality:

P.7 (*Triangle inequality*). For any (S, Y, Z, Z'),

$$UI(S; Y \backslash Z) \leq UI(S; Y \backslash Z') + UI(S; Z' \backslash Z).$$

An intuitive understanding of Property (P.7) can be gained by iterating the fundamental idea of information decomposition as follows: In the presence of a fourth variable Z', we aim to decompose $u := UI(S; Y \backslash Z)$ into two components—a part u_1, which is also known to Z', and the remainder $u_2 = u - u_1$, which Z' does not know. Clearly, u_1 should be upper-bounded by $UI(S; Z' \backslash Z)$, as Z' alone knows what Z' and Y share. Moreover, $u_2 \leq UI(S; Y \backslash Z')$, since what neither Z nor Z' knows is less than what Z' does not know. Together, these observations provide a heuristic argument for why the triangle inequality should hold.

Property P.7 relies on the following monotonicity property: UI can only increase under local operations by Eve.

P.8 (*Monotonicity under local operations of Eve*). For all (S, Y, Z, Z') such that SY–Z–Z' is a Markov chain, $UI(S; Y \backslash Z) \leq UI(S; Y \backslash Z')$.

Using Property P.7 and Property P.2, we conclude:

Corollary 1. *For any* (S, Y, Z, Z'), $UI(S; Y \backslash Z) \leq UI(S; Y \backslash Z') + UI(SY; Z' \backslash Z)$.

We can interpret Corollary 1 like inequality (33): Given $(S, Y, Z, Z') \sim P$, if the induced channel $P_{Z|SY}$ dominates the channel $P_{Z'|SY}$ in the Blackwell sense, then the second term $UI(SY; Z' \backslash Z)$ vanishes (see Lemma 1). One can interpret $UI(SY; Z' \backslash Z)$ as quantifying a deviation from the Blackwell order when we replace $P_{Z|SY}$ with $P_{Z'|SY}$.

5.5. UI-Based Bounds on Secret Key Rates

General properties of upper bounds on secret key rates have been studied within the framework of secrecy or protocol monotones—non-negative real-valued functionals of joint distributions that remain non-increasing throughout the execution of a protocol (see, e.g., [14,67,88,89,107]). For example, the *intrinsic information* in (47) is a protocol monotone [84]. We defer a more general discussion on protocol monotones in the context of *resource theories* to Section 6.

The following theorem gives sufficient conditions for a function to be an upper bound for the secret key rate.

Theorem 7 (Theorem 3.1 in [107], Lemma 2.10 in [88])**.** *Let M be a non-negative real-valued function of the joint distribution of the triple (S, Y, Z) that satisfies Properties P.2–P.6. Then, M is an upper bound for the one-way secret key rate.*

If, in addition, M does not increase under public communication by Bob (Property P.3, with $f(S)$ replaced by $g(Y)$ for some function g over the support of Y), then M is an upper bound for the two-way secret key rate.

Like the UI, $S_{\rightarrow}$ depends only on the marginal distributions of the pairs (S, Y) and (S, Z) [66]. Since UI satisfies Properties P.2–P.6, the following result is immediate from Theorem 7:

Theorem 8. *$UI(S; Y \backslash Z)$ is an upper bound for the one-way secret key rate $S_{\rightarrow}(S; Y|Z)$.*

Corollary 1 implies the following result.

Proposition 12. $UI(S; Y \backslash Z) \leq B_1(S; Y|Z)$.

From Theorem 8 and Proposition 12, we have the following chain of inequalities relating the bounds on the two-way rate.

Theorem 9.

$$\begin{aligned} C_S(S; Y|Z) \leq S_{\rightarrow}(S; Y|Z) \leq UI(S; Y \backslash Z) \leq B_1(S; Y|Z) \\ \leq I(S; Y \downarrow\downarrow Z) \leq I(S; Y \downarrow Z) \leq I(S; Y|Z). \end{aligned} \tag{57}$$

Remark 4. *Given $(S, Y, Z) \sim P$, let*

$$Q^* \in \underset{Q \in \Delta_{P(S,Y,Z)}}{\arg\min} \; I_Q(S; Y|Z). \tag{58}$$

By definition, $I_{Q^*}(S; Y|Z) = UI(S; Y \backslash Z)$. Recall that the distribution Q^* is a minimum synergy *distribution (see Equation (16)). An immediate consequence of Theorem 9 is as follows: choosing $P = Q^*$, all known upper bounds on the two-way rate collapse to the UI and the conditional mutual information, respectively.*

The following example [80,108] shows that there exists a distribution for which $UI(S;Y\backslash Z)$ is *not* lower than $L(S;Y|Z)$, the best-known lower bound on the two-way rate (see (45)).

Example 10 (Doubly symmetric binary erasure (DSBE) source)**.** *Consider the DSBE source with parameters* (p,ϵ) *in Example 3.*

If $\epsilon = 0$*, we have* $Z = SY$ *and* $S_{\leftrightarrow}(S;Y|Z) = 0$*, while if* $\epsilon = 1$*, we have* $Z = \mathsf{e}$ *and* $S_{\leftrightarrow}(S;Y|Z) = I(S;Y)$*.*

For this source, the two-way rate vanishes if and only if (see Theorem 14 in [12]*)*

$$\epsilon \leq \frac{1-p}{p}. \tag{59}$$

On the other hand, both the one-way secret key rate $S_{\rightarrow}(S;Y|Z)$ *and the best-known lower bound* $L(S;Y|Z)$ *vanish if and only if (see Theorem 7 in* [80]*)*

$$\epsilon \leq 4p(1-p), \tag{60}$$

while both $UI(S;Y\backslash Z)$ *and* $UI(Y;S\backslash Z)$ *vanish if and only if*

$$\epsilon \leq 2(1-p). \tag{61}$$

For $p > \frac{1}{2}$*, we have* $\frac{1-p}{p} < 2(1-p) < 4p(1-p)$*. Figure 3 illustrates these bounds for a DSBE*$(0.6, \epsilon)$ *source.*

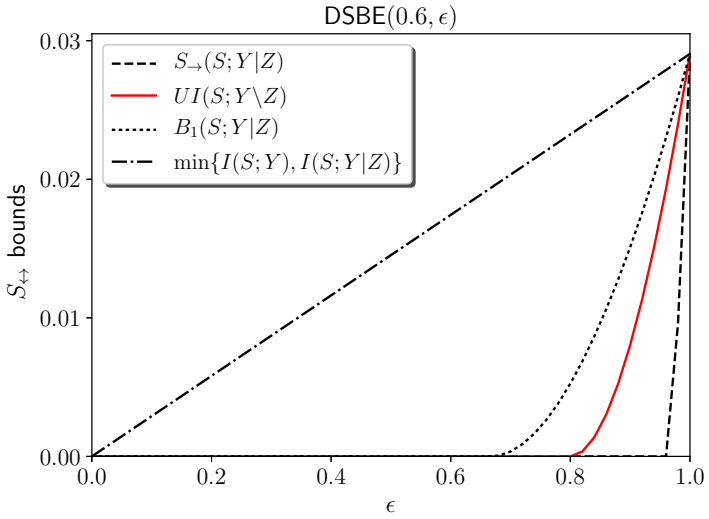

Figure 3. Bounds on the two-way secret key rate for a DSBE$(0.6, \epsilon)$ source.

On the other hand, the following example shows that UI is *not* an upper bound on $S_{\leftrightarrow}$ (see also Example 12).

Example 11 (Satellite source with independent BECs [86])**.** *Let* $R \sim$ *Bernoulli* $\left(\frac{1}{2}\right)$*. We pass* R *through three independent erasure channels with parameter* ϵ *to obtain* S*,* Y*, and* Z*. Thus,* $P_{SYZR}(s,y,z,r) = P_R(r)P_{S|R}(s|r)P_{Y|R}(y|r)P_{Z|R}(z|r)$*. Observe that* $P_{SY}(a,b) = P_{SZ}(a,b) = P_{YZ}(a,b)$ *for all* $a,b \in \{0,1,e\}$*. Therefore, all the UIs vanish. Gohari and Anantharam [86] showed that a secret key agreement protocol exists such that* $S_{\leftrightarrow}(S;Y|Z) = I(S;Y|Z) = \epsilon(1-\epsilon)^2$ *is an achievable rate.* $\epsilon(1-\epsilon)^2$ *is strictly positive for* $\epsilon \in (0,1)$*.*

We make the following conjecture:

Conjecture 1. $UI(S;Y\backslash Z) \leq S_{\leftrightarrow}(S;Y|Z)$.

Remark 5 (Sandwich bound on $S_{\leftrightarrow}(S;Y|Z)$)**.** *If Conjecture 1 is true, then*

$$UI(S;Y\backslash Z) = I_{Q^*}(S;Y|Z) \leq S_{\leftrightarrow}(S;Y|Z) \leq I_P(S;Y|Z). \tag{62}$$

Equation (62) implies that the set of all Q^* as in (58) is a set of distributions for which the UI equals the two-way rate.

A related work [87] gives necessary conditions for when the two-way rate equals the conditional mutual information.

Definition 20. *Define the following functions on $\mathbb{P}_{\mathcal{S}\times\mathcal{Y}\times\mathcal{Z}}$.*

$$B_{sUI}(S;Y|Z) := \inf_{P_{Z'|SYZ}} UI(S;Y\backslash Z') + UI(SY;Z'\backslash Z).$$
$$B_{gUI}(S;Y|Z) := \inf_{P_{Z'|SYZ}} I(S;Y|Z') + UI(SY;Z'\backslash Z).$$

As the following proposition shows, $B_{gUI}(S;Y|Z)$ is a new upper bound on the two-way rate which is juxtaposed between the two best-known bounds B_2 and B_1.

Proposition 13.

$$B_{sUI}(S;Y|Z) = UI(S;Y\backslash Z) \leq B_{gUI}(S;Y|Z) \tag{63}$$
$$B_2(S;Y|Z) \leq B_{gUI}(S;Y|Z) \leq B_1(S;Y|Z) \tag{64}$$

It remains to be seen if there exist scenarios where the B_{gUI} bound is strictly better than the B_1 bound. This remains a scope for future study.

5.6. The Blackwell Property and Secret Key Agreement Against Active Adversaries

In the source model for secret key agreement, we assume that the public channel is authenticated; i.e., Eve is only a passive adversary. In practice, this is guaranteed by authentication schemes that require Alice and Bob to share a short secret key in advance [90]. However, if this assumption is no longer valid and Eve gains both read and write access to the public channel, Maurer and Wolf [109] established an all-or-nothing result: either the same secret key rate as in the authenticated channel case can be achieved, or no key can be established at all. Maurer introduced the following property of a joint distribution to characterize the impossibility of secret key agreement in the presence of an active adversary:

Definition 21. *Given $(S,Y,Z) \sim P$, we say that Y* is simulatable by Z with respect to S *and write $\mathrm{sim}_S(Z \to Y)$ if there exists a random variable Y' such that the pairs (S,Y) and (S,Y') are statistically indistinguishable, and $S - Z - Y'$ is a Markov chain.*

It is immediately apparent that $\mathrm{sim}_S(Z \to Y)$ and $Z \sqsupseteq'_S Y$ in Definition 3 are equivalent. We now restate Maurer's impossibility result in terms of the function UI. We write $S^*_{\leftrightarrow}$ to denote the secret key rate in the active adversary scenario.

Theorem 10 ([109], Theorem 11)**.** *Let (S,Y,Z) be a triple of random variables such that $S_{\leftrightarrow}(S;Y|Z) > 0$. If either $UI(S;Y\backslash Z) = 0$ or $UI(Y;S\backslash Z) = 0$, then $S^*_{\leftrightarrow}(S;Y|Z) = 0$, else $S^*_{\leftrightarrow}(S;Y|Z) = S_{\leftrightarrow}(S;Y|Z)$.*

Remark 6. *Theorem 10 gives another operational interpretation of the vanishing UI; namely, if either S or Y possesses no unique information about each other with respect to Z, then Alice and Bob have no advantage in a secret key agreement task against an active Eve.*

Example 12 shows a distribution for which $S_{\leftrightarrow}(S;Y|Z) > 0$ but $S^*_{\leftrightarrow}(S;Y|Z) = 0$.

Example 12 ([110], Example 4)**.** *Consider the distribution* $P_{SYZ}(0,0,0) = P_{SYZ}(0,0,1) = P_{SYZ}(0,1,0) = P_{SYZ}(1,0,0) = P_{SYZ}(1,1,1) = \frac{1}{5}$. *This distribution has* $I(S;Y\downarrow Z) = 0.02$. *Gisin and Wolf [110] showed that a secret key agreement protocol exists such that* $S_{\leftrightarrow}(S;Y|Z) > 0$. *However, since the pairwise marginal distributions of* (S,Y), (S,Z), *and* (Y,Z) *are all identical, all the unique informations vanish. Thus,* $S^*_{\leftrightarrow}(S;Y|Z) = 0$.

For the passive key agreement scenario, Example 9 shows that two-round protocols can be strictly better than one-round protocols. In general, there exists no upper bound on the number of rounds required to agree on a secret key [111]. Orlitsky and Wigderson [112], however, gave a necessary and sufficient condition for the existence of a secret key: $S_{\leftrightarrow} > 0$, if and only if $S_{\leftrightarrow}$ is positive with only *two rounds* of communication. Property P.3 shows that the UI can never increase in a one-round secret key agreement protocol where Alice sends a public message to Bob. An analysis of the behavior of the UI in two-round protocols, where, in addition, Bob feeds a message back to Alice, is reserved for future study.

6. Resource Theories of Secrecy

In this concluding section, we sketch the resource-theoretic underpinnings behind Theorem 8. Resource theories study a set of objects endowed with a preorder. Classical and quantum information theories can be viewed as examples of resource theories [113]. A resource-theoretic formulation of thermodynamics is implicit in Lieb and Yngvason's axiomatic derivation of the second law of thermodynamics [114]. We refer the reader to [7–11] for detailed exposition on resource theories. We next study the problem of interconvertibility between a given pair of source and target distributions from a resource-theoretic perspective. This is similar in spirit to the work in [52] that briefly studied interconversions between the different partial information terms in (6) under local operations.

6.1. Theories of Resource Convertibility

Resource theories provide an abstract operational framework for studying what physical transformations between objects are possible under a certain class of constraints. The set of all possible operations on these objects can be divided into those that can be implemented in a cheap or simple way (called "free operations"), and those that entail a costly implementation. Given access to the set of free operations. the theory seeks to study the structure that is induced on the objects. We say that objects A and B are ordered as $A \to B$, if A can be converted to B by free operations. An object is *free* if it can be generated from scratch using only free operations; all other objects are *resources*. The resource content of an object cannot increase under free operations.

For example, in the source model for secret key agreement (Section 5.2), the objects of interest are the set of all source distributions. The set of free operations is local operations and public communication (LOPC) by Alice and Bob. The free objects are the set of all distributions under which Alice and Bob's observations are mutually independent; all other objects are *resources*. The basic resource unit is the secret bit Φ (see (56)). Resources are valuable in the sense that when combined with free operations, they can generate other resources or simulate non-free operations. For example, one can simulate a one-time pad (OTP) using LOPC and a secret bit (see Example 5).

Any non-negative, real-valued function M that respects the preorder in the sense that if $A \to B$, then $M(A) \geq M(B)$ is called a *monotone*. M can be interpreted as an assignment of a value to each object in a way that is consistent with the preorder. If $M(A) < M(B)$, then a conversion of A to B is not possible. This property is useful in practice for checking the infeasibility of a conversion.

When an *exact* conversion of A to B is not possible, we can instead ask for an *approximate* conversion at a many-copy level: convert n independent realizations of A to B' which is close to m independent realizations of the desired target B, i.e., $A^{\otimes n} \to B' \simeq B^{\otimes m}$ under the free operations. The *rate* or *yield* of this conversion is $\gamma := \frac{m}{n}$. The existence of a monotone that satisfies certain additivity and continuity properties allows us to obtain an upper bound on the rate of such conversions (Theorem 7).

The UI is a monotone that quantifies the resourcefulness or secrecy content of a source distribution when the set of free operations is local operations by Alice and Bob and one-way public communication by Alice. In particular, UI is non-increasing under this set of free operations. A consequence of this property is that the UI is an upper bound on the one-way secret key rate $S_{\to}$ (Theorem 8).

We now study two other "symmetric" monotones, the total correlation (TC) (see (65)) and the dual total correlation (DTC) (see (66)), which can be viewed as multipartite generalizations of the mutual information.

6.2. Total Correlation and Dual Total Correlation

Given $(S, Y, Z) \sim P$, the *total correlation (TC)* [115] is defined as follows:

$$\begin{aligned} TC(S;Y;Z) &= D(P_{SYZ} \| P_S \times P_Y \times P_Z) \\ &= H(S) + H(Y) + H(Z) - H(SYZ) \\ &= I(S;Y) + I(Y;Z) + I(Z;S) - CoI(S;Y;Z). \end{aligned} \tag{65}$$

TC measures the total amount of correlations between S, Y, and Z. TC is symmetric in its arguments, non-negative, and vanishes if and only if $P_{SYZ} = P_S \times P_Y \times P_Z$. Total correlation is called multi-information in [116,117] and stochastic interaction in [117].

Te Sun [34] defined a related quantity called the *dual total correlation (DTC)* based on the lattice-theoretic duality of Shannon information measures [32]:

$$\begin{aligned} DTC(S;Y;Z) &= H(SYZ) - H(S|YZ) + H(Y|SZ) - H(Z|SY) \\ &= I(S;Y|Z) + I(Y;Z|S) + I(Z;S|Y) + CoI(S;Y;Z). \end{aligned} \tag{66}$$

Like the TC, DTC is symmetric in its arguments, non-negative, and vanishes if and only if $P_{SYZ} = P_S \times P_Y \times P_Z$ [34]. From (65) and (66), we have the following relation between TC and DTC:

$$TC(S;Y;Z) + DTC(S;Y;Z) = I(S;YZ) + I(Y;SZ) + I(Z;SY). \tag{67}$$

TC and DTC capture different aspects of the correlations between S, Y, and Z. To see this, consider the RDN and XOR distributions in Example 1: The correlations in the RDN distribution can be attributed purely to pairwise interactions since S, Y, and Z are identical random variables. On the other hand, correlations in the XOR distribution arise purely due to triplewise interactions, since S, Y, and Z are pairwise independent. For the RDN, we have $TC = 2\log 2 > \log 2 = DTC$, and for the XOR, we have $DTC = 2\log 2 > \log 2 = TC$. For distributions where S, Y, and Z have binary supports, TC is maximized by the RDN distribution, while DTC is maximized by the XOR distribution [117].

Te Sun [34] studied higher-dimensional analogs of these quantities and argued that TC is effective in measuring "local" lower-order correlations, whereas DTC is effective in measuring overall higher-order correlations (see, e.g., Example 6.2 in [34]). For many practical distributions of interest, most of the TC resides in the lower-order correlations [118]. Austin [119] studied the different nature of the structures induced by small values of TC and DTC on a metric space of probability measures: if a joint distribution P has a small

TC, then P is close to a product measure, where closeness is in the sense of the Wasserstein distance; on the other hand, if P has a small DTC, then it is close to a mixture of product measures.

Interconvertibility between probability distributions under LOPC. Of immediate interest to us are the monotonicity properties of TC and DTC in relation to the problem of converting a given probability distribution to another. Concretely, we consider the following setup: Three collaborating parties, Alice, Bob, and Charlie observe i.i.d. copies of random variables S, Y, and Z, resp., distributed according to some known source distribution P. The goal is to convert P into a target distribution P' when the set of free operations is LOPC by Alice, Bob, and Charlie.

Cerf et al. [120] showed that TC and DTC are monotones under LOPC. In particular, in the tripartite case, $TC(S;Y;Z)$, $DTC(S;Y;Z)$, $I(S;YZ)$, $I(Y;SZ)$, and $I(Z;SY)$ are five monotones. From (67), it is shown that these monotones are not all linearly independent. However, none of these monotones can be expressed as a linear combination of the others with only positive coefficients. Hence, for a given source–target pair, these five monotones set independent constraints on the possible interconversions under LOPC.

Table 1 lists the values of the monotones for some distributions in Example 1. We see, for instance, that the conversion XOR $\to$ RDN is not feasible since TC increases from 1 to 2 while going from XOR to RDN. Likewise, RDN $\to$ XOR is not feasible since DTC increases from 1 to 2 while going from RDN to XOR.

Table 1. Values of the five tripartite monotones for the secret bit Φ, and the RDN and XOR distributions [120].

	$I(S;YZ)$	$I(Y;SZ)$	$I(Z;YS)$	$DTC(S;Y;Z)$	$TC(S;Y;Z)$
Φ	1	1	0	1	1
RDN	1	1	1	**1**	**2**
XOR	1	1	1	**2**	**1**

On the other hand, the following conversions are feasible and can be achieved using simple protocols [120]:

- XOR $\to \Phi$: Charlie publicly announces the value of his bit.
- RDN $\to \Phi$: Charlie forgets his bit (e.g., sends Z through a channel that completely randomizes it).
- XOR$^{\otimes 2}$ $\to$ RDN: Alice, Bob, and Charlie observe, resp., the bits (s, s'), (y, y'), and (z, z'), where $z = s \oplus y$ and $z' = s' \oplus y'$. Alice publicly announces s and Bob y'. Since Charlie knows (z, z'), she computes $z \oplus s = y$ and $z' \oplus y' = s'$ and publicly announces $w = y \oplus s'$. Finally, since Alice knows s', she computes $s' \oplus w = y$. Thus, Alice, Bob, and Charlie end up sharing the bit y.
- RDN$^{\otimes 2}$ $\to$ XOR: Alice, Bob, and Charlie observe, resp., the bits (s, s'), (y, y'), and (z, z'). Alice and Bob forget, resp., s and y', while Charlie computes $z \oplus z'$ and forgets the values z and z'.

Cerf et al. [120] considered the more general question of a reversible interconversion between an arbitrary P_{SYZ} and the distributions Φ_{SY}, Φ_{YZ}, Φ_{ZS}, RDN, and XOR, where we write Φ_{SY} for the secret bit shared between S and Y, and likewise for Φ_{YZ} and Φ_{ZS}. More concretely, does there exist yields $\gamma_1, \ldots, \gamma_5$ such that the following reversible conversion is feasible under LOPC?

$$P_{SYZ} \rightleftharpoons \Phi_{SY}^{\otimes \gamma_1} \otimes \Phi_{YZ}^{\otimes \gamma_2} \otimes \Phi_{ZS}^{\otimes \gamma_3} \otimes \text{XOR}^{\otimes \gamma_4} \otimes \text{RDN}^{\otimes \gamma_5}. \tag{68}$$

Cerf et al. [120] showed that the five monotones in Table 1 do not forbid, in principle, the following reversible conversions under LOPC:

- If $CoI = 0$, then $P_{SYZ} \rightleftharpoons \Phi_{SY}^{\otimes \gamma_1} \otimes \Phi_{YZ}^{\otimes \gamma_2} \otimes \Phi_{ZS}^{\otimes \gamma_3}$.
- If $CoI > 0$, then $P_{SYZ} \rightleftharpoons \Phi_{SY}^{\otimes \gamma_1} \otimes \Phi_{YZ}^{\otimes \gamma_2} \otimes \Phi_{ZS}^{\otimes \gamma_3} \otimes \mathrm{RDN}^{\otimes \gamma_5}$.
- If $CoI < 0$, then $P_{SYZ} \rightleftharpoons \Phi_{SY}^{\otimes \gamma_1} \otimes \Phi_{YZ}^{\otimes \gamma_2} \otimes \Phi_{ZS}^{\otimes \gamma_3} \otimes \mathrm{XOR}^{\otimes \gamma_4}$,

where $CoI = SI - CI$ is the coinformation (see (8)). It is, however, plausible that additional monotones exist that might render some of these conversions infeasible (see, e.g., [84,121]). One natural candidate for such a monotone is an "extractable" version of the function SI in Definition 1, which we describe next.

6.3. Extractable Shared Information and Monotonicity Under Local Operations

Rauh et al. [52] and Bertschinger et al. [60] argue that shared information should never increase under local operations (e.g., coarse graining) of the target and/or the predictors. Specifically, for local operations of the predictors, the function SI in Definition 1 satisfies the following property called *right monotonicity* (see A.7 in Appendix A):

$$SI(S;Y,Z) \geq SI(S;f_1(Y),f_2(Z)) \tag{69}$$

for all functions f_1 and f_2. However, for local operations on the target, SI does not exhibit a corresponding property, referred to as *left monotonicity* (see A.8 in Appendix A). Rauh et al. [19] proposed a construction that enforces left monotonicity. Define

$$\overline{SI}(S;Y,Z) = \sup_{f:\mathcal{S}\to\mathcal{S}'} SI(f(S);Y,Z), \tag{70}$$

where the supremum runs over all functions $f\colon \mathcal{S} \to \mathcal{S}'$ from the domain of S to an arbitrary finite set $\mathcal{S}'$. By construction, $\overline{SI}$ satisfies left monotonicity, and $\overline{SI}$ is the smallest function bounded from below by SI that satisfies left monotonicity. One can interpret $\overline{SI}$ as a measure of "extractable" shared information [19]. The intuition is that $\overline{SI}$ is the maximal possible amount of SI one can extract from (Y,Z) by transforming S locally. Furthermore, one can generalize the construction to define a probabilistic version of extractability by replacing f by a stochastic matrix. This leads to the definition

$$\tilde{\overline{SI}}(S;Y,Z) := \sup_{P_{S'|S}\colon YZ-S-S'} SI(S';Y,Z). \tag{71}$$

By definition, $\tilde{\overline{SI}}$ is monotone under local operations. A study of the monotonicity properties of $\tilde{\overline{SI}}$ with respect to public communication is reserved for future study.

Remark 7. *More generally, one can apply the "extractable" construction to arbitrary information measures. Furthermore, by iterating the construction, one can construct an information measure that is monotonic in all arguments* [19]. *An example of this construction is the intrinsic information* $I_\downarrow$ *in* (47). *The use of* min *instead of* max *in Definition* (47) *reflects that* $I_\downarrow$ *can only increase under local operations by Eve, a monotonicity property it shares with the function UI (see Property P.8 and Proposition 11). Work in a similar vein include* [103], *where a construction called "arrowing" is used for building probabilistically extractable versions of a given function (see also* [122]). *Galla and Gühne* [123] *discuss probabilistic extractability for a measure of correlation called the "connected correlation"* [124–128], *which are based on projections onto exponential families* [129].

6.4. Left Monotonic Information Decompositions

Is it possible to construct an information decomposition where all measures satisfy left monotonicity? The seemingly simple strategy of starting with an arbitrary decomposition

and replacing each partial information measure with its extractable counterpart fails, as this would increase all partial measures (unless already extractable), leading to an overall increase in their sum. For instance, if $\widetilde{SI}$ is replaced by a larger function, then $\widetilde{UI}$ must be reduced due to constraint (4).

As argued in [52], it is intuitive that $\widetilde{UI}$ be left monotonic. In particular, the function UI in Definition 1 satisfies left monotonicity (see Property P.2 in Section 5.4). Likewise, it is also desirable that $\widetilde{SI}$ be left monotonic [52,60]. The intuition for synergy is much less clear. The extractable construction cannot be directly generalized to ensure left monotonicity for both unique and shared information. However, such a decomposition may still exist, with left monotonicity affecting the measure of shared information. Suppose that $\widetilde{SI}$, $\widetilde{UI}$, and $\widetilde{CI}$ define a bivariate information decomposition satisfying (4)–(6), and suppose that $\widetilde{SI}$ and $\widetilde{UI}$ satisfy left monotonicity. Then,

$$\widetilde{SI}(f(Y,Z);Y,Z) \leq I(Y;Z) \tag{72}$$

for any function f [19]. Inequality (72) is related to the identity axiom (see A.6 in Appendix A). Indeed, it is easy to derive (72) from the identity axiom and from the assumption that $\widetilde{SI}$ is left monotonic. None of the non-negative information decompositions proposed so far satisfies (72). Griffith et al. [130] proposed a function $I_\curlywedge$ as a measure of shared information that satisfies left monotonicity. However, this function does not induce a non-negative information decomposition (see A.4 in Appendix A).

The next proposition shows that left monotonicity of the shared information is not consistent with the Blackwell property of the unique information:

Proposition 14 ([19,20])**.** *There is no bivariate information decomposition satisfying* (4)–(6) *in which* $\widetilde{UI}$ *satisfies the Blackwell property and* $\widetilde{SI}$ *satisfies left monotonicity.*

A resource-theoretic characterization of the complementary information appears challenging. The problem resides with the fact that it is difficult to postulate how the complementary information should behave if, say, one or more parties perform some local operations on their subsystems. Two studies in this direction deserve notice: Rauh et al.'s Section IV.C in [52] show that the measure CI in Definition 1 can either increase or decrease under local operations of the targets and/or the sources. Another work is a decomposition of the total correlation (TC) due to Amari [124]. The total correlation among three variables can be decomposed into a sum of two non-negative terms that quantify, resp., the amount of correlations arising from purely pairwise and purely triplewise interactions [124] (Equation 78). The latter term can be interpreted as the synergistic component of the total correlation. However, examples are known where this component violates monotonicity under local operations [123]. Finally, an axiomatic approach to information flow in computational systems is motivated in [131], where it is shown that the CI in Definition 1 provides an intuitive and insightful measure of information flow volume. This warrants further investigation.

Funding: A part of this research was funded by the European Research Council (ERC) under the EU's Horizon 2020 research and innovation program (grant agreement no 757983).

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the author.

Acknowledgments: The author gratefully acknowledges Johannes Rauh, Eckehard Olbrich, Jürgen Jost, Guido Montúfar, Nils Bertschinger, Tobias Fritz, and David Wolpert for their valuable insights and helpful discussions throughout the development of this work.

Conflicts of Interest: The author declares no conflicts of interest.

Appendix A. The Axiomatic Approach to Shared Information

An axiomatic approach to the concept of shared information was pioneered by Williams and Beer [46]. Recall that the shared information can be interpreted as mutual information without the unique information; see (4).

For the general case of k finite predictor variables $Y_1, \ldots, Y_k$, Williams and Beer [46] proposed the *partial information lattice* framework to decompose the total mutual information $I(S; Y_1, \ldots, Y_k)$ into a sum of terms (called *partial information terms*) corresponding to the different ways in which combinations of the variables $Y_1, \ldots, Y_k$ convey shared, unique, or complementary information about the target S. When $k = 2$, writing $Y_1 \equiv Y$ and $Y_2 \equiv Z$, the decomposition has the form given by (4)–(6).

The partial information lattice is a consequence of certain natural properties of the shared information (sometimes called the *Williams–Beer axioms*). The underlying idea is that any information about S can be classified according to "who knows what", i.e., which information about S is shared by which subsets of $\{Y_1, \ldots, Y_k\}$. This idea resonates with *secret-sharing schemes*, a fundamental tool used in many cryptographic protocols [132]. A secret-sharing scheme involves a secret (S), a finite set $\mathcal{K} = \{1, \ldots, k\}$ of parties, and a family $\mathcal{A}$ of (nonempty) subsets of $\mathcal{K}$ called the *access structure* that is closed to taking supersets. The goal is to distribute the secret (S) among k parties such that only elements of $\mathcal{A}$ can reconstruct the secret, while all other subsets of $\mathcal{K}$ obtain no information about the secret. There is a one-to-one correspondence between the partial information terms of Williams and Beer's decomposition scheme and the set of access structures of secret-sharing schemes with k parties [59].

Let $\widetilde{SI}(S; Y_1, \ldots, Y_k)$ denote the information about S that is shared among the random variables $Y_1, \ldots, Y_k$. It is natural to demand that $\widetilde{SI}$ satisfies the following properties [46]:

A.1 (*Symmetry*). $\widetilde{SI}(S; Y_1, \ldots, Y_k)$ is symmetric under permutations of $Y_1, \ldots, Y_k$.

A.2 (*Self-redundancy*). $\widetilde{SI}(S; Y_1) = I(S; Y_1)$.

A.3 (*Monotonicity*). $\widetilde{SI}(S; Y_1, \ldots, Y_{k-1}, Y_k) \leq \widetilde{SI}(S; Y_1, \ldots, Y_{k-1})$, with equality if $Y_i = f(Y_k)$ for some $i < k$ and some function f.

We refer to properties A.1–A.3 as the *Williams–Beer axioms*. Any function satisfying these axioms is non-negative [46]. The axioms, however, do not uniquely characterize the function $\widetilde{SI}$. When $\widetilde{SI}$ is defined, we can associate with each element of the partial information lattice a "local" quantity (the partial information term) that is uniquely determined from $\widetilde{SI}$ by a Möbius inversion. The total mutual information $I(S; Y_1, \ldots, Y_k)$ is then a sum of these local terms [46]. In general, however, the local terms can be negative. For a non-negative decomposition, we require the following additional property:

A.4 (*Local positivity*). All the partial information terms in the induced decomposition are non-negative.

Williams and Beer defined a function

$$\begin{aligned} I_{\min}(S; Y_1, \ldots, Y_k) &= \sum_s P_S(s) \min_i \Big\{ \sum_{y_i} P_{Y_i|S}(y_i|s) \log \frac{P_{S|Y_i}(s|y_i)}{P_S(s)} \Big\} \\ &= \sum_s \min_i \Big\{ \sum_{y_i} P_{SY_i}(s, y_i) \log \frac{P_{SY_i}(s, y_i)}{P_S(s) P_{Y_i}(y_i)} \Big\}, \end{aligned} \tag{A1}$$

and showed that $I_{\min}$ satisfies A.1–A.3 and that the decomposition induced by I_{min} satisfies A.4. While the measure $I_{\min}$ has subsequently been criticized for "not measuring the right thing" [26,53,60], there has been no successful attempt to find better measures, except for

the bivariate case ($k = 2$) [6,26]. One problem seems to be the lack of a clear consensus on the values of the shared information for some paradigmatic examples. For example, it seems natural that the shared information is zero for the COPY distribution in Example 1 since Y and Z are independent [6,26,60]. However, $I_{\min}$ assigns 1 bit of shared information in this case. The second problem relates to the difficulty of coming up with a minimal number of "natural" and "essential" properties for the shared information.

Bertschinger et al. [60] proposed the following additional axiom:

A.5 (*Left chain rule*). $\widetilde{SI}(SS'; Y_1, \ldots, Y_k) = \widetilde{SI}(S; Y_1, \ldots, Y_k) + \widetilde{SI}(S'; Y_1, \ldots, Y_k|S)$, where $\widetilde{SI}(S'; Y_1, \ldots, Y_k|S) = \sum_{s \in \mathcal{S}} P_S(s) \widetilde{SI}(S'; Y_1, \ldots, Y_k|s)$.

A.5 is a natural generalization of the chain rule of mutual information (3).

Specializing to the bivariate case, A.4 and A.5 together imply the following property [60], which was first proposed in [26]:

A.6 (*Identity*). $\widetilde{SI}((Y, Z); Y, Z) = I(Y; Z)$.

The identity property states that if S is an identical copy of the predictor variables, i.e., if $S = (Y, Z)$, then the shared information should equal the mutual information between Y and Z. Rauh et al. [52], however, showed that A.6 is incompatible with A.4 for $k \geq 3$. This implies that A.4 and A.5 are not compatible for $k \geq 3$.

Rauh et al. [52] argue that shared information should never increase if the target and/or the predictors perform some *local* operation (e.g., coarse graining) on their subsystems. For local operations of the predictors, the Williams–Beer axioms imply the following property:

A.7 *(Right monotonicity).* $\widetilde{SI}(S; Y_1, \ldots, Y_k) \geq \widetilde{SI}(S; f_1(Y_1), \ldots, f_k(Y_k))$ *for all functions* $f_1, \ldots, f_k$.

On the other hand, the left chain rule A.5 implies the following property [60]:

A.8 (*Left monotonicity*). $\widetilde{SI}(S; Y_1, \ldots, Y_k) \geq \widetilde{SI}(f(S); Y_1, \ldots, Y_k)$ for all functions f.

Appendix B. Deferred Proofs

Proof of Lemma 4. Since $Z \sqsupseteq_S Y$, there exists some $\lambda' \in \mathsf{M}(\mathcal{Z}; \mathcal{Y})$ such that $\kappa = \lambda' \circ \mu$. Hence,

$$\begin{aligned}
\delta_o^\pi(\kappa, \nu) &= \min_{\lambda \in \mathsf{M}(\mathcal{Y}; \mathcal{W})} D(\nu \| \lambda \circ \kappa | \pi_S) \\
&= \min_{\lambda \in \mathsf{M}(\mathcal{Y}; \mathcal{W})} D(\nu \| \lambda \circ \lambda' \circ \mu | \pi_S) \\
&\geq \min_{\lambda \in \mathsf{M}(\mathcal{Z}; \mathcal{W})} D(\nu \| \lambda \circ \mu | \pi_S) = \delta_o^\pi(\mu, \nu).
\end{aligned}$$

Since $Y \sqsupseteq_S W$, there exists some $\lambda' \in \mathsf{M}(\mathcal{Y}; \mathcal{W})$ such that $\nu = \lambda' \circ \kappa$. Hence,

$$\begin{aligned}
\delta_o^\pi(\mu, \nu) &= \min_{\lambda \in \mathsf{M}(\mathcal{Z}; \mathcal{W})} D(\nu \| \lambda \circ \mu | \pi_S) \\
&= \min_{\lambda \in \mathsf{M}(\mathcal{Z}; \mathcal{W})} D(\lambda' \circ \kappa \| \lambda \circ \mu | \pi_S) \\
&\leq \min_{\lambda \in \mathsf{M}(\mathcal{Z}; \mathcal{Y})} D(\lambda' \circ \kappa \| \lambda' \circ \lambda \circ \mu | \pi_S) \\
&\leq \min_{\lambda \in \mathsf{M}(\mathcal{Z}; \mathcal{Y})} D(\kappa \| \lambda \circ \mu | \pi_S) = \delta_o^\pi(\mu, \kappa),
\end{aligned}$$

where the inequality in the last step follows from the data processing inequality for the KL divergence (Theorem 2.2 in [28]). □

Proof of Proposition 11. We shall use the following variational characterization of the UI, which follows from Property P.8:

$$UI(S;Y\backslash Z) = \min_{P_{Z'|Z}:SY-Z-Z'} UI(S;Y\backslash Z'). \tag{A2}$$

Let $(S,Y,Z) \sim P$ and let $(S,Y,Z') \sim P'$ such that $P' = P \cdot P_{Z'|Z} \equiv \sum_{z\in\mathcal{Z}} P_{SYZ}P_{Z'|Z}$. Let $Q \in \Delta_P$ and $Q' = Q \cdot P_{Z'|Z} \in \Delta_{P'}$. We then have

$$\begin{aligned}
UI(S;Y\backslash Z) &= \min_{P_{Z'|Z}:SY-Z-Z'} UI(S;Y\backslash Z') \\
&= \min_{P_{Z'|Z}:SY-Z-Z'} \min_{Q'\in\Delta_{P'}} I_{Q'}(S;Y|Z') \\
&\leq \min_{P_{Z'|Z}:SY-Z-Z'} \min_{Q\in\Delta_P} I_{Q\cdot P_{Z'|Z}}(S;Y|Z') \\
&= \min_{Q\in\Delta_P} \min_{P_{Z'|Z}:SY-Z-Z'} I_{Q\cdot P_{Z'|Z}}(S;Y|Z') \\
&= \min_{Q\in\Delta_P} I_Q(S;Y{\downarrow}Z) \leq I_P(S;Y{\downarrow}Z),
\end{aligned}$$

where the first step is just (A2), and the inequality in the third step follows since for any $Q \in \Delta_P$, Q' lies in $\Delta_{P'}$. □

References

1. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [CrossRef]
2. Rényi, A. On the foundations of information theory. *Rev. L'Institut Int. Stat.* **1965**, *33*, 1–14. [CrossRef]
3. Rényi, A. On measures of entropy and information. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*; University of California Press: Berkeley, CA, USA, 1961; pp. 547–561.
4. Csiszár, I. A class of measures of informativity of observation channels. *Period. Math. Hung.* **1972**, *2*, 191–213. [CrossRef]
5. Csiszár, I. Axiomatic characterizations of information measures. *Entropy* **2008**, *10*, 261–273. [CrossRef]
6. Bertschinger, N.; Rauh, J.; Olbrich, E.; Jost, J.; Ay, N. Quantifying Unique Information. *Entropy* **2014**, *16*, 2161–2183. [CrossRef]
7. Coecke, B.; Fritz, T.; Spekkens, R.W. A mathematical theory of resources. *Inf. Comput.* **2016**, *250*, 59–86. [CrossRef]
8. Fritz, T. Resource convertibility and ordered commutative monoids. *Math. Struct. Comput. Sci.* **2017**, *27*, 850–938. [CrossRef]
9. Del Rio, L.; Kraemer, L.; Renner, R. Resource theories of knowledge. *arXiv* **2015**, arXiv:1511.08818.
10. Goold, J.; Huber, M.; Riera, A.; del Rio, L.; Skrzypczyk, P. The role of quantum information in thermodynamics—A topical review. *J. Phys. Math. Theor.* **2016**, *49*, 143001. [CrossRef]
11. Chitambar, E.; Gour, G. Quantum resource theories. *Rev. Mod. Phys.* **2019**, *91*, 025001. [CrossRef]
12. Maurer, U.M.; Wolf, S. Unconditionally Secure Key Agreement and the Intrinsic Conditional Information. *IEEE Trans. Inf. Theory* **1999**, *45*, 499–514. [CrossRef]
13. Bloch, M.; Barros, J. *Physical-Layer Security: From Information Theory to Security Engineering*; Cambridge University Press: Cambridge, UK, 2011.
14. Narayan, P.; Tyagi, H. Multiterminal secrecy by public discussion. *Found. Trends Commun. Inf. Theory* **2016**, *13*, 129–275. [CrossRef]
15. Tyagi, H.; Watanabe, S. *Information-Theoretic Cryptography*; Cambridge University Press: Cambridge, UK, 2023.
16. Banerjee, P.K.; Olbrich, E.; Jost, J.; Rauh, J. Unique informations and deficiencies. In Proceedings of the 56th Annual Allerton Conference on Communication, Control and Computing, Monticello, IL, USA, 2–5 October 2018; pp. 32–38.
17. Rauh, J.; Banerjee, P.K.; Olbrich, E.; Jost, J. Unique information and secret key decompositions. In Proceedings of the IEEE International Symposium on Information Theory (ISIT), Paris, France, 7–12 July 2019; pp. 3042–3046.
18. Rauh, J.; Banerjee, P.K.; Olbrich, E.; Montúfar, G.; Jost, J. Continuity and additivity properties of information decompositions. *Int. J. Approx. Reason.* **2023**, *161*, 108979. [CrossRef]
19. Rauh, J.; Banerjee, P.K.; Olbrich, E.; Jost, J.; Bertschinger, N. On extractable shared information. *Entropy* **2017**, *19*, 328. [CrossRef]
20. Rauh, J.; Banerjee, P.K.; Olbrich, E.; Jost, J.; Bertschinger, N.; Wolpert, D. Coarse-graining and the Blackwell order. *Entropy* **2017**, *19*, 527. [CrossRef]
21. Banerjee, P.K.; Rauh, J.; Montúfar, G. Computing the unique information. In Proceedings of the IEEE International Symposium on Information Theory (ISIT), Vail, CO, USA, 17–22 June 2018; pp. 141–145.
22. Banerjee, P.K. Unique Information and the Blackwell Order. Ph.D. Thesis, Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany, 2020.
23. Le Cam, L. Sufficiency and approximate sufficiency. *Ann. Math. Stat.* **1964**, *35*, 1419–1455. [CrossRef]
24. Torgersen, E. *Comparison of Statistical Experiments*; Cambridge University Press: Cambridge, UK, 1991; Volume 36.

25. Raginsky, M. Shannon meets Blackwell and Le Cam: Channels, codes, and statistical experiments. In Proceedings of the IEEE International Symposium on Information Theory (ISIT), St. Petersburg, Russia, 31 July–5 August 2011; pp. 1220–1224.
26. Harder, M.; Salge, C.; Polani, D. A Bivariate measure of redundant information. *Phys. Rev. E* **2013**, *87*, 012130. [CrossRef]
27. Csiszár, I.; Körner, J. *Information Theory: Coding Theorems for Discrete Memoryless Systems*; Cambridge University Press: Cambridge, UK, 2011.
28. Polyanskiy, Y.; Wu, Y. Lecture Notes on Information Theory. Lecture Notes for ECE563 (UIUC) and 6.441 (MIT), 2012–2017. Available online: https://ocw.mit.edu/courses/6-441-information-theory-spring-2016/pages/lecture-notes/ (accessed on 12 November 2024).
29. Wehrl, A. General properties of entropy. *Rev. Mod. Phys.* **1978**, *50*, 221. [CrossRef]
30. Bell, A.J. The Co-Information Lattice. In Proceedings of the Fourth International Symposium on Independent Component Analysis and Blind Signal Separation (ICA 03), Charleston, SC, USA, 5–8 March 2003.
31. McGill, W. Multivariate information transmission. *IRE Trans. Inf. Theory* **1954**, *4*, 93–111.
32. Han, T.S. Linear dependence structure of the entropy space. *Inf. Control.* **1975**, *29*, 337–368. [CrossRef]
33. Yeung, R.W. A new outlook on Shannon's information measures. *IEEE Trans. Inf. Theory* **1991**, *37*, 466–474. [CrossRef]
34. Han, T.S. Nonnegative entropy measures of multivariate symmetric correlations. *Inf. Control.* **1978**, *36*, 133–156. [CrossRef]
35. Hayden, P.; Headrick, M.; Maloney, A. Holographic mutual information is monogamous. *Phys. Rev. D* **2013**, *87*, 046003. [CrossRef]
36. Brenner, N.; Strong, S.P.; Koberle, R.; Bialek, W.; Steveninck, R.R.d.R.v. Synergy in a neural code. *Neural Comput.* **2000**, *12*, 1531–1552. [CrossRef] [PubMed]
37. Averbeck, B.B.; Latham, P.E.; Pouget, A. Neural correlations, population coding and computation. *Nat. Rev. Neurosci.* **2006**, *7*, 358. [CrossRef] [PubMed]
38. Gat, I.; Tishby, N. Synergy and redundancy among brain cells of behaving monkeys. *Adv. Neural Inf. Process. Syst.* **1999**, *11*, 111–117.
39. Reich, D.S.; Mechler, F.; Victor, J.D. Independent and redundant information in nearby cortical neurons. *Science* **2001**, *294*, 2566–2568. [CrossRef]
40. Schneidman, E.; Puchalla, J.L.; Segev, R.; Harris, R.A.; Bialek, W.; Berry, M.J. Synergy from silence in a combinatorial neural code. *J. Neurosci.* **2011**, *31*, 15732–15741. [CrossRef] [PubMed]
41. Chechik, G.; Anderson, M.J.; Bar-Yosef, O.; Young, E.D.; Tishby, N.; Nelken, I. Reduction of information redundancy in the ascending auditory pathway. *Neuron* **2006**, *51*, 359–368. [CrossRef] [PubMed]
42. Anastassiou, D. Computational analysis of the synergy among multiple interacting genes. *Mol. Syst. Biol.* **2007**, *3*, 83. [CrossRef]
43. Kontoyiannis, I.; Lucena, B. Mutual information, synergy and some curious phenomena for simple channels. In Proceedings of the IEEE International Symposium on Information Theory (ISIT), Adelaide, SA, Australia, 4–9 September 2005; pp. 1651–1655.
44. Steudel, B.; Ay, N. Information-theoretic inference of common ancestors. *Entropy* **2015**, *17*, 2304–2327. [CrossRef]
45. Jakulin, A.; Bratko, I. Quantifying and Visualizing Attribute Interactions: An Approach Based on Entropy. *arXiv* **2003**, arXiv:cs/0308002.
46. Williams, P.; Beer, R. Nonnegative Decomposition of Multivariate Information. *arXiv* **2010**, arXiv:1004.2515v1.
47. Latham, P.E.; Nirenberg, S. Synergy, Redundancy, and Independence in Population Codes, Revisited. *J. Neurosci.* **2005**, *25*, 5195–5206. [CrossRef] [PubMed]
48. Pola, G.; Thiele, A.; Hoffmann, K.P.; Panzeri, S. An exact method to quantify the information transmitted by different mechanisms of correlational coding. *Network Comput. Neural Syst.* **2003**, *14*, 35–60. [CrossRef] [PubMed]
49. Oizumi, M.; Ishii, T.; Ishibashi, K.; Hosoya, T.; Okada, M. Mismatched decoding in the brain. *J. Neurosci.* **2010**, *30*, 4815–4826. [CrossRef] [PubMed]
50. Steeg, G.V.; Brekelmans, R.; Harutyunyan, H.; Galstyan, A. Disentangled representations via synergy minimization. In Proceedings of the 55th Annual Allerton Conference on Communication, Control and Computing, Monticello, IL, USA, 3–6 October 2017; pp. 180–187.
51. Schneidman, E.; Bialek, W.; Berry, M.J. Synergy, redundancy, and independence in population codes. *J. Neurosci.* **2003**, *23*, 11539–11553. [CrossRef]
52. Rauh, J.; Bertschinger, N.; Olbrich, E.; Jost, J. Reconsidering unique information: Towards a multivariate information decomposition. In Proceedings of the IEEE International Symposium on Information Theory (ISIT), Honolulu, HI, USA, 29 June–4 July 2014; pp. 2232–2236.
53. Griffith, V.; Koch, C. Quantifying Synergistic Mutual Information. In *Guided Self-Organization: Inception*; Emergence, Complexity and Computation; Springer: Berlin/Heidelberg, Germany, 2014; Volume 9, pp. 159–190.
54. Barrett, A.B. Exploration of synergistic and redundant information sharing in static and dynamical Gaussian systems. *Phys. Rev. E* **2015**, *91*, 052802. [CrossRef]
55. Olbrich, E.; Bertschinger, N.; Rauh, J. Information decomposition and synergy. *Entropy* **2015**, *17*, 3501–3517. [CrossRef]

56. Chicharro, D.; Panzeri, S. Synergy and redundancy in dual decompositions of mutual information gain and information loss. *Entropy* **2017**, *19*, 71. [CrossRef]
57. Chicharro, D. Quantifying multivariate redundancy with maximum entropy decompositions of mutual information. *arXiv* **2017**, arXiv:1708.03845.
58. Pica, G.; Piasini, E.; Safaai, H.; Runyan, C.; Harvey, C.; Diamond, M.; Kayser, C.; Fellin, T.; Panzeri, S. Quantifying how much sensory information in a neural code is relevant for behavior. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 3689–3699.
59. Rauh, J. Secret sharing and shared information. *Entropy* **2017**, *19*, 601. [CrossRef]
60. Bertschinger, N.; Rauh, J.; Olbrich, E.; Jost, J. Shared Information—New Insights and Problems in Decomposing Information in Complex Systems. In *Proceedings ECCS 2012*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 251–269.
61. Blackwell, D. Equivalent Comparisons of Experiments. *Ann. Math. Stat.* **1953**, *24*, 265–272. [CrossRef]
62. Bertschinger, N.; Rauh, J. The Blackwell relation defines no lattice. In Proceedings of the IEEE International Symposium on Information Theory (ISIT), Honolulu, HI, USA, 29 June–4 July 2014; pp. 2479–2483.
63. Körner, J.; Marton, K. Comparison of two noisy channels. In *Topics in Information Theory*; Colloquia Mathematica Societatis Janos Bolyai: Keszthely, Hungary, 1975; Volume 16, pp. 411–423.
64. Wyner, A.D. The wire-tap channel. *Bell Syst. Tech. J.* **1975**, *54*, 1355–1387. [CrossRef]
65. Csiszár, I.; Körner, J. Broadcast channels with confidential messages. *IEEE Trans. Inf. Theory* **1978**, *24*, 339–348. [CrossRef]
66. Ahlswede, R.; Csiszár, I. Common randomness in information theory and cryptography–Part I: Secret sharing. *IEEE Trans. Inf. Theory* **1993**, *39*, 1121–1132. [CrossRef]
67. Gohari, A.A.; Anantharam, V. Information-theoretic key agreement of multiple terminals–Part I. *IEEE Trans. Inf. Theory* **2010**, *56*, 3973–3996. [CrossRef]
68. Shannon, C.E. A note on a partial ordering for communication channels. *Inf. Control.* **1958**, *1*, 390–397. [CrossRef]
69. El Gamal, A.; Kim, Y.H. *Network Information Theory*; Cambridge University Press: Cambridge, UK, 2011.
70. El Gamal, A. The capacity of a class of broadcast channels. *IEEE Trans. Inf. Theory* **1979**, *25*, 166–169. [CrossRef]
71. Geng, Y.; Nair, C.; Shitz, S.S.; Wang, Z.V. On broadcast channels with binary inputs and symmetric outputs. *IEEE Trans. Inf. Theory* **2013**, *59*, 6980–6989. [CrossRef]
72. Cohen, J.; Kemperman, J.; Zbăganu, G. *Comparisons of Stochastic Matrices with Applications in Information Theory, Statistics, Economics, and Population Sciences*; Birkhäuser: Basel, Switzerland, 1998.
73. de Oliveira, H. Blackwell's informativeness theorem using diagrams. *Games Econ. Behav.* **2018**, *109*, 126–131. [CrossRef]
74. Dahl, G. Matrix majorization. *Linear Algebra Its Appl.* **1999**, *288*, 53–73. [CrossRef]
75. Rockafellar, R.T. *Convex Analysis*; Princeton University Press: Princeton, NJ, USA, 2015.
76. Nasser, R. Characterizations of Two Channel Orderings: Input-Degradedness and the Shannon Ordering. *IEEE Trans. Inf. Theory* **2018**, *64*, 6759–6770. [CrossRef]
77. Nasser, R. On the input-degradedness and input-equivalence between channels. In Proceedings of the IEEE International Symposium on Information Theory (ISIT), Aachen, Germany, 25–30 June 2017; pp. 2453–2457.
78. Van Dijk, M. On a special class of broadcast channels with confidential messages. *IEEE Trans. Inf. Theory* **1997**, *43*, 712–714. [CrossRef]
79. Nair, C. Capacity Regions of Two New Classes of Two-Receiver Broadcast Channels. *IEEE Trans. Inf. Theory* **2010**, *56*, 4207–4214. [CrossRef]
80. Gohari, A.; Günlü, O.; Kramer, G. Coding for positive rate in the source model key agreement problem. *IEEE Trans. Inf. Theory* **2020**, *66*, 6303–6323. [CrossRef]
81. Banerjee, P.K.; Montúfar, G. The variational deficiency bottleneck. In Proceedings of the International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 19–24 July 2020; pp. 1–8.
82. Csiszár, I.; Matúš, F. Information projections revisited. *IEEE Trans. Inf. Theory* **2003**, *49*, 1474–1490. [CrossRef]
83. Banerjee, P.K.; Olbrich, E.; Jost, J.; Rauh, J. Unique Informations and Deficiencies. *arXiv* **2018**, arXiv:1807.05103.
84. Renner, R.; Wolf, S. New Bounds in Secret-Key Agreement: The Gap between Formation and Secrecy Extraction. In Proceedings of the Advances in Cryptology—EUROCRYPT 2003, Warsaw, Poland, 4–8 May 2003; pp. 562–577.
85. Maurer, U.M. Secret key agreement by public discussion from common information. *IEEE Trans. Inf. Theory* **1993**, *39*, 733–742. [CrossRef]
86. Gohari, A.A.; Anantharam, V. Comments On "Information-Theoretic Key Agreement of Multiple Terminals–Part I". *IEEE Trans. Inf. Theory* **2017**, *63*, 5440–5442. [CrossRef]
87. Chitambar, E.; Fortescue, B.; Hsieh, M.H. Distributions attaining secret key at a rate of the conditional mutual information. In *Proceedings of the Annual Cryptology Conference*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 443–462.
88. Maurer, U.M.; Renner, R.; Wolf, S. Unbreakable keys from random noise. In *Security with Noisy Data*; Springer: Berlin/Heidelberg, Germany, 2007; pp. 21–44.

89. Keykhosravi, K.; Mahzoon, M.; Gohari, A.A.; Aref, M.R. From source model to quantum key distillation: An improved upper bound. In Proceedings of the IEEE IWCIT, Tehran, Iran, 7–8 May 2014; pp. 1–6.
90. Wegman, M.N.; Carter, J.L. New hash functions and their use in authentication and set equality. *J. Comput. Syst. Sci.* **1981**, *22*, 265–279. [CrossRef]
91. Shannon, C.E. Communication theory of secrecy systems. *Bell Syst. Tech. J.* **1949**, *28*, 656–715. [CrossRef]
92. Vernam, G.S. Cipher printing telegraph systems for secret wire and radio telegraphic communications. *Trans. Am. Inst. Electr. Eng.* **1926**, *45*, 295–301. [CrossRef]
93. Diffie, W.; Hellman, M. New directions in cryptography. *IEEE Trans. Inf. Theory* **1976**, *22*, 644–654. [CrossRef]
94. Rivest, R.L.; Shamir, A.; Adleman, L. A method for obtaining digital signatures and public-key cryptosystems. *Commun. ACM* **1978**, *21*, 120–126. [CrossRef]
95. Goldreich, O. *Foundations of Cryptography: Basic Tools*; Cambridge University Press: Cambridge, UK, 2001.
96. Impagliazzo, R. A personal view of average-case complexity. In Proceedings of the Structure in Complexity Theory, Tenth Annual IEEE Conference, Minneapolis, MN, USA, 19–22 June 1995; pp. 134–147.
97. Shor, P.W. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Rev.* **1999**, *41*, 303–332. [CrossRef]
98. Maurer, U.M.; Wolf, S. From Weak to Strong Information-Theoretic Key Agreement. In Proceedings of the IEEE International Symposium on Information Theory (ISIT), Sorrento, Italy, 25–30 June 2000; p. 18.
99. Oppenheim, J.; Spekkens, R.W.; Winter, A. A classical analogue of negative information. *arXiv* **2005**, arXiv:quant-ph/0511247.
100. Maurer, U.M.; Wolf, S. Towards characterizing when information-theoretic secret key agreement is possible. In *Proceedings of the International Conference on the Theory and Application of Cryptology and Information Security*; Springer: Berlin/Heidelberg, Germany, 1996; pp. 196–209.
101. Bennett, C.H.; Brassard, G.; Crépeau, C.; Maurer, U.M. Generalized privacy amplification. *IEEE Trans. Inf. Theory* **1995**, *41*, 1915–1923. [CrossRef]
102. Christandl, M.; Renner, R.; Wolf, S. A property of the intrinsic mutual information. In Proceedings of the IEEE International Symposium on Information Theory (ISIT), Yokohama, Japan, 29 June–4 July 2003; p. 258.
103. Synak-Radtke, B.; Horodecki, M. On asymptotic continuity of functions of quantum states. *J. Phys. A Math. Gen.* **2006**, *39*, L423. [CrossRef]
104. Fannes, M. A continuity property of the entropy density for spin lattice systems. *Commun. Math. Phys.* **1973**, *31*, 291–294. [CrossRef]
105. Winter, A. Tight uniform continuity bounds for quantum entropies: Conditional entropy, relative entropy distance and energy constraints. *Commun. Math. Phys.* **2016**, *347*, 291–313. [CrossRef]
106. Christandl, M.; Winter, A. Squashed entanglement—An additive entanglement measure. *J. Math. Phys.* **2004**, *45*, 829–840. [CrossRef]
107. Christandl, M.; Ekert, A.; Horodecki, M.; Horodecki, P.; Oppenheim, J.; Renner, R. Unifying classical and quantum key distillation. In *Proceedings of the Theory of Cryptography Conference*; Springer: Berlin/Heidelberg, Germany, 2007; pp. 456–478.
108. Maurer, U.M.; Wolf, S. Secret-key agreement over unauthenticated public channels–Part II: The simulatability condition. *IEEE Trans. Inf. Theory* **2003**, *49*, 832–838. [CrossRef]
109. Maurer, U.M.; Wolf, S. Secret-key agreement over unauthenticated public channels–Part I: Definitions and a completeness result. *IEEE Trans. Inf. Theory* **2003**, *49*, 822–831. [CrossRef]
110. Gisin, N.; Wolf, S. Linking classical and quantum key agreement: Is there bound information? In *Proceedings of the Annual International Cryptology Conference*; Springer: Berlin/Heidelberg, Germany, 2000; pp. 482–500.
111. Chitambar, E.; Hsieh, M.H. Round complexity in the local transformations of quantum and classical states. *Nat. Commun.* **2017**, *8*, 2086. [CrossRef]
112. Orlitsky, A.; Wigderson, A. Secrecy enhancement via public discussion. In Proceedings of the IEEE International Symposium on Information Theory (ISIT), San Antonio, TX, USA, 17–22 January 1993; p. 155.
113. Devetak, I.; Harrow, A.W.; Winter, A.J. A resource framework for quantum Shannon theory. *IEEE Trans. Inf. Theory* **2008**, *54*, 4587–4618. [CrossRef]
114. Lieb, E.H.; Yngvason, J. A guide to entropy and the second law of thermodynarnics. *Not. Am. Math. Soc.* **1998**, *45*, 571–581.
115. Watanabe, S. Information theoretical analysis of multivariate correlation. *IBM J. Res. Dev.* **1960**, *4*, 66–82. [CrossRef]
116. Studenỳ, M.; Vejnarová, J. The multiinformation function as a tool for measuring stochastic dependence. In *Learning in Graphical Models*; Springer: Berlin/Heidelberg, Germany, 1998; pp. 261–297.
117. Wennekers, T.; Ay, N. Spatial and temporal stochastic interaction in neuronal assemblies. *Theory Biosci.* **2003**, *122*, 5–18. [CrossRef]
118. Schneidman, E.; Berry, M.J.; Segev, R.; Bialek, W. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* **2006**, *440*, 1007. [CrossRef]
119. Austin, T. Measures of correlation and mixtures of product measures. *arXiv* **2018**, arXiv:1809.10272.

120. Cerf, N.J.; Massar, S.; Schneider, S. Multipartite classical and quantum secrecy monotones. *Phys. Rev. A* **2002**, *66*, 042309. [CrossRef]
121. Prabhakaran, V.M.; Prabhakaran, M.M. Assisted common information with an application to secure two-party sampling. *IEEE Trans. Inf. Theory* **2014**, *60*, 3413–3434. [CrossRef]
122. Horodecki, K.; Horodecki, M.; Horodecki, P.; Oppenheim, J. Locking entanglement with a single qubit. *Phys. Rev. Lett.* **2005**, *94*, 200501. [CrossRef] [PubMed]
123. Galla, T.; Gühne, O. Complexity measures, emergence, and multiparticle correlations. *Phys. Rev. E* **2012**, *85*, 046209. [CrossRef]
124. Amari, S.I. Information geometry on hierarchy of probability distributions. *IEEE Trans. Inf. Theory* **2001**, *47*, 1701–1711. [CrossRef]
125. Schneidman, E.; Still, S.; Berry, M.J., II; Bialek, W. Network information and connected correlations. *Phys. Rev. Lett.* **2003**, *91*, 238701. [CrossRef]
126. Kahle, T.; Olbrich, E.; Jost, J.; Ay, N. Complexity measures from interaction structures. *Phys. Rev. E* **2009**, *79*, 026201. [CrossRef]
127. Linden, N.; Popescu, S.; Wootters, W.K. Almost every pure state of three qubits is completely determined by its two-particle reduced density matrices. *Phys. Rev. Lett.* **2002**, *89*, 207901. [CrossRef]
128. Zhou, D.L. Irreducible multiparty correlations can be created by local operations. *Phys. Rev. A* **2009**, *80*, 022113. [CrossRef]
129. Ay, N.; Jost, J.; Vân Lê, H.; Schwachhöfer, L. *Information Geometry*; Springer: Berlin/Heidelberg, Germany, 2017; Volume 8.
130. Griffith, V.; Chong, E.K.P.; James, R.G.; Ellison, C.J.; Crutchfield, J.P. Intersection Information Based on Common Randomness. *Entropy* **2014**, *16*, 1985–2000. [CrossRef]
131. Venkatesh, P.; Dutta, S.; Grover, P. How should we define information flow in neural circuits? In Proceedings of the IEEE International Symposium on Information Theory (ISIT), Paris, France, 7–12 July 2019; pp. 176–180.
132. Beimel, A. Secret-sharing schemes: A survey. In *Proceedings of the International Conference on Coding and Cryptology*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 11–46.