Article

Channel-Supermodular Entropies: Order Theory and an Application to Query Anonymization [†]

Arthur Américo [‡], MHR Khouzani [‡] and Pasquale Malacaria ^{*,‡}

School of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, UK; a.passosderezende@qmul.ac.uk (A.A.); arman.khouzani@qmul.ac.uk (M.K.)

* Correspondence: p.malacaria@qmul.ac.uk

[†] Presented at the IEEE Information Theory Workshop, Visby, Gotland, Sweden, 25–28 August 2019.

[‡] These authors contributed equally to this work.

Abstract: This work introduces channel-supermodular entropies, a subset of quasi-concave entropies. Channel-supermodularity is a property shared by some of the most commonly used entropies in the literature, including Arimoto–Rényi conditional entropies (which include Shannon and min-entropy as special cases), k-tries entropies, and guessing entropy. Based on channel-supermodularity, new preorders for channels that strictly include degradedness and inclusion (or Shannon ordering) are defined, and these preorders are shown to provide a sufficient condition for the more-capable and capacity ordering, not only for Shannon entropy but also regarding analogous concepts for other entropy measures. The theory developed is then applied in the context of query anonymization. We introduce a greedy algorithm based on channel-supermodularity for query anonymization and prove its optimality, in terms of information leakage, for all symmetric channel-supermodular entropies.

Keywords: information theory; quantitative information flow; channel ordering; more-capable; less-noisy; broadcast channels; anonymity

Citation: Américo, A.; Khouzani, M.; Malacaria, P. Channel-Supermodular Entropies: Order Theory and an Application to Query Anonymization. *Entropy* **2022**, *24*, 39. <https://doi.org/10.3390/e24010039>

Academic Editor: Raúl Alcaraz

Received: 17 November 2021

Accepted: 20 December 2021

Published: 25 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The idea of preorders over channels goes back a long way in the history of information theory. For instance, in [1], Shannon introduced the “inclusion” preorder to compare the capacities of discrete memoryless channels. Several authors, such as El Gamal [2], Korner and Marton [3], and many more, made further significant contributions to the study of channel preorders.

Such preorders are of practical importance in information theory. For example, the “more capable” preorder [3] is used in calculating the capacity region of broadcast channels [2], or in deciding whether a system is more secure than another [4,5]. As discussed in the book by Cohen, Kempermann, and Zbaganu [6], the applications of preorders over stochastic matrices goes beyond the field of information theory, for instance, to statistics, economics, and population sciences.

In this work, which is an extension of the results in our previous work [7], we introduce a new preorder over channels. To illustrate the key idea, consider the following channel:

$$\begin{pmatrix} 0.3 & 0.5 & 0.2 \\ 0.6 & 0.4 & 0 \\ 0.2 & 0.3 & 0.5 \end{pmatrix}.$$

Now build a new channel from it as follows: take the first two columns, and for each row, rearrange their pairwise entries such that the larger element is moved to the first column, and the smaller element is in the second column. This yields:

$$\begin{pmatrix} 0.5 & 0.3 & 0.2 \\ 0.6 & 0.4 & 0 \\ 0.3 & 0.2 & 0.5 \end{pmatrix}.$$

We will refer to this pairwise operation on columns as a *Join-Meet* operation. We prove that, for most commonly used entropy measures, a *Join-Meet* operation always increases the posterior entropy. More precisely, the posterior entropy of the derived channel is never less than the posterior entropy of the original channel, for any probability distribution on the input. That is, the original channel is *more capable* [3] than the derived channel. We name the entropies respecting this property *channel-supermodular*, and prove they entail Arimoto–Rényi entropies (including Shannon and min-entropy), and the guessing entropy, as well as some other entropies that are motivated from security and privacy contexts.

We define the *supermodular* preorder ($\geq_s$) based on the Join-Meet operator. In particular, given channels K_1 and K_2 , we say $K_1 \geq_s K_2$ iff K_2 can be obtained from K_1 via a finite sequence of Join-Meet operations. We establish that the supermodular preorder is neither included nor does it include the *degradedness* [8] or *inclusion* (Shannon) [1] preorders. Motivated by this, we define two other channel preorders ($\geq_{ds}$ and $\geq_{shs}$) that strictly include the aforementioned ones, respectively. The relation $K_1 \geq_{ds} K_2$ implies that K_1 is “more capable” than K_2 . Moreover, whenever $K_1 \geq_{shs} K_2$, then the capacity of K_1 is higher than that of K_2 . Several such new channel ordering results are proven in this paper based on channel-supermodularity.

Next, we will consider the applications of channel-supermodularity in the context of security and privacy. The starting point will be the channel design problem, which is the problem of designing a channel that leaks the least amount of confidential information while respecting a set of operational constraints. This kind of problem arises in many security systems, such as authentication systems [9], operating systems functions [10], scheduling protocols [11], bucketing schemes in cryptography [12], anonymity (Section 6.2), and so on. In the context of these applications, the problem is particularly interesting for deterministic systems and deterministic solutions. Solutions which are unique across many measures of leakage are also of interest because they are robust against how the knowledge or abilities of the attackers are modeled. In this work, we present a robust anonymity mechanism. The algorithm is based on a result from [13], which uses the properties of channel-supermodularity presented in this paper to derive a greedy channel design algorithm that is provably optimal and unique for all channel-supermodular measures of leakage. We apply our robust anonymity mechanism to query anonymization: we consider the problem in which the real query itself is the secret, and we also consider the scenario where a related attribute is the secret. We provide optimal solutions for these two problems based on channel-supermodularity.

1.1. Related Literature

In the “information theory” literature, the degradedness order was introduced by Cover [8] in the study of broadcasting channels. Cover conjectured a solution for the capacity region of broadcast channels that satisfy the degradedness ordering, which was proved by Bergmans [14,15] and Gallager [16]. The problem of determining the capacity region of broadcast channels also motivated Korner and Marton to introduce the H_1 -less noisy and H_1 -more capable orderings [3]. In the same paper, Korner and Marton established the capacity region for broadcast channels that respect the H_1 -less noisy ordering. A similar result for broadcast channels respecting the H_1 -more capable ordering was later established by El Gamal [2].

Those orderings also play an important role in the field of *quantitative information flow* (QIF), which is concerned with quantifying information leakage in computational systems (we refer to [17] for a review of QIF). Malacaria [18] makes use of degradedness ordering to reason about the security of deterministic programs, proving it is equivalent to the H_1 , H_∞ , and H_G -more capable orderings for deterministic channels. This ordering also

appears in the work of Alvim et al. [19], in which it is shown to imply the more capable ordering for the g -entropy family, a generalizing framework for information measures used in QIF. In the same paper, they conjectured that if two channels satisfy the more capable ordering for all members of the g -entropy family, they satisfy the degradedness order. This conjecture, which was proven by McIver et al. [20], turns out to be equivalent to Blackwell's theorem in the finite setting [21]. The less noisy ordering also appears recently in QIF literature, especially the H_∞ -less noisy ordering, in the study of *Dalenius leakage* by Bordenabe and Smith [22]. This last work is also closely related to the classical implications of the work by Buscemi [23], which is mainly focused on the H_∞ -less noisy ordering in quantum information theory.

Shannon ordering (or inclusion), which generalizes degradedness ordering, was first introduced by Shannon when studying channels that could be perfectly simulated by other channels [1]. In the same paper, Shannon established that inclusion implies H_1 -capacity ordering. This ordering has been the object of study of several recent works. Inspired by Le Cam's concept of deficiency [24], which is itself related to the degradedness order and Blackwell's theorem, Raginsky [25] defines *Shannon deficiency*, which may be seen as a measure of how far two channels are from satisfying the inclusion order. Techniques to verify Shannon ordering, both algebraic and computational, were studied by Zhang and Tepedelenlioğlu [26], and Nasser [27] gave two different characterizations of the Shannon ordering.

The abstract problem of designing a system that leaks the least amount of information under some generalized form of operational constraints has been the objective of recent exploration in the literature [28–30]. The problem of optimal system design in security settings has been studied within specific contexts, including secure multi-party computation systems [31] and countermeasures against timing attacks [12,32]. The general channel design problem is of particular significance in QIF, as it represents a paradigm shift from the earlier foundational research in the area, which focuses mostly on measuring information leakage for existing channels or systems [4,5,19,33,34].

Query obfuscation and anonymity have been investigated by several authors and even implemented in commercial products. Related to our work are algorithms presented in [35–37]. Compared to those works, our approach follows an order-theoretical methodology and is based on a more general notion of entropy.

All the results in this paper rely on the concept of core-concave entropies, a generalizing framework that has been recently developed and can be shown to generalize the most commonly used conditional entropy measures in the literature [28]. Core-concavity is a generalization of concavity, which has also been considered as a defining or desirable property for generalized information measures in QIF [38] and in information theory [39–42].

1.2. Notational Conventions

Throughout the paper, $X, Y, Z, \dots$ represent discrete random variables with (nonempty, finite) alphabets $\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \dots$. We assume that the elements of each alphabet are ordered, denoting by $x_1, \dots, x_{|\mathcal{X}|}$ the elements of $\mathcal{X}$, by $y_1, \dots, y_{|\mathcal{Y}|}$ the elements of $\mathcal{Y}$, and so on. Given $x_i \in \mathcal{X}$, we write $p(x_i)$ or p_i to mean $\Pr\{X = x_i\}$, and use p to refer to the (categorical) distribution (as a vector). We may specify the r.v. with a subscript, for example, writing $p_X(x)$, if it is not clear from the context.

We denote by $\Delta_n \subset \mathbb{R}^n$ the $(n - 1)$ -dimensional probability simplex. Given a probability distribution p over $\{x_1, \dots, x_n\}$, we overload the notation and use p to refer to its probability vector $(p_1, \dots, p_n) \in \Delta_n$. We write $(p_{[1]}, p_{[2]}, \dots, p_{[n]})$ for the nonincreasing rearrangement of $p = (p_1, \dots, p_n)$ —that is, $p_{[1]}$ denotes largest element of p , $p_{[2]}$ the second largest, and so forth. Given a vector $\mathbf{r} = (r_1, \dots, r_n) \in \mathbb{R}^n$, we denote by $\|\mathbf{r}\|_\alpha$ its α -norm $(\sum r_i^\alpha)^{1/\alpha}$ (note that this is a slight abuse of nomenclature, since it is not a norm when $\alpha < 1$).

Given a function F over Δ_n and a random variable X with distribution $p = (p_1, \dots, p_n)$, we use $F(X)$, $F(p_1, \dots, p_n)$ and $F(p)$ interchangeably.

A channel K is a row stochastic matrix with rows indexed by $\mathcal{X}$ and columns indexed by $\mathcal{Y}$. The value $K(x, y)$ is equal to $p(y|x) = \Pr\{Y = y \mid X = x\}$, that is, the conditional probability that y is produced by the channel K when x is the input value. The notation $K : \mathcal{X} \rightarrow \mathcal{Y}$ means that the channel K has $\mathcal{X}$ and $\mathcal{Y}$ as input and output alphabets, respectively.

$$p(x, y) = p_X(x)K(y|x).$$

Channels are represented in a table format, or simply in the matrix notation, for example:

$$\begin{array}{c|cc} K & y_1 & y_2 \\ \hline x_1 & 0.5 & 0.5 \\ x_2 & 0.6 & 0.4 \\ x_3 & 0.2 & 0.8 \end{array}, \begin{pmatrix} 0.5 & 0.5 \\ 0.6 & 0.4 \\ 0.2 & 0.8 \end{pmatrix},$$

with the understanding that the i th row corresponds to x_i , and the j th column to y_j .

2. Preliminaries

2.1. Core-Concave Entropies

The main result this work provides is the monotonicity of conditional entropy with regard to the Join-Meet operator, which will be introduced in Section 3. This holds not only for Shannon entropy, but also for a number of different entropy measures, including the Arimoto–Rényi entropies (which include Shannon and min-entropy as limit cases) [43], and the guessing entropy [44].

Most of the results in this paper concern the aforementioned entropies, which are instances of what we call *channel-supermodular* entropies. To define a channel supermodular entropy, however, we first need a generalizing framework. To this end, we introduce the core-concave entropies based on the framework introduced in [28]. Besides the ones aforementioned, they also include the Tsallis [45] and Sharma–Mittal [46] entropies.

Definition 1. A “core-concave” entropy H is a pair (η, F) such that

- $F : \Delta_n \rightarrow \mathbb{R}$ is a concave and continuous function,
- η is a strictly increasing continuous real-valued function, defined on the image of the function F .

Given a core-concave $H = (\eta, F)$, we define $H(X) = \eta(F(p_X))$. The set of core-concave entropies will be denoted by $\mathcal{H}$.

Besides the value $H(X)$, which we refer to as the *unconditional form*, we also define a conditional form for the core-concave entropies, with relation to two random variables X, Y .

Definition 2. The “conditional form” of a core-concave entropy $H = (\eta, F)$ is defined as

$$H(X|Y) = \eta \left(\sum_{y \in \text{supp}(Y)} p(y) F(X|y) \right)$$

where $\text{supp}(Y)$ is the support of Y .

As claimed before, core-concave entropies encompass the most common entropies in the literature. Some of these are summarized in Table 1, together with their conditional form. Notice that H is used to denote an arbitrary core-concave entropy, while H_1 refers to Shannon entropy.

Table 1. Some examples of core-concave entropies.

	$\eta(r)$	$F(p)$	Conditional Form $H(X Y)$	
Shannon (H_1)	r	$-\sum_i p_i \log(p_i)$	$\sum_y p(y) H_1(X y)$	
Min-entropy (H_∞)	$-\log(-r)$	$-\max_i p_i$	$-\log\left(\sum_y p(y) \max_x p_{X Y}(x)\right)$	
Guessing [44] (H_G)	r	$\sum_i i p_{[i]}$	$\sum_y p(y) H_G(X y)$	
Arimoto–Rényi [43] (H_α)	$\frac{\alpha}{1-\alpha} \log(r)$ $\frac{\alpha}{1-\alpha} \log(-r)$	$\ p\ _\alpha$ $-\ p\ _\alpha$	if $0 < \alpha < 1$ if $\alpha > 1$	$\frac{\alpha}{1-\alpha} \log\left(\sum_y p(y) \ p_{X Y}\ _\alpha\right)$
Hayashi–Rényi [47] (H'_α)	$\frac{1}{1-\alpha} \log(r)$ $\frac{1}{1-\alpha} \log(-r)$	$\ p\ _\alpha^\alpha$ $-\ p\ _\alpha^\alpha$	if $0 < \alpha < 1$ if $\alpha > 1$	$\frac{1}{1-\alpha} \log\left(\sum_y p(y) \ p_{X Y}\ _\alpha^\alpha\right)$
Tsallis [45] ($H_{(\alpha,\alpha)}$)	$\frac{1}{\alpha-1}(1-r)$ $\frac{1}{\alpha-1}(1+r)$	$\ p\ _\alpha^\alpha$ $-\ p\ _\alpha^\alpha$	if $0 < \alpha < 1$ if $\alpha > 1$	$\frac{1}{\alpha-1} \left(1 - \left(\sum_y p(y) \ p_{X Y}\ _\alpha^\alpha\right)\right)$
Sharma–Mittal [46] ($H_{(\alpha,\beta)}$)	$\frac{1}{\beta-1} \left(1 - r^{\frac{1-\beta}{1-\alpha}}\right)$ $\frac{1}{\beta-1} \left(1 - (-r)^{\frac{1-\beta}{1-\alpha}}\right)$	$\ p\ _\alpha^\alpha$ $-\ p\ _\alpha^\alpha$	if $0 < \alpha < 1$ if $\alpha > 1$	$\frac{1}{\beta-1} \left(1 - \left(\sum_y p(y) \ p_{X Y}\ _\alpha^\alpha\right)^{\frac{1-\beta}{1-\alpha}}\right)$

Based on Definitions 1 and 2, we define a notion of mutual information and channel capacity for each $H \in \mathcal{H}$.

Definition 3. Let $H \in \mathcal{H}$. The H -mutual information is defined as

$$I_H(X; Y) = H(X) - H(X|Y).$$

When X and Y are respectively the input and output of a channel K , the H -channel capacity of K is defined as

$$C_H(K) = \max_{p_X} I_H(X; Y).$$

Notice that, in general, H -mutual information is not symmetric. Core-concave entropies satisfy the data-processing inequality.

Theorem 1 ([28] Proposition 2(b)). Let X, Y, Z be random variables such that $X \rightarrow Y \rightarrow Z$ (i.e., X and Z are conditionally independent given Y). Then, for all $H \in \mathcal{H}$,

$$H(X|Y) \leq H(X|Z).$$

2.2. Preorders over Channels

Let $K_1 : \mathcal{X} \rightarrow \mathcal{Y}$, $K_2 : \mathcal{X} \rightarrow \mathcal{Z}$ be channels which share an input X and produce outputs Y, Z .

A channel K_2 is degraded from K_1 [8], written as $K_1 \geq_d K_2$, if there exists a channel $R : \mathcal{Y} \rightarrow \mathcal{Z}$ such that $K_2 = K_1 R$.

In [1], Shannon introduced a preorder which includes the one above. A channel K_1 includes K_2 , written as $K_1 \geq_{sh} K_2$, if there exists a family of tuples $\{(g_i, T_i, R_i)\}_i$ of channels T_i, R_i and non-negative real numbers g_i such that:

$$K_2 = \sum_i g_i T_i K_1 R_i \text{ and } \sum_i g_i = 1. \quad (1)$$

As noted by Shannon, any channel can be expressed as a convex combination of deterministic channels. Thus, whenever $K_1 \geq_{sh} K_2$, it is possible to choose $\{(g_i, T_i, R_i)\}_i$ such that T_i, R_i are deterministic channels and (1) holds.

The two preorders defined above are not dependent on a particular entropy but on the structure of the channel. The next preorders depend on the choice of a core-concave entropy H and generalize preorders introduced in [3].

A channel K_1 is H -less noisy than K_2 , denoted as $K_1 \geq_{\text{ln}}^H K_2$, if for all random variables U with finite support such that $U \rightarrow X \rightarrow (Y, Z)$,

$$I_H(U; Y) \geq I_H(U; Z).$$

K_1 is H -more capable than K_2 , written as $K_1 \geq_{\text{mc}}^H K_2$, if for all distributions of the input, X ,

$$I_H(X; Y) \geq I_H(X; Z).$$

Finally, $K_1 \geq_{\text{c}}^H K_2$ stands for

$$C_H(K_1) \geq C_H(K_2).$$

If $\mathcal{A} \subset \mathcal{H}$ is a subset of core-concave entropies, $K_1 \geq_{\text{ln}}^{\mathcal{A}} K_2$ is defined as:

$$\forall H \in \mathcal{A}, K_1 \geq_{\text{ln}}^H K_2.$$

This is similar for $\geq_{\text{mc}}^{\mathcal{A}}$ and $\geq_{\text{c}}^{\mathcal{A}}$.

2.3. Relationships between Preorders

We now explore some relationships between the preorders defined above.

Proposition 1. For any $H \in \mathcal{H}$:

$$K_1 \geq_{\text{d}} K_2 \Rightarrow K_1 \geq_{\text{ln}}^H K_2 \Rightarrow K_1 \geq_{\text{mc}}^H K_2 \Rightarrow K_1 \geq_{\text{c}}^H K_2.$$

Proof. The first implication follows from Theorem 1, the second implication follows by choosing $U = X$ in the definition of $\geq_{\text{ln}}^H$, and the third is straightforward. $\square$

While the reverse of the above implications is false, the following is true:

Theorem 2 ([23]). $K_1 \geq_{\text{ln}}^{H_\infty} K_2 \Rightarrow K_1 \geq_{\text{d}} K_2$.

The following important theorem can be traced back to Blackwell's result on comparison of experiments [21] and relates to more recent results in [20,48].

Theorem 3. $K_1 \geq_{\text{d}} K_2 \Leftrightarrow K_1 \geq_{\text{mc}}^{\mathcal{H}} K_2$.

3. Channel-Supermodular Entropies

Note that any specific core-concave entropy H induces a H -more capable preorder over channels. However, this preorder might not be preserved for a different choice of conditional entropy. Theorem 3 characterizes a channel preorder that is "consistent" for all core-concave entropies. As strong as this result is, it still leaves the question whether there exists a preorder that is consistent for a class of entropies of interest. This is motivated by the fact that the class of core-concave entropies include far more entropies than the conventionally used ones, so it may include some eccentric ones that can be excluded for a stronger result. Moreover, the degradedness relation between channels seems very restrictive: there are many channels that cannot be compared with respect to degradedness, but have consistent ordering with respect to all typically used entropies.

With these motivations in mind, we introduce *channel-supermodular entropies*. Channel-supermodularity is a property satisfied by a significant portion of commonly used entropies, being a helpful tool in optimization problems (as shown in Section 6.1).

The characterization of channel-supermodular entropies is linked to *supermodular functions* over the real lattice. These functions and some basic properties are introduced next. For details about supermodular functions please refer to [49] and [50] (Chapter 6.D).

Consider the set $\mathbb{R}_{\geq 0}^n$ of all n -dimensional vectors with no negative entries (i.e., the non-negative orthant of $\mathbb{R}^n$). Let $\preceq$ represent the element-wise inequality, that is, given $\mathbf{r} = (r_1, \dots, r_n)$ and $\mathbf{s} = (s_1, \dots, s_n)$, $\mathbf{r} \preceq \mathbf{s}$ iff $r_i \leq s_i$ for all i . $\preceq$ is a partial order on $\mathbb{R}_{\geq 0}^n$. In fact $\preceq$ defines a lattice over $\mathbb{R}_{\geq 0}^n$, whose join $\vee$ and meet $\wedge$ operations are defined as:

$$\begin{aligned}\mathbf{r} \vee \mathbf{s} &= (\max(r_1, s_1), \dots, \max(r_n, s_n)), \\ \mathbf{r} \wedge \mathbf{s} &= (\min(r_1, s_1), \dots, \min(r_n, s_n)).\end{aligned}$$

Recall that:

Definition 4. A function $\phi : \mathbb{R}_{+}^n \rightarrow \mathbb{R}$ is supermodular (over a lattice) if, for all $\mathbf{r}, \mathbf{s} \in \mathbb{R}_{+}^n$,

$$\phi(\mathbf{r} \vee \mathbf{s}) + \phi(\mathbf{r} \wedge \mathbf{s}) \geq \phi(\mathbf{r}) + \phi(\mathbf{s}).$$

Next, we introduce some fundamental definitions for this work:

Definition 5. Let $H = (\eta, F)$ be a core-concave entropy. Define the function $G_F : \mathbb{R}_{\geq 0}^n \rightarrow \mathbb{R}$ as

$$G_F(\mathbf{r}) := \|\mathbf{r}\|_1 F\left(\frac{\mathbf{r}}{\|\mathbf{r}\|_1}\right)$$

if $\mathbf{r}$ is not the null vector, and $G_F(0, \dots, 0) = 0$. (Notice that, as F is continuous over a compact set, $\lim_{\mathbf{r} \rightarrow (0, \dots, 0)} G_F(\mathbf{r}) = 0$.)

Definition 6. An entropy $H = (\eta, F) \in \mathcal{H}$ is said to be channel-supermodular if G_F is supermodular. The set of channel-supermodular entropies is noted by $\mathcal{S} \subset \mathcal{H}$.

The motivation for defining channel-supermodularity in terms of G_F might seem arbitrary, but it is justified for its relationship with conditional entropies, given by

$$H(X|Y) = \eta \left(\sum_{y \in \mathcal{Y}} G_F(p_{(X,Y)}(x_1, y), \dots, p_{(X,Y)}(x_n, y)) \right). \quad (2)$$

Together with (2), the supermodularity of G_F can be a powerful tool for deriving results regarding conditional entropy and mutual information for entropies in $\mathcal{S}$.

3.1. Examples of Channel-Supermodular and Non-Channel-Supermodular Entropies

In the next sections, we will study the implications of channel-supermodularity. The inequality in Definition 4 implies interesting behaviors regarding H -mutual information, as will be seen in Section 4. This property has immediate consequences for channel ordering and channel design, as will be explored in Sections 5 and 6.

One appealing aspect of channel-supermodularity is that some of the most commonly used entropies in the literature belong to $\mathcal{S}$, including Shannon and min-entropy, and more generally the Arimoto–Rényi entropies. In this section, we prove that these (and other entropies) indeed belong to $\mathcal{S}$, and provide examples of entropies that do not. First, we state a useful characterization of supermodular functions, which is an immediate consequence of Corollary 2.6.1 in [49].

Let $\phi : \mathbb{R}_{\geq 0}^n \rightarrow \mathbb{R}$ and let $\mathbf{e}_1, \dots, \mathbf{e}_n$ denote the canonical basis of $\mathbb{R}^n$. The function ϕ is supermodular if and only if, for all $\mathbf{r} \in \mathbb{R}_{+}^n$, all $\delta_1, \delta_2 \geq 0$ and all i, j with $i \neq j$,

$$\phi(\mathbf{r} + \delta_1 \mathbf{e}_i + \delta_2 \mathbf{e}_j) + \phi(\mathbf{r}) \geq \phi(\mathbf{r} + \delta_1 \mathbf{e}_i) + \phi(\mathbf{r} + \delta_2 \mathbf{e}_j). \quad (3)$$

Moreover, if ϕ has second partial derivatives, ϕ is supermodular if and only if, for all $\mathbf{r} \in \mathbb{R}_+^n$ and all i, j with $i \neq j$,

$$\frac{\partial^2 \phi(\mathbf{r})}{\partial r_i \partial r_j} \geq 0. \quad (4)$$

The property characterized by Equations (3) and (4) is known in the economics literature as *increasing differences* [49]. This name is due to the effect an increase on a coordinate has on the value of ϕ being monotonically increasing with regard to the other coordinates. This is readily noticeable if we rearrange the terms of (3):

$$\phi(\mathbf{r} + \delta_1 \mathbf{e}_i + \delta_2 \mathbf{e}_j) - \phi(\mathbf{r} + \delta_2 \mathbf{e}_j) \geq \phi(\mathbf{r} + \delta_1 \mathbf{e}_i) - \phi(\mathbf{r}).$$

That is, the change of ϕ prompted by an increase of δ_1 in coordinate i is greater the greater the value of coordinate j . Equation (3) is thus just the statement that on the lattice $(\mathbb{R}_{\geq 0}^n, \preceq)$ increasing differences and supermodularity are equivalent concepts ([49] Corollary 2.6.1).

Using this result, as well as appealing directly to Definition 4, we now prove channel-supermodularity for a number of commonly used entropies. Throughout, given $\mathbf{r} \in \mathbb{R}_{\geq 0}^n$, we denote by r_i its i th coordinate.

Proposition 2.

1. Shannon entropy is channel-supermodular.
2. Arimoto–Rényi entropies are channel-supermodular for all $\alpha \in (0, 1) \cup (1, \infty)$.
3. For any k , the k -tries entropy is channel-supermodular. In particular, min-entropy is channel-supermodular.
4. Guessing entropy is channel-supermodular.

Proof. See Appendix A. $\square$

Items 3 and 4 of Proposition 2 are of particular interest to security applications, as guessing entropy and k -tries entropies have found interesting applications in the field of quantitative information flow.

Guessing entropy is especially useful in scenarios modelling brute-force attacks, as it models the expected number of attempts necessary for an adversary to obtain the value of a secret when trying one by one. On the other hand, k -tries entropy reflects the probability of guessing a value correctly when k guesses are allowed (see e.g., [19] (Section III.C)). It is defined as

$$H_{k\text{-tries}}(p) = -\log \sum_{i=1}^k p_{[i]},$$

and can be readily seen to be core-concave by taking $\eta(x) = -\log(-x)$ and $F(p) = -\sum_{i=1}^k p_{[i]}$. Notice that min entropy is equal to $H_{k\text{-tries}}$ when $k = 1$.

The results in Proposition 2 justify our interest in channel-supermodularity, as any property derived for entropies in $\mathcal{S}$ will also hold for this set of commonly used entropy measures. However, not all entropies are channel-supermodular.

This includes another interesting entropy family useful in security, the *partition entropies* [19]. Let $\mathcal{P}$ be a partition of the set $\{1, \dots, n\}$. The partition entropy with regard to $\mathcal{P}$ is given by

$$H_{\mathcal{P}}(X) = -\log \max_{\mathcal{A} \in \mathcal{P}} \sum_{i \in \mathcal{A}} p(x_i).$$

It is easy to see that $H_{\mathcal{P}}$ is core-concave, by taking $\eta(x) = -\log(-x)$ and $F(p) = -\max_{\mathcal{A} \in \mathcal{P}} \sum_{i \in \mathcal{A}} p_i$. The partition entropy $H_{\mathcal{P}}$ is useful for capturing the uncertainty of an *adversary* that is interested in knowing only to which subset the realization of X pertains. This is an appropriate model for adversaries that are interested in obtaining some specific partial knowledge about some sensitive information (e.g., obtaining the home town or the DOB of a user).

Proposition 3.

1. Hayashi–Rényi, Tsallis and Sharma–Mittal entropies (with conditional forms as in Table 1) are not channel-supermodular for all $\alpha > 1$ whenever the input set is of size greater than 2. Moreover, they are also not channel-supermodular for all $\alpha \in (0, 1)$ for some size of input set.
2. The partition entropy is not, in general, channel-supermodular.

Proof. See Appendix A. $\square$

Notice that, for some choices of $\mathcal{P}$, $H_{\mathcal{P}}$ is channel-supermodular. In particular, if $\mathcal{P} = \{\{i\} \mid i \in \{1, \dots, n\}\}$, $H_{\mathcal{P}}$ coincides with min-entropy.

A generalization of partition entropy is the *weighted partition entropies*,

$$H_{\mathcal{P}, \mathbf{w}}(X) = -\log \max_{\mathcal{A} \in \mathcal{P}} \sum_{i \in \mathcal{A}} p(x_i) w_i$$

where $\mathbf{w} = (w_1, \dots, w_n) \in \mathbb{R}_{\geq 0}^n$ is a set of weights. Being a generalization of partition entropies, weighted partition entropies are also not channel-supermodular in general.

4. The Join-Meet Operator and a New Structural Order

In this section we address the claim made in Section 1, proving that the *Join-Meet* operation is monotonic with regard to conditional entropy for all channel-supermodular entropies.

Let $K : \mathcal{X} \rightarrow \mathcal{Y}$ be a channel, with $\mathcal{Y} = \{y_1, \dots, y_m\}$, and let K^i be the column of K corresponding to output y_i . Define, for $i \neq j$, the *Join-Meet* operator $\diamond_{i,j}$ as follows:

$$(\diamond_{i,j} K)^l = \begin{cases} K^i \vee K^j & \text{if } l = i, \\ K^i \wedge K^j & \text{if } l = j, \\ K^l & \text{otherwise.} \end{cases}$$

The next result proves that the Join-Meet operator is monotonic with I_H if $H \in \mathcal{S}$.

Theorem 4. For all channels K_1 and all i, j , $K_1 \geq_{\text{mc}}^{\mathcal{S}} \diamond_{i,j} K_1$.

Proof. Let $H = (\eta, F) \in \mathcal{S}$ and define G_F as in Definition 5. Let $K_2 = \diamond_{i,j} K_1$, and denote by Y_1, Y_2 the outputs of K_1, K_2 . Notice that, for any distribution on the input, we have $p_{X,Y_2}(x_k, y_i) = \max(p_{X,Y_1}(x_k, y_i), p_{X,Y_1}(x_k, y_j))$, and, similarly, $p_{X,Y_2}(x_k, y_j) = \min(p_{X,Y_1}(x_k, y_i), p_{X,Y_1}(x_k, y_j))$. Thus,

$$\begin{aligned} & \sum_l G_F(p_{X,Y_2}(x_1, y_l), \dots, p_{X,Y_2}(x_n, y_l)) \\ & \geq \sum_{l=i,j} G_F(p_{X,Y_1}(x_1, y_l), \dots, p_{X,Y_1}(x_n, y_l)) \\ & \quad + \sum_{l \neq i,j} G_F(p_{X,Y_2}(x_1, y_l), \dots, p_{X,Y_2}(x_n, y_l)) \\ & = \sum_l G_F(p_{X,Y_1}(x_1, y_l), \dots, p_{X,Y_1}(x_n, y_l)) \end{aligned}$$

where the inequality follows from G_F being supermodular. From Equation (2) and η being increasing, it follows that $H(X|Y_1) \leq H(X|Y_2)$, which is equivalent to $I_H(X; Y_1) \geq I_H(X; Y_2)$. $\square$

In light of Theorem 4, one might wonder if the Join-Meet operator completely defines $\mathcal{S}$ —that is, whether $H \in \mathcal{S}$ whenever $K \geq_{\text{mc}}^H \diamond_{i,j} K$ for all channels K and all i, j . In fact, an even stronger statement can be made by only considering a subset of channels.

Definition 7. Let $K_{(k,l,\epsilon_1,\epsilon_2)}$ denote the channel with input alphabet $\{x_1, \dots, x_n\}$ and output alphabet $\{y_1, y_2\}$, given by

$$K_{(k,l,\epsilon_1,\epsilon_2)}(y|x) = \begin{cases} 1 - \epsilon_1 & \text{if } x = x_k, y = y_1, \\ \epsilon_1 & \text{if } x = x_k, y = y_2, \\ \epsilon_2 & \text{if } x = x_l, y = y_1, \\ 1 - \epsilon_2 & \text{if } x = x_l, y = y_2, \\ 1/2 & \text{otherwise.} \end{cases}$$

Theorem 5. Let $H = (\eta, F) \in \mathcal{H}$. If for all $k, l \leq n$, and all $\epsilon_1, \epsilon_2 \in [0, 1/2)$, $K_{(k,l,\epsilon_1,\epsilon_2)} \geq_{\text{mc}}^H \diamond_{1,2} K_{(k,l,\epsilon_1,\epsilon_2)}$, then $H = (\eta, F) \in \mathcal{S}$.

Proof. We prove the contrapositive. Suppose that $H = (\eta, F) \notin \mathcal{S}$. Then, from (3), there are $\mathbf{r} = (r_1, \dots, r_n) \in \mathbb{R}_{\geq 0}^n$, $i, j \leq n$ with $i \neq j$ and $\delta_1, \delta_2 > 0$ such that

$$G_F(\mathbf{r} + \delta_1 \mathbf{e}_i + \delta_2 \mathbf{e}_j) + G_F(\mathbf{r}) < G_F(\mathbf{r} + \delta_1 \mathbf{e}_i) + G_F(\mathbf{r} + \delta_2 \mathbf{e}_j).$$

Let $\gamma = (2\|\mathbf{r}\|_1 + \delta_1 + \delta_2)^{-1}$ and define a probability distribution over X by

$$p_X(x) = \begin{cases} \gamma(2r_i + \delta_1), & \text{if } x = x_i, \\ \gamma(2r_j + \delta_2), & \text{if } x = x_j, \\ 2\gamma r_l, & \text{if } x = x_l, l \neq i, j. \end{cases}$$

Let $\epsilon_1 = \frac{r_i}{(2r_i + \delta_1)}$ and $\epsilon_2 = \frac{r_j}{(2r_j + \delta_2)}$, and let $K_1 = K_{(i,j,\epsilon_1,\epsilon_2)}$, $K_2 = \diamond_{1,2} K_1$. Then,

$$\begin{aligned} H(X|Y_1) &= \eta\left(\gamma(G_F(\mathbf{r} + \delta_1 \mathbf{e}_i) + G_F(\mathbf{r} + \delta_2 \mathbf{e}_j))\right), & \text{and} \\ H(X|Y_2) &= \eta\left(\gamma(G_F(\mathbf{r} + \delta_1 \mathbf{e}_i + \delta_2 \mathbf{e}_j) + G_F(\mathbf{r}))\right), \end{aligned}$$

and thus, as η is strictly increasing, $H(X|Y_1) > H(X|Y_2)$, which concludes the proof. $\square$

An immediate consequence of Theorems 4 and 5 is that the Join-Meet operator completely characterizes $\mathcal{S}$.

Corollary 1. Let $H \in \mathcal{H}$. $H \in \mathcal{S}$ if, and only if, $K \geq_{\text{mc}}^H \diamond_{i,j} K$ for all channels K and all i, j .

A New Structural Ordering

Theorem 4 yields some immediate new results for reasoning about channel ordering, as, whenever $|\mathcal{X}| > 2$, the Join-Meet operator is not, in general, captured by the degradedness ordering. Consider, for instance, the following channels K_1, K_2 (notice that $K_1 = K_{(1,2,0,0)}$).

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1/2 & 1/2 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1/2 & 1/2 \end{pmatrix} \quad (5)$$

Then, we have that $K_2 = \diamond_{1,2} K_1$, but clearly $K_1 \not\geq_d K_2$. To see this, fix a channel $R : \mathcal{Y}_1 \rightarrow \mathcal{Y}_2$. Because R is a channel, there are $p, q \in [0, 1]$ such that

$$R = \begin{pmatrix} p & 1-p \\ q & 1-q \end{pmatrix}.$$

Then, we have

$$K_1 R = \begin{pmatrix} p & 1-p \\ q & 1-q \\ (1/2)(p+q) & (1/2)(2-p-q) \end{pmatrix}.$$

Therefore, $K_2 \neq K_1 R$ for any choice of p, q .

We formalize this observation in the next result.

Proposition 4.

1. If $|\mathcal{X}| = 2$, then, for all $K : \mathcal{X} \rightarrow \mathcal{Y}$ and all $i, j \leq |\mathcal{Y}|$, $K \geq_d \diamond_{i,j} K$.
2. If $|\mathcal{X}| > 2$, then there are $K : \mathcal{X} \rightarrow \mathcal{Y}$ and $i, j \leq |\mathcal{Y}|$ such that $K \not\geq_d \diamond_{i,j} K$.

Proof. We first prove (1). As it is possible to reorder columns by degrading a channel, without loss of generality let $i = 1$ and $j = 2$. Fix $K : \mathcal{X} \rightarrow \mathcal{Y}$. Let $K_{kl} = K(y_l|x_k)$, and suppose, again without loss of generality, that $K_{11} \geq K_{12}$ and $K_{22} \geq K_{21}$.

If $K_{11} = K_{12}$ or $K_{22} = K_{21}$, then $\diamond_{1,2} K$ is obtainable by permutating columns of K , and therefore $K \geq_d \diamond_{1,2} K$. Otherwise, we have $K_{11}K_{22} - K_{12}K_{21} > 0$, and $\diamond_{1,2} K = KR$ where R is the following channel:

R	y_1	y_2	y_3	$\cdots$	y_m
y_1	$\frac{K_{22}K_{11}-K_{22}K_{12}}{K_{11}K_{22}-K_{12}K_{21}}$	$\frac{K_{22}K_{12}-K_{12}K_{21}}{K_{11}K_{22}-K_{12}K_{21}}$	0	$\cdots$	0
y_2	$\frac{K_{11}K_{22}-K_{11}K_{21}}{K_{11}K_{22}-K_{12}K_{21}}$	$\frac{K_{11}K_{21}-K_{12}K_{21}}{K_{11}K_{22}-K_{12}K_{21}}$	0	$\cdots$	0
y_3	0	0	1	$\cdots$	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
y_m	0	0	0	$\cdots$	1

For the proof of (2), it suffices to notice that, whenever $|\mathcal{X}| > 2$, $K_{(1,2,0,0)} \not\geq_d \diamond_{1,2} K_{(1,2,0,0)}$. The proof for general $|\mathcal{X}|$ is along the same lines of the argument after (5). $\square$

Next we define the channel-supermodularity preorder over channels, which is based on the Join-Meet operators $\diamond_{i,j}$.

Definition 8. $K_1 \geq_s K_2$ if there is a finite collection of tuples (i_k, j_k) such that

$$K_2 = \diamond_{i_1, j_1} (\diamond_{i_2, j_2} (\cdots \diamond_{i_m, j_m} K_1)).$$

An induced preorder can be then defined by combining $\geq_d$ and $\geq_s$ as follows:

Definition 9. $K_1 \geq_{ds} K_2$ if there are channels $W_1, \dots, W_n$ such that $K_1 \geq_0 W_1 \geq_1 \cdots \geq_{n-1} W_n \geq_n K_2$, where each $\geq_i$ stands for $\geq_d$ or $\geq_s$.

5. Relations between Preorders for Channel-Supermodular Entropies

Throughout this section, let $K_1 : \mathcal{X} \rightarrow \mathcal{Y}$ and $K_2 : \mathcal{X} \rightarrow \mathcal{Z}$. First, note that Proposition 1 and Theorem 2 are still meaningful under $\mathcal{S}$. The next proposition summarizes the relationship between $(\geq_{ds})$ and the other preorders.

Proposition 5.

1. $K_1 \geq_d K_2 \Rightarrow K_1 \geq_{ds} K_2$ and $K_1 \geq_s K_2 \Rightarrow K_1 \geq_{ds} K_2$,
2. $K_1 \geq_{ds} K_2 \not\Rightarrow K_1 \geq_d K_2$,
3. $K_1 \geq_{ds} K_2 \not\Rightarrow K_1 \geq_s K_2$,
4. $K_1 \geq_{ds} K_2 \not\Rightarrow K_1 \geq_{\ln}^S K_2$,

5. $K_1 \geq_{\ln}^S K_2 \Rightarrow K_1 \geq_{ds} K_2$,
6. $K_1 \geq_{\ln}^{H_1} K_2 \not\Rightarrow K_1 \geq_{ds} K_2$,
7. $K_1 \geq_{ds} K_2 \Rightarrow K_1 \geq_{mc}^S K_2$,
8. $K_1 \geq_{ds} K_2 \not\Rightarrow K_1 \geq_{sh} K_2$,
9. $K_1 \geq_{sh} K_2 \not\Rightarrow K_1 \geq_{ds} K_2$.

Proof. See Appendix B. $\square$

Proposition 5.7 can be used to decide whether $K_1 \geq_{mc}^H K_2$, for $H \in \mathcal{S}$, by only using structural properties of the channel. Consider, for example, the following channels K_1, K_2 .

$$\begin{pmatrix} 1/2 & 0 & 1/2 \\ 0 & 1/2 & 1/2 \\ 1/2 & 1/2 & 0 \end{pmatrix} \quad \begin{pmatrix} 1/4 & 3/4 \\ 1/4 & 3/4 \\ 3/5 & 2/5 \end{pmatrix}.$$

In [19], the authors claimed to have no proof that $K_1 \geq_{mc}^{H_1} K_2$. By Proposition 5 (7), $K_1 \geq_{mc}^{H_1} K_2$ can be proven as follows:

$$\begin{pmatrix} 1/2 & 0 & 1/2 \\ 0 & 1/2 & 1/2 \\ 1/2 & 1/2 & 0 \end{pmatrix} \geq_s \begin{pmatrix} 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{pmatrix} \geq_d \begin{pmatrix} 1/4 & 3/4 \\ 1/4 & 3/4 \\ 3/5 & 2/5 \end{pmatrix}.$$

Notice that if $\mathcal{H}$ is substituted for $\mathcal{S}$, Theorem 3 does not hold:

Proposition 6. $K_1 \geq_{mc}^S K_2 \not\Rightarrow K_1 \geq_d K_2$ and $K_1 \geq_{mc}^S K_2 \not\Rightarrow K_1 \geq_{\ln}^S K_2$.

Proof. From Proposition 5 (2), there are channels K_1, K_2 such that $K_1 \geq_{ds} K_2$ and $K_1 \not\geq_d K_2$. For such channels, Proposition 5 (7) implies $K_1 \geq_{mc}^S K_2$, and the first result follows. The second result then follows by noting that Proposition 1 and Theorem 2 imply $K_1 \geq_{\ln}^S K_2 \Leftrightarrow K_1 \geq_d K_2$. $\square$

Proposition 7. $K_1 \geq_{mc}^S K_2 \not\Rightarrow K_1 \geq_{sh} K_2$.

Results on Channel Capacity

In [1], Shannon proved that $K_1 \geq_{sh} K_2 \Rightarrow K_1 \geq_c^{H_1} K_2$. We can use Theorem 4 to prove similar results for the preorder $\geq_{shs}$, which is an extension of $\geq_{sh}$ with $\geq_s$.

Definition 10. $K_1 \geq_{shs} K_2$ if there are channels $W_1, \dots, W_n$ such that $K_1 \geq_0 W_1 \geq_1 \dots \geq_{n-1} W_n \geq_n K_2$ where each $\geq_i$ stands for $\geq_{sh}$ or $\geq_s$.

We have:

Proposition 8. For all channels K_1, K_2 ,

1. $K_1 \geq_{sh} K_2 \Rightarrow K_1 \geq_{shs} K_2$,
2. $K_1 \geq_{ds} K_2 \Rightarrow K_1 \geq_{shs} K_2$,
3. $K_1 \geq_{shs} K_2 \not\Rightarrow K_1 \geq_s K_2$,
4. $K_1 \geq_{shs} K_2 \not\Rightarrow K_1 \geq_{sh} K_2$.

Proof. (1) and (2) follow immediately from the definitions of $\geq_{ds}$ and $\geq_{shs}$, and from observing that $\geq_d \Rightarrow \geq_{sh}$. (3) and (4) follow from (2) and Propositions 5 (3) and 5 (8). $\square$

In the remainder of this section, we prove that $K_1 \geq_{shs} K_2$ is a sufficient condition for establishing that both the Shannon and min-capacity of K_1 are at least as large as that of K_2 .

Lemma 1. Let $H \in \mathcal{S}$. If $K_1 \geq_{sh} K_2 \Rightarrow K_1 \geq_c^H K_2$, then $K_1 \geq_{shs} K_2 \Rightarrow K_1 \geq_c^H K_2$.

Proof. Let $H \in \mathcal{S}$. From Propositions 1, 5 (1), and 5 (7), $K_1 \geq_s K_2 \Rightarrow K_1 \geq_c^H K_2$. The result then follows from Definition 10. $\square$

Proposition 9.

1. $K_1 \geq_{\text{shs}} K_2 \Rightarrow K_1 \geq_c^{H_1} K_2$,
2. $K_1 \geq_{\text{shs}} K_2 \Rightarrow K_1 \geq_c^{H_\infty} K_2$,
3. $K_1 \geq_c^{H_1} K_2 \not\Rightarrow K_1 \geq_{\text{shs}} K_2$ and $K_1 \geq_c^{H_\infty} K_2 \not\Rightarrow K_1 \geq_{\text{shs}} K_2$.

Proof. See Appendix B. $\square$

Figure 1 summarizes the implications between orderings. As can be seen, there are some open questions that need to be established, designated by dotted lines with a question mark. Note that the absence of an arrow means that the implication is known to be false.

Figure 1. The four implication graphs summarize the relation between preorders. It is known that a preorder $\geq_i$ implies a preorder $\geq_j$ if and only if there is a path of solid arrows from $\geq_i$ to $\geq_j$. Preorders that are equivalent are grouped together, and the dotted arrows represent an implication whose validity is an open question. When no path is present, the implication is known not to hold.

6. Channel Design

Core-concavity was originally introduced in [51] in the context of universally optimal channel design—that is, the problem of finding, given some operational constraints, a channel leaking the minimum amount of confidential information (optimality), for all entropy-based measures of leakage (universality). This section shows how core-concavity and channel-supermodularity can be used in this context.

If X and Y are the input and output of a channel and H is a core-concave entropy, then the leakage about X through Y as measured by H is defined to be the H -mutual information $I_H(X; Y)$, as in Definition 3.

The concept of leakage is relevant in security/privacy contexts where X is some confidential data, K is modelling a system (e.g., a cryptographic computation or a database query system), and Y some observable generated by the system (e.g., the computation time or the result of a statistical query). Different H corresponds to different attackers' models,

and universal channel solutions identify countermeasures to leakage which are robust with regard to all attackers in that universe.

Minimizing leakage of sensitive information is usually a desirable goal. However, when designing systems, it is often the case that some leakage is unavoidable. With that in mind, some recent works in QIF aimed at obtaining channels that leak the least amount of information subject to some operational constraints [13,28,29]. From Definition 3, the problem can be rephrased as finding the channel which, subject to some operational constraints, maximizes $H(X|Y)$ — or, as η is increasing, maximizes $\sum_y p(y)F(X|y)$. In a recent work [29], which considered a generalizing framework for these operational constraints, it was shown that this problem can be solved by convex optimization techniques, for a given core-concave H . However, it was also proven that the solution to the problem is in general not *universal*—that is, the optimal channel given a set of constraints may vary with the choice of H .

Despite this negative result, it was shown in [29] that some classes of problems admit a universal solution. As different entropies model different attackers, these results provide a very strong security guarantee—namely, that the optimal system in these situations is the most secure possible regardless of the attacker model. In the next few sections, we show how channel-supermodularity can be a useful tool in obtaining solutions that, while not universally optimal for all core-concave entropies, are the most secure for all symmetric entropies in $\mathcal{S}$.

6.1. Deterministic Channel Design Problem: A Universal Solution by Channel-Supermodularity

In many applications, such as repeated queries, it is either undesirable or impractical to consider a “probabilistic” system. This motivates the study of the channel design problem restricted to *deterministic* channels, which has been recently investigated in [13].

It was proven in [13] that, similarly to the general channel design problem, the deterministic version does not in general accept a universal solution. Moreover, the problem was also shown to be, in general, NP-hard. However, it was also proven in this work that a specific class of problems, called the *deterministic complete k -hypergraph design problem*, admit a solution that is optimal for all symmetric channel-supermodular entropies. This problem can be defined as follows.

Definition 11. Let $H \in \mathcal{H}$, $k \in \mathbb{N}_{>0}$ and let $\mathcal{X}, \mathcal{Y}$ be finite sets with $|\mathcal{Y}| \geq |\mathcal{X}|/k$

The *deterministic complete k -hypergraph design problem (CKDP)* is to find a channel $K : \mathcal{X} \rightarrow \mathcal{Y}$ that maximizes $H(X|Y)$, subject to the following constraints:

- $\forall x, y, \quad K(y|x) \in \{0, 1\}$, and
- $\forall y, \quad \sum_x K(y|x) \leq k$.

That is, the deterministic CKDP is the problem of finding the most secure deterministic channel, subject to the constraint that each output can only be generated by at most k inputs.

For the remainder of this section, let $k \in \mathbb{N}_{>0}$, and fix a distribution p_X such that, without loss of generality, $p_X(x_i) \geq p_X(x_j)$ whenever $i < j$.

The greedy solution proposed by [13] is described in Algorithm 1. The algorithm is straightforward: it associates the k most likely secrets with the first observable; it then associates the k most likely secrets among the remaining secrets with the second observable, and so on. The solution for $\mathcal{X} = \{x_1, \dots, x_8\}$ and $k = 3$ is depicted in Figure 2.

Algorithm 1 Greedy algorithm for the k -complete hypergraph problem

Input: Input set $\mathcal{X}$, prior p_X and integer $k \leq |\mathcal{X}|$

Output: Matrix of optimal deterministic channel K_k

- 1: **initialize:** K_k as a matrix of 0s, with $|\mathcal{X}|$ rows and $\lceil |\mathcal{X}|/k \rceil$ columns.
 - 2: **for** $i \in \{1, \dots, |\mathcal{X}|\}$ **do**
 - 3: $K_k(i, \lceil i/k \rceil) = 1$
 - 4: **return** K_k
-

K_k	y_1	y_2	y_3
x_1	1	0	0
x_2	1	0	0
x_3	1	0	0
x_4	0	1	0
x_5	0	1	0
x_6	0	1	0
x_7	0	0	1

Figure 2. The solution given by Algorithm 1 for $k = 3$.

Theorem 6 ([13]). *Given a complete k -hypergraph channel design problem, the solution given by Algorithm 1 is optimal for any symmetric channel-supermodular entropy.*

Proof. We reproduce the proof of this theorem from [13], as it provides an interesting application of channel-supermodularity. Let us consider the following joint matrix J_k obtained by Algorithm 1 and the prior:

$$\begin{pmatrix} p_1 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ p_k & 0 & & \vdots \\ 0 & p_{k+1} & & \vdots \\ \vdots & \vdots & & \vdots \\ \vdots & p_{2k} & & \vdots \\ \vdots & 0 & & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & & 0 \\ \vdots & \vdots & & p_{mk+1} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & p_n \end{pmatrix}$$

We now prove that any matrix J satisfying the constraints can be transformed into J_k by a sequence of steps each increasing (or keeping equal) any supermodular entropy. Each step consists of the following three sub-steps:

- Select two columns c_i, c_j and align the non-zero coefficients in c_i, c_j ;
- Perform $\wedge, \vee$ operations on the aligned columns and replace c_i, c_j with $c_i \vee c_j, c_i \wedge c_j$;
- Disalign the two columns $c_i \vee c_j, c_i \wedge c_j$.

The following example illustrates one step (i.e., the three sub-steps above):

$$\begin{pmatrix} 0 & 0.4 & 0 \\ 0 & 0 & 0.3 \\ 0 & 0.15 & 0 \\ 0 & 0 & 0.1 \\ 0.05 & 0 & 0 \end{pmatrix} \Rightarrow \begin{pmatrix} 0 & 0.4 & 0.1 \\ 0 & 0.15 & 0.3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0.05 & 0 & 0 \end{pmatrix} \Rightarrow \begin{pmatrix} 0 & 0.4 & 0.1 \\ 0 & 0.3 & 0.15 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0.05 & 0 & 0 \end{pmatrix} \Rightarrow \begin{pmatrix} 0 & 0.4 & 0 \\ 0 & 0.3 & 0 \\ 0 & 0 & 0.15 \\ 0 & 0 & 0.1 \\ 0.05 & 0 & 0 \end{pmatrix}$$

from left to right we have:

- selected c_2, c_3 , which are the two columns containing the two most likely priors (i.e., 0.4 and 0.3) and aligned c_2, c_3 so that 0.4, 0.3 appear in different rows;
- replaced c_2, c_3 with $c_2 \vee c_3, c_2 \wedge c_3$;
- disaligned columns 2 and 3, that is, position values in $c_2 \vee c_3, c_2 \wedge c_3$, so that each row has the same probability it had before the step.

Notice that aligning (and disaligning) is a permutation of a column; hence, they do not change the value of the posterior of symmetric channel-supermodular entropies because $G_F(c_i) = G_F(c'_i)$ for any permutation c'_i of the column c_i .

Next, for the remaining sub-step, where we replace c_i, c_j with $c_i \vee c_j, c_i \wedge c_j$, by supermodularity of G we have $G_F(c_i) + G_F(c_j) \leq G_F(c_i \vee c_j) + G_F(c_i \wedge c_j)$; hence, that sub-step increases (or keeps equal) the posterior entropy. Notice also that the matrix at the end of the step has in each row the same probabilities as it had before that step; hence, it is still a joint matrix that respects the complete k -hypergraph constraints for the same prior.

The selection and alignment of columns is as follows: at the initial step select c_i such that c_i contains the first r elements with the highest probabilities, say $p_1, \dots, p_r$; if $r < k$, then select c_j as the column containing p_{r+1} ; align c_i, c_j so that p_{r+1} is not on the same row as any of the $p_1, \dots, p_r$ (and $c_i \vee c_j$ has no more than k non-zero terms). Then, $c_i \vee c_j$ will contain $p_1, \dots, p_{r+1}$. Repeat until $r = k$. Then repeat the process considering the probabilities $p_{k+1}, \dots, p_n$.

By repeating these steps, we will reach a matrix J' with columns $c'_1, \dots, c'_n$ such that each element of column c'_i has higher probability than all elements of column c'_{i+1} . This is exactly the solution given by the greedy algorithm (modulo column permutations), that is, $J' = J_k$. $\square$

If H is not channel-supermodular, the greedy solution may not be optimal. Consider, for example, the Hayashi–Rényi entropies, which are not channel-supermodular. Let $\mathcal{X} = \{x_1, \dots, x_4\}$, $p_X = (0.3, 0.3, 0.2, 0.2)$ and $k = 3$, and consider the channels K_1, K_2 with outputs Y_1, Y_2 below.

K_1	y_1	y_2
x_1	1	0
x_2	1	0
x_3	1	0
x_4	0	1

K_2	y_1	y_2
x_1	1	0
x_2	1	0
x_3	0	1
x_4	0	1

Then, K_1 is the greedy solution. However, for Hayashi–Rényi entropies, the following limit holds [52]:

$$\lim_{\alpha \rightarrow \infty} H'_\alpha(X|Y) = -\log \left(\max_{y \in \sim_{||}(\mathcal{Y}), x \in \mathcal{X}} p_{X|y}(x) \right),$$

Therefore,

$$\lim_{\alpha \rightarrow \infty} H'_\alpha(X|Y_1) = 0,$$

whereas

$$\lim_{\alpha \rightarrow \infty} H'_\alpha(X|Y_2) = 1.$$

Thus, $H'_\alpha(X|Y_1) < H'_\alpha(X|Y_2)$ for large enough α , and the greedy solution is not optimal.

Another example of core-concave functions for which Algorithm 1 is not optimal is provided by “partition” entropies. For example, if B is a partition of the possible values of X then

$$H_B(X) = -\log \max_{S \in B} \sum_{x \in S} p_X(x)$$

is a core-concave entropy which is not channel-supermodular.

6.2. An Application to Query Anonymity

Let us consider the following anonymity mechanism problem: we want to design an anonymity mechanism where in order to conceal a secret query from an eavesdropper, the user sends to a server a set of k queries which includes the secret query. Then, once received from the server the response to all the k queries, the user retrieves the response to the secret query. In our setting, this corresponds to each observable having a pre-image of size exactly k .

As an illustrative example consider a Twitter user who wants to visit some other Twitter user page but wants to keep this query secret. To solve this problem, he decides to use the following protocol: whenever he visits the desired user page, he also sends $k - 1$ other queries to the pages of other Twitter users. Suppose further that this user frequently visits this user's page, meaning that a random choice of the other queries is not a wise strategy, since multiple observations would end up revealing more and more information about the query, eventually completely revealing the secret query. The problem is then: which set of k Twitter pages will leak the least information about the user secret query?

We assume the attacker has no background information about the user, and hence we set the probability of a Twitter query for that user as the probability that a general member of the public requests that Twitter page (a good proxy to this measure can be derived by the number of followers of that Twitter page). Let n be the number of possible queries (i.e., the input set). Considering the scenario that n is divisible by k , we can use Algorithm 1 to solve this problem.

Notice that for n secrets, there are $n! \left(\frac{n}{k}\right)!^{-1} (k!)^{-\frac{n}{k}}$ possible ways to satisfy these anonymity constraints. For example, there are about 7×10^{85} possible solutions when $n = 100$ and $k = 10$, and about $4 \times 10^{19,704}$ for $n = 10,000$ and $k = 100$. We will now compare the greedy algorithm in Section 6 against other possible anonymity solutions, and we will measure the goodness of the solutions using min and Shannon posterior entropies. Let us consider the three anonymity solutions below:

1. the solution from the greedy algorithm (Algorithm 1) (i.e., pick the $k - 1$ queries closest in probability to the real query);
2. a random solution (i.e., pick k random queries);
3. a non-optimal solution where the secrets with the highest probabilities, instead of being grouped in the first bin, are distributed in the other bins.

For example, for 6 secrets and $k = 2$, the greedy solution would be $\{\{x_1, x_2\}, \{x_3, x_4\}, \{x_5, x_6\}\}$, whereas the non-optimal solution (3) would be $\{\{x_1, x_4\}, \{x_2, x_5\}, \{x_3, x_6\}\}$.

The difference between these solutions can be very substantial. Figure 3 shows the values when the distribution over the input set is a binomial distribution, with parameter $p = 0.5$: in this scenario, supposing that there is a universe of 350 Twitter pages and that the user sends 19 queries selected using the greedy algorithm, the probability of an attacker guessing the secret query correctly would be over 7 times smaller than if the user had opted instead for the non-optimal solution (using $2^{-H_\infty(X|Y)}$ as the conversion formula from posterior min entropy to probability of guessing). In fact, it is easy to define an input probability distribution such that the leakage gap between the non-optimal solution and the optimal solution given by the greedy algorithm is arbitrarily large.

Figure 3. Posterior Shannon and min-entropy for different anonymity solutions. k is $\sqrt{n}$, where n is the size of the input set. For the random solution, the values in the plot are the average of 1000 samples.

Note that, by Theorem 6, the greedy solution is in fact optimal for *all* channel-supermodular entropies. Hence, the user knows that the greedy solution is optimal against an attacker trying to guess his secret query in a fixed number of guesses, or using guesswork or guessing using a twenty-questions-style guesswork (reflecting Shannon entropy), and so on.

6.3. Query Anonymity for Related Secrets

Consider the scenario where a user who wants to query the Twitter page of some political commentators is at the same time interested in hiding his own political affiliation, which could be leaked by his queries. In this scenario, the solution from Algorithm 1 might be sub-optimal. To see this, suppose that the k queries in the real query's cover end up being all affiliated to the same party. This would thus reveal the user's political party to the attacker with certainty even though the real intended query is still uncertain. This is not a contradiction to the optimality of the algorithm. As such, an adversary would be better modeled by a partition entropy—with political commentators (the queried users) grouped by party affiliation—and, as established in Proposition 3, this type of entropy is not channel-supermodular in general.

Motivated by this scenario, we now give an optimal solution for all channel supermodular entropies (even if not symmetric) to deal with this kind of problem. Suppose there are k political parties, and l commentators aligned with each party. Suppose, further, that the user is affiliated to one of the parties, and would like to check the profiles of his party's commentators without revealing his own affiliation.

To achieve this aim, the user decides to group the political commentators in covers of size k , each cover containing exactly one commentator from each party, and then proceed to use these covers similarly to the mechanism described in Section 6.2, by querying the entire cover (fetching the pages of all the commentators in the cover). The question is: what is the set of covers that reveal the least amount of information about the user's political affiliation?

Let $\mathcal{X} = \{P_1, \dots, P_k\}$ be the set of parties, $\mathcal{Y} = \{c_1^1, \dots, c_1^l, c_2^1, \dots, c_2^l, \dots, c_k^1, \dots, c_k^l\}$ be the set of commentators, wherein c_i^j is the j th commentator of the i th party, and let $\mathcal{Z} \subset 2^{\mathcal{Y}}$ be the set of covers. Let $K : \mathcal{X} \rightarrow \mathcal{Y}$ be the channel giving the conditional probability of a user choosing to query a commentator given the user's party inclination. We assume that the user only chooses commentators that share the same affiliation as his, that is, $K(c_j^i | P_m) = 0$ whenever $j \neq m$. For simplicity, we assume that the commentators are organized decreasingly with regard to this probability—that is,

$$K(c_j^i | P_j) \geq K(c_j^m | P_j) \text{ whenever } m > i. \quad (6)$$

We claim that the optimal mechanism, for all channel-supermodular entropies (with regards to the political parties, not the commentators), is to group the most popular commentators for each party in the first cover, then the second-most popular commentators for each party on the second, and so on. That is, the covers would be: $\{c_1^1, c_2^1, \dots, c_k^1\}$, $\{c_1^2, c_2^2, \dots, c_k^2\}$, $\dots$, $\{c_1^l, c_2^l, \dots, c_k^l\}$.

Let $R : \mathcal{Y} \rightarrow \mathcal{Z}$ be the channel mapping each commentator to their cover, modelling the optimal solution described above. The matrix of this channel can be seen as a vertical concatenation of identity matrices:

R	$Cover_1$	$Cover_2$	$\dots$	$Cover_l$
c_1^1	1	0	$\dots$	0
c_1^2	0	1	$\dots$	0
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
c_1^l	0	0	$\dots$	1
$\vdots$		$\vdots$		
$\vdots$		$\vdots$		
c_k^1	1	0	$\dots$	0
c_k^2	0	1	$\dots$	0
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
c_k^l	0	0	$\dots$	1

KR	$Cover_1$	$Cover_2$	$\dots$	$Cover_l$
P_1	$K(c_1^1 P_1)$	$K(c_1^2 P_1)$	$\dots$	$K(c_1^l P_1)$
P_2	$K(c_2^1 P_2)$	$K(c_2^2 P_2)$	$\dots$	$K(c_2^l P_2)$
$\vdots$	$\vdots$	$\vdots$		$\vdots$
P_k	$K(c_k^1 P_k)$	$K(c_k^2 P_k)$	$\dots$	$K(c_k^l P_k)$

(7)

The channel $KR : \mathcal{X} \rightarrow \mathcal{Z}$ above is then obtained by postprocessing K by R .

The claim of optimality can be more formally stated as follows: given any channel-supermodular H , any distribution p_X over the parties, and any other deterministic covering $R' : \mathcal{Y} \rightarrow \mathcal{Z}'$ in which there is exactly one commentator from each party per cover, the resulting channel KR' will never leak less information than the channel KR in (7). That is,

$$H(X|Z) \geq H(X|Z').$$

The proof that the covering R above is indeed optimal for all channel-supermodular entropies is similar to the proof of Theorem 6, but even simpler. Suppose the channel $R' : \mathcal{Y} \rightarrow \mathcal{Z}$ is any covering that satisfies the restriction that each cover has exactly one commentator from each party. Now, consider the channel KR' , and proceed as follows: first, do the Join-Meet operation of the first column with all the other columns (that is, obtaining the channel $\diamond_{1,l} \diamond_{1,l-1} \dots \diamond_{1,2}(KR')$). Now, disregarding the first column, the process is repeated for the second column with all the remaining ones. It is easy to see that the resulting channel will be exactly KR in (7).

As an example, let $k = 3$, $l = 3$, and suppose the non-zero values of K are:

$$\begin{aligned} K(c_1^1|P_1) &= 0.5, \quad K(c_1^2|P_1) = 0.3, \quad K(c_1^3|P_1) = 0.2 \\ K(c_2^1|P_2) &= 0.6, \quad K(c_2^2|P_2) = 0.3, \quad K(c_2^3|P_2) = 0.1 \\ K(c_3^1|P_3) &= 0.5, \quad K(c_3^2|P_3) = 0.4, \quad K(c_3^3|P_3) = 0.1 \end{aligned}$$

and suppose that the covers that are defined by R' are $\{c_1^2, c_1^2, c_3^2\}$, $\{c_1^1, c_2^3, c_3^1\}$, and $\{c_1^3, c_2^2, c_3^3\}$. We have

KR'	$Cover_1$	$Cover_2$	$Cover_3$
P_1	0.3	0.5	0.2
P_2	0.6	0.1	0.3
P_3	0.4	0.5	0.1

By doing the Join-Meet operations on the first column with the others, we obtain

$\diamond_{1,3}\diamond_{1,2}(KR')$	$Cover_1$	$Cover_2$	$Cover_3$
P_1	0.5	0.3	0.2
P_2	0.6	0.1	0.3
P_3	0.5	0.4	0.1

Finally, by doing the Join-Meet of the second column with the remaining one, we have

$\diamond_{2,3}\diamond_{1,3}\diamond_{1,2}(KR')$	$Cover_1$	$Cover_2$	$Cover_3$
P_1	0.5	0.3	0.2
P_2	0.6	0.3	0.1
P_3	0.5	0.4	0.1

which is exactly the optimal solution given by (7).

7. Conclusions

In this work, we introduced the notion of channel-supermodular entropies, as a subset of core-concave entropies, which include guessing and Arimoto–Rényi entropies. We demonstrated that, for this new classification of entropies, the Join-Meet operator on channel columns decreases the H -mutual information. This property prompted us to define structural preorders of channels ($\geq_{ds}$, $\geq_{sds}$), providing novel sufficient conditions for establishing whether two channels are in the H -more capable any channel-supermodular H or in the (H_1, H_∞) -capacity ordering. Moreover, this work establishes some relationships of these new structural preorders with other existing preorders from the literature.

As an example application, we used channel-supermodularity to prove an optimality result of a greedy query anonymization algorithm.

It is our belief that the connection between supermodular functions and some commonly used entropy measures, made in Section 3.1, will prove useful for posterior investigations in information theory (for example, given the vast literature of supermodular functions over Euclidean space [50] (Chapter 6.D) and [49]). Further directions of work include investigating other useful properties of channel-supermodular entropies, and further applications of channel-supermodularity to anonymity.

Author Contributions: All authors contributed to the results of this paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding

Data Availability Statement: Not applicable.

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

Appendix A. Proofs of Section 3

Appendix A.1. Proof of Proposition 2

(1) For Shannon entropy, Definition 5 yields

$$G_F(\mathbf{r}) = -\|\mathbf{r}\|_1 \sum_i^n \frac{r_i}{\|\mathbf{r}\|_1} \log\left(\frac{r_i}{\|\mathbf{r}\|_1}\right) = -\sum_i^n r_i \log\left(\frac{r_i}{\|\mathbf{r}\|_1}\right).$$

Suppose $\mathbf{r} \in \mathbb{R}_{>0}^n$. Then, for all i, j with $i \neq j$ we have

$$\frac{\partial^2 G_F(\mathbf{r})}{\partial r_i \partial r_j} = \frac{1}{\|\mathbf{r}\|_1 \ln(2)} \geq 0. \quad (\text{A1})$$

Supermodularity of G_F restricted to $\mathbf{r} \in \mathbb{R}_{>0}^n$ then follows from (4). If, however, $\mathbf{r} \in \mathbb{R}_{\geq 0}^n \setminus \mathbb{R}_{>0}^n$, the second derivative may not exist. In such a case, supermodularity can be established by a limiting argument. Let $\epsilon > 0$ and $\mathbf{r}' = \mathbf{r} + \epsilon(\sum_k \mathbf{e}_k)$.

Then (4) and (A1) imply that

$$G_F(\mathbf{r}' + \delta_1 \mathbf{e}_i + \delta_2 \mathbf{e}_j) + G_F(\mathbf{r}') \geq G_F(\mathbf{r}' + \delta_1 \mathbf{e}_i) + G_F(\mathbf{r}' + \delta_2 \mathbf{e}_j).$$

As G_F is continuous, taking $\epsilon \rightarrow 0$ in both sides of the inequality above, we obtain

$$G_F(\mathbf{r} + \delta_1 \mathbf{e}_i + \delta_2 \mathbf{e}_j) + G_F(\mathbf{r}) \geq G_F(\mathbf{r} + \delta_1 \mathbf{e}_i) + G_F(\mathbf{r} + \delta_2 \mathbf{e}_j).$$

Thus, G_F is supermodular following (3).

(2) For Arimoto–Rényi entropies of order $\alpha \in (0, 1)$, we have

$$G_F(\mathbf{r}) = \|\mathbf{r}\|_1 F\left(\frac{\mathbf{r}}{\|\mathbf{r}\|_1}\right) = \|\mathbf{r}\|_1 \left\| \frac{\mathbf{r}}{\|\mathbf{r}\|_1} \right\|_\alpha = \|\mathbf{r}\|_\alpha$$

and similarly, $G_F(\mathbf{r}) = -\|\mathbf{r}\|_\alpha$ for entropies of order $\alpha > 1$.

Now, for all $\mathbf{r} \in \mathbb{R}_{>0}^n$ and i, j with $i \neq j$, we have

$$\frac{\partial^2 \|\mathbf{r}\|_\alpha}{\partial r_i \partial r_j} = (1 - \alpha) r_i^{\alpha-1} r_j^{\alpha-1} \left(\sum_{k=1}^n r_k^\alpha \right)^{\frac{1}{\alpha}-2}$$

which is negative if $\alpha > 1$ and positive if $0 < \alpha < 1$. Thus, from Equation (4), G_F restricted to $\mathbb{R}_{>0}^n$ is supermodular for all $\alpha \in (0, 1) \cup (1, \infty)$. Supermodularity of G_F can then be established by a limiting argument similar to the one in the proof for Shannon entropy.

(3) For the k -tries entropy, we have

$$G_F(\mathbf{r}) = -\|\mathbf{r}\|_1 \left(\sum_{i=1}^k \frac{r_{[i]}}{\|\mathbf{r}\|_1} \right) = -\sum_{i=1}^k r_{[i]}.$$

Let $\mathbf{r} \in \mathbb{R}_{\geq 0}^n$, let i, j be such that $i \neq j$ and let $\delta_1, \delta_2 > 0$. From Equation (3), it suffices to prove that

$$G_F(\mathbf{r} + \delta_1 \mathbf{e}_i) - G_F(\mathbf{r}) \leq G_F(\mathbf{r} + \delta_1 \mathbf{e}_i + \delta_2 \mathbf{e}_j) - G_F(\mathbf{r} + \delta_2 \mathbf{e}_j).$$

Let $\mathbf{r}' = \mathbf{r} + \delta_1 \mathbf{e}_i$, $\mathbf{r}'' = \mathbf{r} + \delta_2 \mathbf{e}_j$ and $\mathbf{r}''' = \mathbf{r} + \delta_1 \mathbf{e}_i + \delta_2 \mathbf{e}_j$. The inequality above is equivalent to

$$\sum_{l=1}^k r'_{[l]} - \sum_{l=1}^k r_{[l]} \geq \sum_{l=1}^k r'''_{[l]} - \sum_{l=1}^k r''_{[l]}. \quad (\text{A2})$$

If $r_i + \delta_1$ is not amongst the k largest elements in $\mathbf{r}'''$, then the right-hand side of (A2) is equal to 0, and there is nothing to prove. If $r_i + \delta_1$ is amongst the k largest elements in $\mathbf{r}'''$, then it is also amongst the k largest elements in $\mathbf{r}'$, and either:

- r_i is amongst the k largest elements in $\mathbf{r}$ and $\mathbf{r}''$, and thus both sides of (A2) are equal to δ_1 ;
- r_i is amongst the k largest elements in $\mathbf{r}$ but not in $\mathbf{r}''$. In which case the left-hand side of (A2) is equal to δ_1 , and the right-hand side of (A2) is less than δ_1 ;
- r_i not amongst the k largest elements in $\mathbf{r}$ and neither in $\mathbf{r}''$. In this case the left-hand side of (A2) is larger than the right-hand side, as the k th largest element of $\mathbf{r}''$ is greater or equal than the k th largest element of $\mathbf{r}$.

In any case, (A2) holds.

(4) For guessing entropy, Definition 5 yields

$$G_F(\mathbf{r}) = \|\mathbf{r}\|_1 \sum_i \frac{ir_{[i]}}{\|\mathbf{r}\|_1} = \sum_i ir_{[i]}.$$

Let i, j be such that $i \neq j$ and let $\delta_1, \delta_2 > 0$. Suppose, without loss of generality, that $r_i + \delta_1 \geq r_j + \delta_2$. We will prove that

$$G_F(\mathbf{r} + \delta_1 \mathbf{e}_i + \delta_2 \mathbf{e}_j) - G_F(\mathbf{r} + \delta_1 \mathbf{e}_i) \geq G_F(\mathbf{r} + \delta_2 \mathbf{e}_j) - G_F(\mathbf{r}). \quad (\text{A3})$$

Let $\mathcal{I}_{\mathbf{r}+\delta_1 \mathbf{e}_i} = \{k \leq n \mid r_j < r_k < r_j + \delta_2\}$ —that is, $\mathcal{I}_{\mathbf{r}+\delta_1 \mathbf{e}_i}$ is the set of coordinates of $\mathbf{r} + \delta_1 \mathbf{e}_i$ whose value is between r_j and $r_j + \delta_2$.

Let $m - 1$ be the number of entries of $\mathbf{r} + \delta_1 \mathbf{e}_i + \delta_2 \mathbf{e}_j$ that are greater than or equal to $r_j + \delta_2$. Then, there are $(m + |\mathcal{I}_{\mathbf{r}+\delta_1 \mathbf{e}_i}| - 1)$ coordinates of $\mathbf{r} + \delta_1 \mathbf{e}_i$ that are strictly greater than r_j .

The difference $G_F(\mathbf{r} + \delta_1 \mathbf{e}_i + \delta_2 \mathbf{e}_j) - G_F(\mathbf{r} + \delta_1 \mathbf{e}_i)$ is then $m(r_j + \delta_2) - (m + |\mathcal{I}_{\mathbf{r}+\delta_1 \mathbf{e}_i}|)r_j$, plus the sum of the entries given by the set $\mathcal{I}_{\mathbf{r}+\delta_1 \mathbf{e}_i}$, as the integers by which they are multiplied are decreased by one. As we assume $r_i + \delta_1 \geq r_j + \delta_2$, $i \notin \mathcal{I}_{\mathbf{r}+\delta_1 \mathbf{e}_i}$, and the value of the right hand-side of (A3) reduces to

$$\begin{aligned} & G_F(\mathbf{r} + \delta_1 \mathbf{e}_i + \delta_2 \mathbf{e}_j) - G_F(\mathbf{r} + \delta_1 \mathbf{e}_i) \\ &= m(r_j + \delta_2) - (m + |\mathcal{I}_{\mathbf{r}+\delta_1 \mathbf{e}_i}|)r_j + \sum_{k \in \mathcal{I}_{\mathbf{r}+\delta_1 \mathbf{e}_i}} r_k \\ &= -|\mathcal{I}_{\mathbf{r}+\delta_1 \mathbf{e}_i}|r_j + m\delta_2 + \sum_{k \in \mathcal{I}_{\mathbf{r}+\delta_1 \mathbf{e}_i}} r_k. \end{aligned}$$

We now turn to the right-hand side of (A3), the difference $G_F(\mathbf{p} + \delta_2 \mathbf{e}_j) - G_F(\mathbf{p})$. We define $\mathcal{I}_{\mathbf{p}}$ similarly to $\mathcal{I}_{\mathbf{r}+\delta_1 \mathbf{e}_i}$ and divide the proof in three cases.

Case 1 ($r_i \geq r_j + \delta_2$): In this case, there are $m - 1$ entries of $\mathbf{r} + \delta_2 \mathbf{e}_j$ greater or equal to $r_j + \delta_2$, and $(m + |\mathcal{I}_{\mathbf{r}}| - 1)$ entries of $\mathbf{r}$ strictly greater than r_j . Moreover, $\mathcal{I}_{\mathbf{r}} = \mathcal{I}_{\mathbf{r}+\delta_1 \mathbf{e}_i}$, and we obtain

$$\begin{aligned} & G_F(\mathbf{r} + \delta_2 \mathbf{e}_j) - G_F(\mathbf{r}) \\ &= m(r_j + \delta_2) - (m + |\mathcal{I}_{\mathbf{r}}|)r_j + \sum_{k \in \mathcal{I}_{\mathbf{r}}} r_k \\ &= G_F(\mathbf{r} + \delta_1 \mathbf{e}_i + \delta_2 \mathbf{e}_j) - G_F(\mathbf{r} + \delta_1 \mathbf{e}_i). \end{aligned}$$

Case 2 ($r_j < r_i < r_j + \delta_2$): In this case, there are $m - 1$ entries of $\mathbf{r} + \delta_2 \mathbf{e}_j$ greater or equal to $r_j + \delta_2$, and $(m + |\mathcal{I}_{\mathbf{r}}| - 2)$ entries of $\mathbf{r}$ strictly greater than r_j . This time, $\mathcal{I}_{\mathbf{r}} = \mathcal{I}_{\mathbf{r}+\delta_1 \mathbf{e}_i} \cup \{i\}$, and we have

$$\begin{aligned} & G_F(\mathbf{r} + \delta_2 \mathbf{e}_j) - G_F(\mathbf{r}) \\ &= (m - 1)(r_j + \delta_2) + \sum_{k \in \mathcal{I}_{\mathbf{r}}} r_k - (m - 1 + |\mathcal{I}_{\mathbf{r}}|)r_j \\ &= (m - 1)(r_j + \delta_2) + \sum_{k \in \mathcal{I}_{\mathbf{r}+\delta_1 \mathbf{e}_i}} r_k + r_i - (m + |\mathcal{I}_{\mathbf{r}+\delta_1 \mathbf{e}_i}|)r_j \\ &= -|\mathcal{I}_{\mathbf{r}+\delta_1 \mathbf{e}_i}|r_j + m\delta_2 + \sum_{k \in \mathcal{I}_{\mathbf{r}+\delta_1 \mathbf{e}_i}} r_k + r_i - (r_j + \delta_2) \\ &\leq -|\mathcal{I}_{\mathbf{r}+\delta_1 \mathbf{e}_i}|r_j + m\delta_2 + \sum_{k \in \mathcal{I}_{\mathbf{r}+\delta_1 \mathbf{e}_i}} r_k \\ &= G_F(\mathbf{r} + \delta_1 \mathbf{e}_i + \delta_2 \mathbf{e}_j) - G_F(\mathbf{r} + \delta_1 \mathbf{e}_i), \end{aligned}$$

where the inequality comes from the assumption that $r_i < r_j + \delta_2$.

Case 3 ($r_i \leq r_j$): In this case, again there are $m - 1$ entries of $\mathbf{r} + \delta_2 \mathbf{e}_j$ greater or equal to $r_j + \delta_2$, and $(m + |\mathcal{I}_{\mathbf{r}}| - 2)$ entries of $\mathbf{r}$ strictly greater than r_j . This time, $\mathcal{I}_{\mathbf{r}} = \mathcal{I}_{\mathbf{r} + \delta_1 \mathbf{e}_i}$, and we have

$$\begin{aligned} & G_F(\mathbf{r} + \delta_2 \mathbf{e}_j) - G_F(\mathbf{r}) \\ &= (m - 1)(r_j + \delta_2) + \sum_{k \in \mathcal{I}_{\mathbf{r}}} r_k - (m - 1 + |\mathcal{I}_{\mathbf{r}}|)r_j \\ &= -|\mathcal{I}_{\mathbf{r}}|r_j + (m - 1)\delta_2 + \sum_{k \in \mathcal{I}_{\mathbf{r}}} r_k \\ &\leq -|\mathcal{I}_{\mathbf{r}}|r_j + m\delta_2 + \sum_{k \in \mathcal{I}_{\mathbf{r}}} r_k \\ &= G_F(\mathbf{r} + \delta_1 \mathbf{e}_i + \delta_2 \mathbf{e}_j) - G_F(\mathbf{r} + \delta_1 \mathbf{e}_i). \end{aligned}$$

Therefore, G_F is supermodular.

Appendix A.2. Proof of Proposition 3

(1) For all these entropies, Definition 5 yields

$$G_F(\mathbf{r}) = \|\mathbf{r}\|_1 \left\| \frac{\mathbf{r}}{\|\mathbf{r}\|_1} \right\|_{\alpha}^{\alpha} = \|\mathbf{r}\|_{\alpha}^{\alpha} \|\mathbf{r}\|_1^{(1-\alpha)},$$

if $0 < \alpha < 1$, and $G_F(\mathbf{r}) = -\|\mathbf{r}\|_{\alpha}^{\alpha} \|\mathbf{r}\|_1^{(1-\alpha)}$ if $\alpha > 1$.

Now, if $\mathbf{r} \in \mathbb{R}_{>0}^n$, we have

$$\frac{\partial^2 \|\mathbf{r}\|_{\alpha}^{\alpha} \|\mathbf{r}\|_1^{(1-\alpha)}}{\partial r_i \partial r_j} = \alpha(1 - \alpha) \|\mathbf{r}\|_1^{-\alpha} (r_i^{\alpha-1} + r_j^{\alpha-1} - \|\mathbf{r}\|_{\alpha}^{\alpha} \|\mathbf{r}\|_1^{-1}),$$

and by (4), G_F is supermodular only if the partial derivative above is positive if $\alpha \in (0, 1)$ and negative if $\alpha > 1$. Equivalently, if G_F is supermodular then for all $\mathbf{r} \in \mathbb{R}_{>0}^n$,

$$\forall i, j \text{ such that } i \neq j \quad r_i^{\alpha-1} + r_j^{\alpha-1} - \|\mathbf{r}\|_{\alpha}^{\alpha} \|\mathbf{r}\|_1^{-1} \geq 0. \quad (\text{A4})$$

Thus, it suffices to provide, for each α , one vector $\mathbf{r}$ such that (A4) does not hold.

Case 1 ($\alpha \in (0, 1)$): Let $\mathbf{r} \in \mathbb{R}_{>0}^n$ be the vector with coordinates $r_1 = r_2 = 1$ and $r_j = \epsilon$ for all $j > 2$. Then,

$$r_1^{\alpha-1} + r_2^{\alpha-1} - \|\mathbf{r}\|_{\alpha}^{\alpha} \|\mathbf{r}\|_1^{-1} = 2 - \frac{2 + (n-2)\epsilon^{\alpha}}{2 + (n-2)\epsilon}.$$

Now, $\lim_{n \rightarrow \infty} \frac{2 + (n-2)\epsilon^{\alpha}}{2 + (n-2)\epsilon} = \epsilon^{\alpha-1}$, and $\epsilon^{\alpha-1}$ can be made arbitrarily large by choosing a suitably small ϵ . Thus, for an appropriate choice of n and ϵ , $\frac{2 + (n-2)\epsilon^{\alpha}}{2 + (n-2)\epsilon} > 2$, and (A4) does not hold.

Case 2 ($\alpha > 1$): Fix $n > 2$. Let $\alpha > 1$, $\epsilon > 0$ and pick $a > 0$ such that

$$a^{\alpha} - 2a - 2 > (n-3)(2\epsilon - \epsilon^{\alpha}).$$

Notice that such a is guaranteed to exist because $\alpha > 1$. Let $\mathbf{r} \in \mathbb{R}_{>0}^n$ be the vector with $r_1 = r_2 = 1$, $r_3 = a$ and $r_k = \epsilon$ for all $k > 3$. Then

$$r_1^{\alpha-1} + r_2^{\alpha-1} - \|\mathbf{r}\|_{\alpha}^{\alpha} \|\mathbf{r}\|_1^{-1} = 2 - \frac{2 + a^{\alpha} + (n-3)\epsilon^{\alpha}}{2 + a + (n-3)\epsilon} = \frac{2 + 2a - a^{\alpha} + (n-3)(2\epsilon - \epsilon^{\alpha})}{2 + a + (n-3)\epsilon} < 0.$$

and (A4) does not hold. Therefore, G_F is not supermodular.

(2) By Definition 5, we have

$$G_F(\mathbf{r}) = -\max_{\mathcal{A} \in \mathcal{P}} \sum_{i \in \mathcal{A}} r_i.$$

Let $n = 3$, $\mathcal{P} = \{\{1, 2\}, \{3\}\}$, and let $\mathbf{r} = (3, 1, 3)$, $\mathbf{s} = (1, 3, 3) \in \mathbb{R}_{\geq 0}^3$. Then,

$$G_F(\mathbf{r} \vee \mathbf{s}) + G_F(\mathbf{r} \wedge \mathbf{s}) = -9 < -8 = G_F(\mathbf{r}) + G_F(\mathbf{s}).$$

Appendix B. Proofs of Section 5

Appendix B.1. Proof of Proposition 5

(1) follows immediately from Definition 9, and (2) from Proposition 4.2. Statement (3) is proven by the following channels:

$$\begin{pmatrix} 1 & \\ & 1 \end{pmatrix} \geq_d \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}.$$

For (4), consider the channels K_1, K_2 as follows:

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1/2 & 1/2 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 \\ 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}.$$

These channels were introduced in [3], where it is proved that $K_1 \not\geq_{\ln}^{H_1} K_2$. However $K_1 \geq_{\text{ds}} K_2$, as

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1/2 & 1/2 \end{pmatrix} \geq_d \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 1/2 \\ 1/2 & 1/4 & 1/4 \end{pmatrix} \geq_s \begin{pmatrix} 1 & 0 & 0 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/4 & 1/4 \end{pmatrix} \geq_d \begin{pmatrix} 1 & 0 \\ 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}.$$

Statement (5) follows from (1) and from Theorem 2.

To prove (6), consider the following channels, for some $r \in (0, 1)$ and $\delta, \epsilon > 0$:

$$K_1 = \begin{pmatrix} r & 1-r \\ r+\delta & 1-r-\delta \end{pmatrix} \quad K_2 = K_1 \begin{pmatrix} 1+\epsilon & -\epsilon \\ 1/2 & 1/2 \end{pmatrix}.$$

Notice that K_1 and K_2 are indeed channels if δ and ϵ are small enough. These channels were introduced in [3] as an example for which (for δ, ϵ small enough) $K_1 \geq_{\ln}^{H_1} K_2$ but $K_1 \not\geq_d K_2$. As their input set is of size 2, Proposition 4.1 implies that $K_1 \not\geq_{\text{ds}} K_2$.

Statement (7) follows directly from Definition 9, Proposition 1, and Theorem 4.

For (8), consider the following K_1, K_2 :

$$\begin{pmatrix} 0 & 1/25 & 24/25 \\ 1/2 & 0 & 1/2 \\ 1/3 & 1/3 & 1/3 \end{pmatrix} \quad \begin{pmatrix} 1/25 & 0 & 24/25 \\ 1/2 & 0 & 1/2 \\ 1/3 & 1/3 & 1/3 \end{pmatrix}.$$

In this case, $K_2 = \diamond_{1,2} K_1$, and thus $K_1 \geq_{\text{ds}} K_2$. Suppose, to derive a contradiction, that $K_1 \geq_{\text{sh}} K_2$. Then there is a finite collection $\{(g_i, T_i, R_i)\}$ such that each T_i, R_i is a deterministic channel, each $g_i > 0$, $\sum g_i = 1$ and $K_2 = \sum_i g_i T_i K_1 R_i$. Notice that, for all i , we must have

$$(T_i K_1 R_i)(y_2|x_1) = (T_i K_1 R_i)(y_2|x_2) = 0. \quad (\text{A5})$$

Let $\mathcal{I} = \{i \mid (T_i K_1 R_i)(y_1|x_1) \leq 1/25\}$. We claim that

$$\text{for all } i \in \mathcal{I}, \quad (T_i K_1 R_i)(y_1|x_3) + (T_i K_1 R_i)(y_3|x_2) + (T_i K_1 R_i)(y_3|x_3) \geq \frac{4}{3}. \quad (\text{A6})$$

To see why, notice that channels of the form $T_i K_1$ are those channels whose rows are equal to rows of K_1 (with possible reordering and duplicates), and channels of the form $(T_i K_1) R_i$ are those obtaining by permutating or summing columns of $(T_i K_1)$.

If the first row of $T_i K_1$ is either $(1/2, 0, 1/2)$ or $(1/3, 1/3, 1/3)$. Then, in order that $i \in \mathcal{I}$, the third column of $T_i K_1 R_i$ must be either 1) equal to the sum of the first and third columns of $T_i K_1$, 2) or equal to the sum of all columns of $T_i K_1$. In either case, we have $(T_i K_1 R_i)(y_3|x_2) \geq 2/3$ and $(T_i K_1 R_i)(y_3|x_3) \geq 2/3$, so (A6) holds.

If the first row of $T_i K_1$ is $(0, 1/25, 24/25)$, the proof is not as straightforward, and we divide it into two subcases. The first subcase is when the second row is either $(1/3, 1/3, 1/3)$ or $(1/2, 0, 1/2)$, then each column of $T_i K_1$ will be mapped to the first or to the third column of $T_i K_1 R_i$, and therefore $(T_i K_1 R_i)(y_1|x_3) + (T_i K_1 R_i)(y_3|x_3) = 1$. Then, (A6) follows from observing that $(T_i K_1 R_i)(y_3|x_2) \geq 1/3$. The second subcase to consider is when the second row is $(0, 1/25, 24/25)$. In this case, we have $(T_i K_1 R_i)(y_3|x_2) \geq 24/25$, and, by considering each of the three possible choices for the third row of $T_i K_1$, it is easy to see that $(T_i K_1 R_i)(y_1|x_3) + (T_i K_1 R_i)(y_3|x_3) \geq 1/2$.

As $K_2(y_1|x_3) + K_2(y_3|x_2) + K_2(y_3|x_3) = 7/6$, it follows from (A6) that

$$\begin{aligned} \frac{7}{6} &= \sum_i g_i \left((T_i K_1 R_i)(y_1|x_3) + (T_i K_1 R_i)(y_3|x_2) + (T_i K_1 R_i)(y_3|x_3) \right) \\ &\geq \sum_{i \in \mathcal{I}} g_i \left((T_i K_1 R_i)(y_1|x_3) + (T_i K_1 R_i)(y_3|x_2) + (T_i K_1 R_i)(y_3|x_3) \right) \\ &\geq \frac{4}{3} \sum_{i \in \mathcal{I}} g_i. \end{aligned}$$

Thus, it follows that $\sum_{i \in \mathcal{I}} g_i \leq 7/8$, and therefore $\sum_{i \notin \mathcal{I}} g_i \geq 1/8$. Since $\min_{i \notin \mathcal{I}} (T_i K_1 R_i)(y_1|x_1) \geq 1/3$, we have

$$\begin{aligned} \frac{1}{25} &= K_2(y_1|x_1) = \sum_{i=1}^k g_i (T_i K_1 R_i)(y_1|x_1) \\ &\geq \sum_{i \notin \mathcal{I}} g_i (T_i K_1 R_i)(y_1|x_1) \\ &\geq \frac{1}{3} \sum_{i \notin \mathcal{I}} g_i \\ &\geq \frac{1}{3} \times \frac{1}{8} = \frac{1}{24} \end{aligned}$$

which is absurd.

For (9), let K_1, K_2 be the following channels:

$$\begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Notice that K_2 is obtained from K_1 by row permutation; hence, $K_1 \geq_{\text{sh}} K_2$. However, $K_1 \not\geq_{\text{mc}}^{H_1} K_2$ (take, for example, $p_X = (1/2, 1/2, 0)$). Thus, (7) yields $K_1 \not\geq_{\text{ds}} K_2$.

Appendix B.2. Proof of Proposition 9

(1) Follows from Lemma 1 and Shannon's result in [1].

(2) In light of Lemma 1, we need only to prove that $K_1 \geq_{\text{sh}} K_2 \Rightarrow K_1 \geq_{\text{c}}^{H_\infty} K_2$.

Firstly, we prove that, given channels W_1 , W_2 and $W_3 = tW_1 + (1-t)W_2$, we have $C_{H_\infty}(W_3) \leq \max(C_{H_\infty}(W_1), C_{H_\infty}(W_2))$. Keeping in mind the identity $C_{H_\infty}(W_3) = \log \sum_y \max_x W_3(x, y)$ ([43] Remark 3),

$$\begin{aligned} C_{H_\infty}(W_3) &= \log \left(\sum_y \max_x (tW_1(x, y) + (1-t)W_2(x, y)) \right) \\ &\leq \log \left(t \sum_y \max_x W_1(x, y) + (1-t) \sum_y \max_x W_2(x, y) \right) \\ &\leq \max(C_{H_\infty}(W_1), C_{H_\infty}(W_2)) \end{aligned}$$

where the last inequality follows from quasi-convexity of $\log$. Now, if $K_2 = (\sum_i g_i R_i K_1 T_i)$, then $C_{H_\infty}(K_2) = C_{H_\infty}(\sum_i g_i R_i K_1 T_i) \leq \max_i C_{H_\infty}(R_i K_1 T_i) \leq C_{H_\infty}(K_1)$, where the last inequality follows from ([33] Theorem 6).

(3) This is readily seen from the fact that $\geq_c^{H_1}$ and $\geq_c^{H_\infty}$ are total preorders that do not coincide. To make the argument concrete, consider channels K_1, K_2 as follows:

$$\begin{pmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \end{pmatrix} \quad \begin{pmatrix} 2/5 & 2/5 & 1/5 \\ 1/5 & 2/5 & 2/5 \\ 2/5 & 1/5 & 2/5 \end{pmatrix}.$$

Then, from ([53] Theorem 2.7.1) $C_{H_\infty}(K_1) > C_{H_1}(K_2)$, whereas, from ([43] Remark 3), $C_{H_\infty}(K_1) < C_{H_\infty}(K_2)$. Thus, (1) and (2) imply that $K_1 \not\geq_{\text{shs}} K_2$ and $K_2 \not\geq_{\text{shs}} K_1$.

References

- Shannon, C.E. A note on a partial ordering for communication channels. *Inf. Control* **1958**, *1*, 390–397. [\[CrossRef\]](#)
- El Gamal, A.A. The capacity of a class of broadcast channels. *IEEE Trans. Inf. Theory* **1979**, *25*, 166–169. [\[CrossRef\]](#)
- Korner, J.; Marton, K. Comparison of two noisy channels. *Coll. Math. Soc. J. Bolyai* **1977**, *16*, 411–423.
- Clark, D.; Hunt, S.; Malacaria, P. Quantitative Information Flow, Relations and Polymorphic Types. *J. Log. Comput.* **2005**, *15*, 181–199. [\[CrossRef\]](#)
- Smith, G. On the Foundations of Quantitative Information Flow. In Proceedings of the 12th International Conference on Foundations of Software Science and Computational Structures (FOSSACS), York, UK, 22–29 March 2009; Springer: Berlin/Heidelberg, Germany, 2009; Volume 5504, pp. 288–302.
- Cohen, J.; Kempermann, J.H.B.; Zbaganu, G. *Comparisons of Stochastic Matrices with Applications in Information Theory, Statistics, Economics and Population*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 1998.
- Américo, A.; Malacaria, P.; Khouzani, A.M. Channel Ordering and Supermodularity. In Proceedings of the 2019 IEEE Information Theory Workshop (ITW) (IEEE ITW 2019), Visby, Sweden, 25–28 August 2019.
- Cover, T.M. Broadcast channels. *IEEE Trans. Inf. Theory* **1972**, *18*, 2–14. [\[CrossRef\]](#)
- Pasareanu, C.S.; Phan, Q.; Malacaria, P. Multi-run Side-Channel Analysis Using Symbolic Execution and Max-SMT. In Proceedings of the IEEE 29th Computer Security Foundations Symposium, CSF 2016, Lisbon, Portugal, 27 June–1 July 2016; pp. 387–400. [\[CrossRef\]](#)
- Heusser, J.; Malacaria, P. Quantifying information leaks in software. In Proceedings of the Twenty-Sixth Annual Computer Security Applications Conference, ACSAC 2010, Austin, TX, USA, 6–10 December 2010; pp. 261–269. [\[CrossRef\]](#)
- Gong, X.; Kiyavash, N. Quantifying the information leakage in timing side channels in deterministic work-conserving schedulers. *IEEE/ACM Trans. Netw.* **2016**, *24*, 1841–1852. [\[CrossRef\]](#)
- Köpf, B.; Smith, G. Vulnerability Bounds and Leakage Resilience of Blinded Cryptography under Timing Attacks. In Proceedings of the 2010 23rd IEEE Computer Security Foundations Symposium, Edinburgh, UK, 17–19 July 2010; pp. 44–56.
- Américo, A.; Khouzani, M.; Malacaria, P. Deterministic Channel Design for Minimum Leakage. In Proceedings of the IEEE 32nd Computer Security Foundations Symposium (CSF), Hoboken, NJ, USA, 25–28 June 2019; pp. 428–441. [\[CrossRef\]](#)
- Bergmans, P. Random coding theorem for broadcast channels with degraded components. *IEEE Trans. Inf. Theory* **1973**, *19*, 197–207. [\[CrossRef\]](#)
- Bergmans, P. A simple converse for broadcast channels with additive white Gaussian noise (Corresp.). *IEEE Trans. Inf. Theory* **1974**, *20*, 279–280. [\[CrossRef\]](#)
- Gallager, R.G. Capacity and coding for degraded broadcast channels. *Probl. Peredachi Informatsii* **1974**, *10*, 3–14.
- Smith, G. Recent Developments in Quantitative Information Flow (Invited Tutorial). In Proceedings of the LICS '15 2015 30th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS), Kyoto, Japan, 6–10 July 2015; IEEE Computer Society: Washington, DC, USA, 2015; pp. 23–31. [\[CrossRef\]](#)

18. Malacaria, P. Algebraic foundations for quantitative information flow. *Math. Struct. Comput. Sci.* **2015**, *25*, 404–428. [[CrossRef](#)]
19. Alvim, M.S.; Chatzikokolakis, K.; Palamidessi, C.; Smith, G. Measuring Information Leakage Using Generalized Gain Functions. In Proceedings of the IEEE 25th Computer Security Foundations Symposium (CSF), Cambridge, MA, USA, 25–27 June 2012; pp. 265–279. [[CrossRef](#)]
20. McIver, A.; Morgan, C.; Smith, G.; Espinoza, B.; Meinicke, L. Abstract Channels and Their Robust Information-Leakage Ordering. In Proceedings of the International Conference on Principles of Security and Trust (POST), Grenoble, France, 5–13 April 2014; Springer: Berlin/Heidelberg, Germany, 2014; Volume 8414, pp. 83–102.
21. Blackwell, D. The comparison of experiments. In *Second Berkeley Symposium on Mathematical Statistics and Probability*; Neyman, J., Ed.; Univ. of California Press: Berkeley, CA, USA, 1951; pp. 93–102.
22. Bordenabe, N.E.; Smith, G. Correlated Secrets in Quantitative Information Flow. In Proceedings of the 2016 IEEE 29th Computer Security Foundations Symposium (CSF), Lisbon, Portugal, 27 June–1 July 2016; pp. 93–104. [[CrossRef](#)]
23. Buscemi, F. Degradable channels, less noisy channels, and quantum statistical morphisms: An equivalence relation. *Probl. Inf. Transm.* **2016**, *52*, 201–213. [[CrossRef](#)]
24. Cam, L.L. Sufficiency and Approximate Sufficiency. *Ann. Math. Stat.* **1964**, *35*, 1419–1455. [[CrossRef](#)]
25. Raginsky, M. Shannon meets Blackwell and Le Cam: Channels, codes, and statistical experiments. In Proceedings of the 2011 IEEE International Symposium on Information Theory Proceedings, St. Petersburg, Russia, 31 July–5 August 2011; pp. 1220–1224. [[CrossRef](#)]
26. Zhang, Y.; Tepedelenlioglu, C. Analytical and Numerical Characterizations of Shannon Ordering for Discrete Memoryless Channels. *IEEE Trans. Inf. Theory* **2014**, *60*, 72–83. [[CrossRef](#)]
27. Nasser, R. Characterizations of Two Channel Orderings: Input-Degradedness and the Shannon Ordering. *IEEE Trans. Inf. Theory* **2018**, *64*, 6759–6770. [[CrossRef](#)]
28. Khouzani, M.; Malacaria, P. Generalised Entropies and Metric-Invariant Optimal Countermeasures for Information Leakage under Symmetric Constraints. *IEEE Trans. Inf. Theory* **2018**, *65*, 888–901. [[CrossRef](#)]
29. Khouzani, M.; Malacaria, P. Leakage-Minimal Design: Universality, Limitations, and Applications. In Proceedings of the IEEE 30th Computer Security Foundations Symposium (CSF), Santa Barbara, CA, USA, 21–25 August 2017; pp. 305–317.
30. Khouzani, M.; Malacaria, P. Optimal Channel Design: A Game Theoretical Analysis. *Entropy* **2018**, *20*, 675. [[CrossRef](#)] [[PubMed](#)]
31. Ah-Fat, P.; Huth, M. Optimal Accuracy-Privacy Trade-Off for Secure Computations. *IEEE Trans. Inf. Theory* **2018**, *65*, 3165–3182. [[CrossRef](#)]
32. Köpf, B.; Basin, D.A. An information-theoretic model for adaptive side-channel attacks. In Proceedings of the 14th ACM Conference on Computer and Communications Security, Alexandria, VA, USA, 2 November–31 October 2007; pp. 286–296. [[CrossRef](#)]
33. Espinoza, B.; Smith, G. Min-entropy as a resource. *Inf. Comp.* **2013**, *226*, 57–75. [[CrossRef](#)]
34. Alvim, M.S.; Chatzikokolakis, K.; McIver, A.; Morgan, C.; Palamidessi, C.; Smith, G. Additive and Multiplicative Notions of Leakage, and Their Capacities. In Proceedings of the IEEE 27th Computer Security Foundations Symposium (CSF), Vienna, Austria, 19–22 July 2014; pp. 308–322. [[CrossRef](#)]
35. Gervais, A.; Shokri, R.; Singla, A.K.; Capkun, S.; Lenders, V. Quantifying Web-Search Privacy. In Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, Scottsdale, AZ, USA, 3–7 November 2014.
36. Beigi, G.; Guo, R.; Nou, A.; Zhang, Y.; Liu, H. Protecting User Privacy: An Approach for Untraceable Web Browsing History and Unambiguous User Profiles. *arXiv* **2018**, arXiv:cs.CR/1811.09340.
37. Rebollo-Monedero, D.; Forne, J. Optimized Query Forgery for Private Information Retrieval. *IEEE Trans. Inf. Theory* **2010**, *56*, 4631–4642. [[CrossRef](#)]
38. Alvim, M.S.; Chatzikokolakis, K.; McIver, A.; Morgan, C.; Palamidessi, C.; Smith, G. An axiomatization of information flow measures. *Theor. Comput. Sci.* **2019**, *777*, 32–54.
39. Arimoto, S. Information-theoretical considerations on estimation problems. *Inf. Control* **1971**, *19*, 181–194. [[CrossRef](#)]
40. Salicru, M.; Menendez, M.; Morales, D.; Pardo, L. Asymptotic distribution of (h, Φ) -entropies. *Commun. Stat.-Theory Methods* **1993**, *22*, 2015–2031. [[CrossRef](#)]
41. Csiszár, I. Information-type measures of difference of probability distributions and indirect observation. *Stud. Sci. Math. Hung.* **1967**, *2*, 229–318.
42. Ho, S.W.; Verdú, S. Convexity/concavity of rényi entropy and α -mutual information. In Proceedings of the 2015 IEEE International Symposium on Information Theory (ISIT), Hong Kong, China, 14–19 June 2015; pp. 745–749. [[CrossRef](#)]
43. Arimoto, S. Information measures and capacity of order α for discrete memoryless channels. In *Topics in Information Theory*; North-Holland Publishing Company: Amsterdam, The Netherlands, 1977.
44. Massey, J.L. Guessing and entropy. In Proceedings of the IEEE Int. Symposium on Information Theory (ISIT), Trondheim, Norway, 27 June–1 July 1994; p. 204. [[CrossRef](#)]
45. Tsallis, C. Possible generalization of Boltzmann-Gibbs statistics. *J. Stat. Phys.* **1988**, *52*, 479–487. [[CrossRef](#)]
46. Sharma, B.; Mittal, D. New non-additive measures of entropy for discrete probability distributions. *J. Math. Sci.* **1975**, *10*, 28–40.
47. Hayashi, M. Exponential decreasing rate of leaked information in universal random privacy amplification. *IEEE Trans. Inf. Theory* **2011**, *57*, 3989–4001. [[CrossRef](#)]
48. Dahl, G. Matrix majorization. *Linear Algebra Its Appl.* **1999**, *288*, 53–73. [[CrossRef](#)]

-
49. Topkis, D.M. *Supermodularity and Complementarity*; Princeton University Press: Princeton, NJ, USA, 1998.
 50. Marshall, A.W.; Olkin, I.; Arnold, B.C. *Inequalities: Theory of Majorization and Its Applications*; Mathematics in Science and Engineering; Academic Press: Cambridge, MA, USA, 1979; Volume 143.
 51. Khouzani, M.H.R.; Malacaria, P. Relative Perfect Secrecy: Universally Optimal Strategies and Channel Design. In Proceedings of the IEEE 29th Computer Security Foundations Symposium, CSF 2016, Lisbon, Portugal, 27 June–1 July 2016; pp. 61–76. [[CrossRef](#)]
 52. Iwamoto, M.; Shikata, J. Information Theoretic Security for Encryption Based on Conditional Rényi Entropies. In *Information Theoretic Security*; Padró, C., Ed.; Springer International Publishing: Cham, Switzerland, 2014; pp. 103–121.
 53. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2006.