Channel Ordering and Supermodularity

Arthur Américo, Pasquale Malacaria, MHR. Khouzani
School of Electronic Engineering and Computer Science
Queen Mary University of London
London, United Kingdom
emails: {a.passosderezende, p.malacaria, arman.khouzani}@qmul.ac.uk

***Abstract*—This work introduces a new preorder over channels that is monotonic with Shannon's mutual information for all distributions over the input alphabet. Moreover, this monotonicity also holds when substituting mutual information for quantities relative to Arimoto-Rényi conditional entropies and guessing entropy. Several results connecting this new preorder with others from the literature are proven. This work also discusses an extension of Shannon ordering based on this new preorder, and establishes that channels ordered this way are also ordered with regards to both Shannon and min-capacity.**

***Index Terms*—Channel ordering, Supermodularity, Security**

I. Introduction

Preorders over noisy channels have been extensively studied in the Information Theory literature starting with Shannon [17] who introduced the inclusion preorder. Significant contributions have been made by [6], [8], [12]. The topic has both theoretical and practical interest: for example establishing whether a channel is *more capable* [12] than another is of practical interest in calculating the capacity region of broadcast channels [8], or in deciding whether a system is more secure than another [5], [18].

Let K_1, K_2 be two discrete memoryless channels (hereafter referred to simply as *channels*) that share an input X over the alphabet $\mathcal{X} = \{x_1, \dots x_n\}$ and produce outputs Y_1, Y_2 over the alphabet $\mathcal{Y} = \{y_1, \dots, y_m\}$. Let $K_i(y|x)$ denote the entries of their probability transition matrices. This work proves that if there are $i, j \leq m$ such that for all $k \leq n$:

$$\begin{aligned} K_2(y_i|x_k) &= \max(K_1(y_i|x_k), K_1(y_j|x_k)), \\ K_2(y_j|x_k) &= \min(K_1(y_i|x_k), K_1(y_j|x_k)), \quad \text{and} \\ K_2(y_l|x_k) &= K_1(y_l|x_k), \quad \forall l \neq i, j, \end{aligned} \tag{1}$$

then, for all distributions p_X over $\mathcal{X}$, we have:

$$I_{H_1}(X;Y_1) \geq I_{H_1}(X;Y_2),$$

where I_{H_1} stands for Shannon mutual information. In other words, (1) implies that K_1 is more capable than K_2. Moreover, this relation holds even if we substitute mutual information with similar quantities relative to other commonly used entropy measures, such as min-entropy, Arimoto-Rényi conditional entropies [3] and guessing entropy [14].

We prove that K_1 and K_2 that respect (1) are not, in general, ordered by either *degradedness* [7] or *inclusion* (also known as *Shannon ordering*) [17]. This allows us to define two channel preorders ($\geq_{\text{ds}}$ and $\geq_{\text{shs}}$) that strictly include the aforementioned ones, yielding novel analytical approaches to establishing whether a channel is more capable or has greater capacity than another.

In Sections II and III, we introduce a generalized entropy measure and some channel preorders generalized from the literature. In Section IV, channel-supermodularity and the JoinMeet operator are introduced, and the claim in the beginning of this section is proved. Section V explores the relationship between $\geq_{\text{ds}}$ and the preorders defined in Section III. Finally, Section VI proves that the preorder $\geq_{\text{shs}}$ can be used to establish whether a channel has smaller capacity than another.

A. Conventions and Notation

We will use $X, Y, Z, \dots$ for random variables, which will respectively take values on finite sets (alphabets) $\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \dots$. The elements of these sets will be endowed with an enumeration, i.e. $\mathcal{X} = \{x_1, \dots, x_n\}$. The notation $K : \mathcal{X} \to \mathcal{Y}$ means that K is a channel with $\mathcal{X}$ and $\mathcal{Y}$ as input and output alphabets. Probability distributions will be denoted by p, with $p(x_i)$ or p_i representing elements of the (categorical) distribution. We may specify the random variable, e.g., write $p_X(x)$, if it is not clear from the context.

We also use the probability distribution name p to refer to the probability vector $(p_1, \dots, p_n) \in \Delta_n$, where Δ_n denotes the $(n-1)$-dimensional probability simplex. Given a function F over Δ_n and a random variable X with distribution p, we may use $F(X)$, $F(p_1, \dots, p_n)$ and $F(p)$ interchangeably. We write $X \to Y \to Z$ if X and Z are conditionally independent from each other given Y, i.e., the process $X \to Y \to Z$ constitutes a discrete Markov chain.

II. Core-concave Entropies

This section introduces a generalized entropy measure, similar to the one in [11], for which our results will be derived. In this work we only compare channels with the same input set $\mathcal{X} = \{x_1, \dots, x_n\}$, thus we define our generalized entropy measure as a function over Δ_n.

Definition 1: A *core-concave entropy* H is a pair (η, F) such that:

1) $F : \Delta_n \to \mathbb{R}$ is a concave and continuous function,
2) $\eta : image(F) \to \mathbb{R}$ is a continuous, strictly increasing function whose domain is the image of F.

Given $p \in \Delta_n$, we define $H(p)$ to mean $\eta(F(p))$, and we denote by $\mathcal{H}$ the set of all core concave entropies.

Definition 2: Given a core-concave entropy $H = (\eta, F)$, its *conditional* or *posterior* counterpart is given by

$$H(X|Y) = \eta\left(\sum_{y\in\mathcal{Y}^+} p_Y(y)F(X|y)\right)$$

where $\mathcal{Y}^+$ is the support of p_Y and $X|y$ stands for $X|Y{=}y$.

For each $H \in \mathcal{H}$, we define the *H-mutual information* as

$$I_H(X;Y) = H(X) - H(X|Y). \tag{2}$$

and the *H-capacity* of a channel $K : \mathcal{X} \to \mathcal{Y}$ as

$$C_H(K) = \sup_{p_X} I_H(X,Y). \tag{3}$$

Definitions 1 and 2 are very general, subsuming most entropy measures used in the literature, such as[1]

- Shannon entropy, $H_1 = (x, -\sum p_i \log p_i)$,
- min-entropy, $H_\infty = (-\log(-x), -\max_i p_i)$,
- guessing entropy, $H_G = (x, \sum i p_{[i]})$, where $p_{[1]}, \ldots, p_{[n]}$ is a non-increasing rearrangement of p,
- g-entropies [1], $H_g = (-\log(-x), -V_g)$.

They also generalize the Rényi family of entropies [16], by choosing (for $\alpha > 1$) either $\eta(x) = \frac{\alpha}{1-\alpha}\log(-x)$, $F(X) = -\|p\|_\alpha$; or $\eta(x) = \frac{1}{1-\alpha}\log(-x)$, $F(X) = -\|p\|_\alpha^\alpha$.[2] The first choice yields the conditional entropy defined by Arimoto [3]:

$$H_\alpha(X|Y) = \frac{\alpha}{1-\alpha}\log\sum_{y\in\mathcal{Y}^+} p(y)\left\|p_{X|y}\right\|_\alpha \tag{4}$$

and the second yields another conditional entropy also defined in the literature [9]:

$$H'_\alpha(X|Y) = \frac{1}{1-\alpha}\log\sum_{y\in\mathcal{Y}^+} p(y)\left\|p_{X|y}\right\|_\alpha^\alpha. \tag{5}$$

Amongst these, only for (4) both $\lim_{\alpha\to 1} H_\alpha(X|Y) = H_1(X|Y)$ and $\lim_{\alpha\to\infty} H_\alpha(X|Y) = H_\infty(X|Y)$ hold [10].

An important property of core-concave entropies is that they respect the data processing inequality [11].

Theorem 1: For all core concave H, if $X \to Y \to Z$, then $H(X|Y) \le H(X|Z)$.

III. Preorders over Channels

This section introduces preorders over channels that share the same input alphabet. Throughout this section, let $K_1 : \mathcal{X} \to \mathcal{Y}$ and $K_2 : \mathcal{X} \to \mathcal{Z}$ share an input X and produce outputs Y, Z.

Channel K_2 is *degraded from* K_1 [7], written as $K_1 \ge_{\mathrm{d}} K_2$, if exists a channel $R : \mathcal{Y} \to \mathcal{Z}$ such that $K_2 = K_1 R$, i.e.,

$$K_2(z|x) = \sum_{y\in\mathcal{Y}} K_1(y|x)R(z|y) \quad \forall x \in \mathcal{X}, z \in \mathcal{Z}.$$

Another preorder introduced by Shannon [17] includes the one above. A channel K_1 is said to *include* K_2, indicated by $K_1 \ge_{\mathrm{sh}} K_2$, if there is a family of tuples $\{(g_i, T_i, R_i)\}_i$ of non-negative real numbers g_i and channels T_i, R_i such that

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

in which the convex sum of channels $tW_1 + (1-t)W_2$ is defined as $(tW_1{+}(1{-}t)W_2)(y|x) = tW_1(y|x){+}(1{-}t)W_2(y|x)$.

The two preorders defined above are independent of entropy measure, and for that reason we sometimes refer to them as *structural* preorders. Next, given $H \in \mathcal{H}$, we introduce three preorders that are sensitive over the choice of H. The first two are generalizations over the preorders introduced by [12].[3]

We say that a channel K_1 is *H-less noisy* than K_2, and write $K_1 \ge_{\mathrm{ln}}^H K_2$, if for all random variables U with finite support such that $U \to X \to (Y, Z)$, we have:

$$I_H(U;Y) \ge I_H(U;Z),$$

K_1 is said to be *H-more capable* than K_2, denoted as $K_1 \ge_{\mathrm{mc}}^H K_2$, if for all distributions of the input, X, we have:

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

Finally, we write $K_1 \ge_{\mathrm{c}}^H K_2$ to mean $C_H(K_1) \ge C_H(K_2)$.

Given $\mathcal{A} \subset \mathcal{H}$, we write $K_1 \ge_{\mathrm{ln}}^{\mathcal{A}} K_2$ if $\forall H \in \mathcal{A}, K_1 \ge_{\mathrm{ln}}^H K_2$. The relations $\ge_{\mathrm{mc}}^{\mathcal{A}}$ and $\ge_{\mathrm{c}}^{\mathcal{A}}$ are defined similarly.

A. Relationships between preorders

In this section, some relationships between the preorders defined above are explored. These results are similar to others in the literature, but generalized to $\mathcal{H}$.

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

$$K_1 \ge_{\mathrm{d}} K_2 \Rightarrow K_1 \ge_{\mathrm{ln}}^H K_2 \Rightarrow K_1 \ge_{\mathrm{mc}}^H K_2 \Rightarrow K_1 \ge_{\mathrm{c}}^H K_2.$$

Proof: The first implication is a direct consequence of Theorem 1. The other implications are straightforward. ∎

The converses of these implications do not hold in general, as shown for H_1 in [12]. However, the *H_∞-less noisy* ordering is sufficient to imply the degradedness ordering [4].

Theorem 2: $K_1 \ge_{\mathrm{ln}}^{H_\infty} K_2 \Rightarrow K_1 \ge_{\mathrm{d}} K_2$.

There is no known specific choice of a single core-concave entropy that would yield a similar result for the "more capable" orderings (or if one exists). However, the Theorem 9 in [15] shows that it is possible to recover degradedness ordering if the "more capable" orderings hold for a subset of core-concave entropies.

Theorem 3: Let $\mathcal{H}_g$ be the set of all g-entropies as defined in [1]. Then $K_1 \ge_{\mathrm{mc}}^{\mathcal{H}_g} K_2 \Rightarrow K_1 \ge_{\mathrm{d}} K_2$

Theorem 3 together with Proposition 1 yield the following equivalence.

Proposition 2: $K_1 \ge_{\mathrm{d}} K_2 \Leftrightarrow K_1 \ge_{\mathrm{ln}}^{\mathcal{H}} K_2 \Leftrightarrow K_1 \ge_{\mathrm{mc}}^{\mathcal{H}} K_2$

[1] All logarithms are taken to the base 2

[2] For $0 < \alpha < 1$, one may take instead: $\eta(x) = \frac{\alpha}{1-\alpha}\log(x)$, $F(X) = \|p\|_\alpha$ for the fist case and $\eta(x) = \frac{1}{1-\alpha}\log(x)$, $F(X) = \|p\|_\alpha^\alpha$ for the second.

[3] To be precise, they generalize a specific characterization of these orders, given by Propositions 1 and 2(ii) in [12]

IV. Channel-supermodular entropies

We now introduce the concept of *channel-supermodular* entropies, which were studied in [2] under the name "leakage-supermodular". They are based on *supermodular* functions over the euclidean vector space. A full account of supermodularity can be found in [19] and in [13, Chapter 6.D].

Consider the set $\mathbb{R}^n_+$ 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, i.e., given $\mathbf{r} = (r_1, \ldots, r_n)$ and $\mathbf{s} = (s_1, \ldots, s_n)$, $\mathbf{r} \preceq \mathbf{s}$ iff $r_i \leq s_i$ for all i. The $\preceq$ relation induces a partial order on $\mathbb{R}^n_+$, moreover, $\mathbb{R}^n_+$ and $\preceq$ form a lattice, with its "join" (least upper-bound) and "meet" (greatest lower-bound) respectively given by:

$$\mathbf{r} \vee \mathbf{s} = (\max(r_1, s_1), \ldots, \max(r_n, s_n))$$
$$\mathbf{r} \wedge \mathbf{s} = (\min(r_1, s_1), \ldots, \min(r_n, s_n))$$

Definition 3: A function $\phi : \mathbb{R}^n_+ \to \mathbb{R}$ is *supermodular* (over the above lattice) if, for all $\mathbf{r}, \mathbf{s} \in \mathbb{R}^n_+$,

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

The following useful characterization of supermodular functions is an immediate consequence of Corollary 2.6.1 in [19].

Theorem 4: Let $\phi : \mathbb{R}^n_+ \to \mathbb{R}$ and let $\mathbf{e}_1, \ldots \mathbf{e}_n$ 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 $i \neq j$,

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

Whenever the function is smooth, the inequality above reduces to a property on the second derivatives [19].

Theorem 5: Let $\phi : \mathbb{R}^n_+ \to \mathbb{R}$ have second derivatives. Then, ϕ is supermodular if and only if, for all $\mathbf{r} \in \mathbb{R}^n_+$ and all i, j with $i \neq j$,

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

The relationship between some core-concave entropies and supermodularity is established by the following definition.

Definition 4: Let $H = (\eta, F)$ be a core concave entropy. Define, for all $\mathbf{r} \in \mathbb{R}^n_+$:

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

where by convention $G_F(0, \ldots, 0) = 0$. H is *channel-supermodular* if G_F is supermodular. We denote by $\mathcal{S} \subset \mathcal{H}$ the set of all channel-supermodular entropies.

The supermodular property of the functions G_F turns out to be a powerful tool for analytically reasoning about channels. In fact, for each $H \in \mathcal{S}$, the conditional entropy of X given Y can be defined in terms of G_F and the joint probability distribution over X, Y, as follows

$$H(X|Y) = \eta\left(\sum_{y \in \mathcal{Y}} G_F\left(p_{X,Y}(x_1, y), \ldots, p_{X,Y}(x_n, y)\right)\right) \tag{6}$$

The next result from [2] confirms that some of the most usual entropy measures are channel-supermodular.

Theorem 6: Shannon and min-entropy, and more generally, Rényi entropies (as in (4)) are channel-supermodular. Guessing entropy is also channel-supermodular.

Proof: Due to space restrictions, we only provide the proof for Rényi's entropies. A separate proof for Shannon entropy can be derived in a very similar way and a proof for min-entropy can be obtained directly from Definition 3. A proof for guessing entropy can be obtained using Theorem 4.

For Rényi entropies of order $\alpha < 1$, we have,

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

and similarly, $G_F(\mathbf{r}) = -\|\mathbf{r}\|_\alpha$ if $\alpha > 1$. 0 For $\mathbf{r} \in \mathbb{R}^n_{>0}$ and i, j with $i \neq j$, we have

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

which is negative for $\alpha > 1$ and positive for $\alpha < 1$. Thus, from Theorem 5, G_F restricted to $\mathbb{R}^n_{>0}$ is supermodular .

If $\mathbf{r}$ has components equal to zero, the second derivatives may not exist. Let $\epsilon > 0$ and $\mathbf{r}' = \mathbf{r} + \epsilon(\sum_k \mathbf{e}_k)$. As $\mathbf{r}' \in \mathbb{R}^n_{>0}$, Theorem 4 yields, for all $\delta_1, \delta_2 > 0$

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

By taking $\epsilon \to 0$, we obtain that G_F is supermodular over $\mathbb{R}^n_+$. ∎

A. The JoinMeet Operator

Let $K : \mathcal{X} \to \mathcal{Y}$ be a channel, with $\mathcal{Y} = \{y_1, \ldots, y_m\}$, and let K^i be the column of K corresponding to output y_i. Define, for $i \neq j$, the *JoinMeet* 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, for all $H \in \mathcal{S}$, the operator $\Diamond_{i,j}$ is monotonic with I_H.

Theorem 7: Given K_1 and i, j, $K_1 \geq^{\mathcal{S}}_{\text{mc}} \Diamond_{i,j} K_1$

Proof: Let $H = (\eta, F) \in \mathcal{S}$ and define G_F as in Definition 4. 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, $p_{X,Y_2}(x_k, y_i) = \max(p_{X,Y_1}(x_k, y_i), p_{X,Y_1}(x_k, y_j))$, 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,

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

where the inequality follows from G_F being supermodular. From (6) and η being increasing, it follows that $H(X|Y_1) \leq H(X|Y_2)$, which is equivalent to $I_H(X; Y_1) \geq I_H(X; Y_2)$. ∎

B. A new structural ordering

Theorem 7 yields some immediate new results for reasoning about channel ordering, as the JoinMeet operator is not, in general, captured by the "degradedness" relation. Consider, for instance, the following channels K_1, K_2

$$\begin{pmatrix} 0.5 & 0.5 \\ 0.6 & 0.4 \\ 0.2 & 0.8 \end{pmatrix} \qquad \begin{pmatrix} 0.5 & 0.5 \\ 0.6 & 0.4 \\ 0.8 & 0.2 \end{pmatrix} \tag{7}$$

Then, we have that $K_2 = \Diamond_{1,2} K_1$, but $K_1 \not\geq_{\mathrm{d}} K_2$, as one can establish that there is no channel R such that $K_2 = K_1 R$. Therefore, for all $H \in \mathcal{S}$, Theorem 7 gives a novel sufficient condition for a channel being H-more capable than another, and one which is straightforward to check.

This leads us to define two new structural preorders over channels. The first is given by the JoinMeet operators $\Diamond_{i,j}$. The subscript s is used to associate it with supermodularity.

Definition 5: $K_1 \geq_{\mathrm{s}} K_2$ if there is a finite collection of tuples (i_k, j_k) such that $K_2 = \Diamond_{i_1,j_1}(\Diamond_{i_2,j_2}(\dots \Diamond_{i_m,j_m} K_1))$.

We also define a preorder induced by combining $\geq_{\mathrm{d}}$ and $\geq_{\mathrm{s}}$. That is, a preorder that relates channels that are obtained by a sequence of degrading and JoinMeet operations. From Proposition 1 and Theorem 7, this preorder also implies the "more capable" preorder for channel-supermodular entropies.

Definition 6: $K_1 \geq_{\mathrm{ds}} K_2$ if there are channels $W_1, \dots, W_n$ such that $K_1 \geq_0 W_1 \geq_1 \dots \geq_{\mathrm{n}-1} W_n \geq_{\mathrm{n}} K_2$, where each $\geq_{\mathrm{i}}$ stands for $\geq_{\mathrm{d}}$ or $\geq_{\mathrm{s}}$.

V. Relations between preorders for channel-supermodular entropies

This section establishes the relationships amongst the preorders introduced in Sections III and IV-B, when the entropies considered are in the set $\mathcal{S}$.

TABLE I
IMPLICATIONS BETWEEN PREORDERS: CIRCLED LETTERS ARE RESULTS PROVEN IN THIS PAPER.

$\Rightarrow$	$\geq_{\mathrm{d}}$	$\geq_{\mathrm{ds}}$	$\geq_{\mathrm{ln}}^{\mathcal{S}}$	$\geq_{\mathrm{mc}}^{\mathcal{S}}$	$\geq_{\mathrm{sh}}$
$\geq_{\mathrm{d}}$		(Y) Prop 3	Y	Y	Y
$\geq_{\mathrm{ds}}$	(N) Prop 3		(N) Prop 3	(Y) Prop 3	(N) Prop 3
$\geq_{\mathrm{ln}}^{\mathcal{S}}$	Y	Y		Y	Y
$\geq_{\mathrm{mc}}^{\mathcal{S}}$	(N) Prop 4	?	(N) Prop 4		(N) Prop 5
$\geq_{\mathrm{sh}}$	N	(N) Prop 3	N	N	

Table I summarizes the relations between the different preorders, with new results encircled. For example the value (N) in the cell $(\geq_{\mathrm{ds}}, \geq_{\mathrm{d}})$ means that $\geq_{\mathrm{ds}}$ does not imply $\geq_{\mathrm{d}}$.

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

Proposition 3:

1) $K_1 \geq_{\mathrm{d}} K_2 \Rightarrow K_1 \geq_{\mathrm{ds}} K_2$ and $K_1 \geq_{\mathrm{s}} K_2 \Rightarrow K_1 \geq_{\mathrm{ds}} K_2$,
2) $K_1 \geq_{\mathrm{ds}} K_2 \not\Rightarrow K_1 \geq_{\mathrm{d}} K_2$,
3) $K_1 \geq_{\mathrm{ds}} K_2 \not\Rightarrow K_1 \geq_{\mathrm{s}} K_2$,
4) $K_1 \geq_{\mathrm{ds}} K_2 \not\Rightarrow K_1 \geq_{\mathrm{ln}}^{\mathcal{S}} K_2$,
5) $K_1 \geq_{\mathrm{ln}}^{\mathcal{S}} K_2 \Rightarrow K_1 \geq_{\mathrm{ds}} K_2$,
6) $K_1 \geq_{\mathrm{ds}} K_2 \Rightarrow K_1 \geq_{\mathrm{mc}}^{\mathcal{S}} K_2$,
7) $K_1 \geq_{\mathrm{ds}} K_2 \not\Rightarrow K_1 \geq_{\mathrm{sh}} K_2$,
8) $K_1 \geq_{\mathrm{sh}} K_2 \not\Rightarrow K_1 \geq_{\mathrm{ds}} K_2$.

Proof: 1) follows immediately from Definition 6, and 2) from the example channels in (7). Statement 3) is proved by the following channels

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix} \geq_{\mathrm{d}} \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}$$

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 [12], where it is proven that $K_1 \not\geq_{\mathrm{ln}}^{H_1} K_2$. However $K_1 \geq_{\mathrm{ds}} K_2$, as

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1/2 & 1/2 \end{pmatrix} \geq_{\mathrm{d}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 1/2 \\ 1/2 & 1/4 & 1/4 \end{pmatrix} \geq_{\mathrm{s}} \begin{pmatrix} 1 & 0 & 0 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/4 & 1/4 \end{pmatrix} \geq_{\mathrm{d}} \begin{pmatrix} 1 & 0 \\ 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}.$$

Statement 5) follows from 1) and from Theorem 2. Statement 6) follows directly from Definition 6, Proposition 1 and Theorem 7. For 7), consider the following channels K_1, K_2

$$\begin{pmatrix} 1/2 & 1/2 & 0 \\ 0 & 2/3 & 1/3 \\ 1/2 & 1/8 & 3/8 \end{pmatrix} \qquad \begin{pmatrix} 1/2 & 1/2 & 0 \\ 2/3 & 0 & 1/3 \\ 1/2 & 1/8 & 3/8 \end{pmatrix}$$

Then, $K_2 = \Diamond_{1,2} K_1$. However, $K_1 \not\geq_{\mathrm{sh}} K_2$, as can be checked by linear programming [20]. Finally, for 8) consider channels K_1, K_2 as follows

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

As K_2 is obtained from K_1 by row permutation, $K_1 \geq_{\mathrm{sh}} K_2$. However, it is readily seen that $K_1 \not\geq_{\mathrm{mc}}^{H_1} K_2$ (take, for example, $p_X = (1/2, 1/2, 0)$). Thus, 6) yields $K_1 \not\geq_{\mathrm{ds}} K_2$. ∎

Proposition 3.6 yields a new analytical approach to deciding whether $K_1 \geq_{\mathrm{mc}}^{H} K_2$, for $H \in \mathcal{S}$, which relies only in the structural properties of the channel. To illustrate this technique, consider the following channels K_1, K_2 introduced in [1].

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

While these channels are not ordered w.r.t. $\geq_{\mathrm{d}}$, the authors claimed they had experimental evidence, but no proof, that $K_1 \geq_{\mathrm{mc}}^{H_1} K_2$. In light of Proposition 3.6, this follows from

$$\begin{pmatrix} 1/2 & 0 & 1/2 \\ 0 & 1/2 & 1/2 \\ 1/2 & 1/2 & 0 \end{pmatrix} \geq_{\mathrm{s}} \begin{pmatrix} 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{pmatrix} \geq_{\mathrm{d}} \begin{pmatrix} 1/4 & 3/4 \\ 1/4 & 3/4 \\ 3/5 & 2/5 \end{pmatrix}$$

Proposition 3 yields immediate consequences for the relationship amongst the other preorders, when entropies in $\mathcal{S}$ are considered. First, note that the equivalent of Proposition 2 does not hold if $\mathcal{H}$ is substituted for $\mathcal{S}$.

Proposition 4: $K_1 \geq_{\mathrm{mc}}^{\mathcal{S}} K_2 \not\Rightarrow K_1 \geq_{\mathrm{d}} K_2$ and $K_1 \geq_{\mathrm{mc}}^{\mathcal{S}} K_2 \not\Rightarrow K_1 \geq_{\mathrm{ln}}^{\mathcal{S}} K_2$

Proof: From Proposition 3.2, there are channels K_1,K_2 such that $K_1 \geq_{\mathrm{ds}} K_2$ and $K_1 \not\geq_{\mathrm{d}} K_2$. For such channels, Proposition 3.6 implies $K_1 \geq_{\mathrm{mc}}^{\mathcal{S}} K_2$, and the first result follows. The second result then follows by noting that Proposition 1 and Theorem 2 imply $K_1 \geq_{\mathrm{ln}}^{\mathcal{S}} K_2 \Leftrightarrow K_1 \geq_{\mathrm{d}} K_2$. ∎

Proposition 4 has some interesting ramifications in the field of Quantitative Information Flow. It has been recently argued that the degradedness order, in light of Theorem 3, ought to be the standard to establishing whether a channel is more secure than another (see the discussion at the end of Section 6 in [15]). Our result, however, shows that this order is too strong if one can limit their considerations only to entropies in $\mathcal{S}$.

It follows immediately from Propositions 3.6 and 3.7 that the more capable ordering, even when quantified over $\mathcal{S}$, does not imply Shannon ordering.

Proposition 5: $K_1 \geq_{\mathrm{mc}}^{\mathcal{S}} K_2 \not\Rightarrow K_1 \geq_{\mathrm{sh}} K_2$.

VI. Results on channel capacity

In the same paper that Shannon introduced the inclusion ordering, he established that $K_1 \geq_{\mathrm{sh}} K_2 \Rightarrow K_1 \geq_{\mathrm{c}}^{H_1} K_2$ [17]. Theorem 7 allows us to derive similar results for a new preorder $\geq_{\mathrm{shs}}$, which is an extension of $\geq_{\mathrm{sh}}$ with $\geq_{\mathrm{s}}$.

Definition 7: $K_1 \geq_{\mathrm{shs}} K_2$ if there are channels $W_1, \ldots, W_n$ such that $K_1 \geq_0 W_1 \geq_1 \ldots \geq_{\mathrm{n}-1} W_n \geq_{\mathrm{n}} K_2$ where each $\geq_{\mathrm{i}}$ stands for $\geq_{\mathrm{sh}}$ or $\geq_{\mathrm{s}}$

First, we establish the relationship between $\geq_{\mathrm{shs}}$ and the other structural preorders.

Proposition 6: For all channels K_1, K_2

1) $K_1 \geq_{\mathrm{sh}} K_2 \Rightarrow K_1 \geq_{\mathrm{shs}} K_2$
2) $K_1 \geq_{\mathrm{ds}} K_2 \Rightarrow K_1 \geq_{\mathrm{shs}} K_2$
3) $K_1 \geq_{\mathrm{shs}} K_2 \not\Rightarrow K_1 \geq_{\mathrm{s}} K_2$
4) $K_1 \geq_{\mathrm{shs}} K_2 \not\Rightarrow K_1 \geq_{\mathrm{sh}} K_2$

Proof: 1) is clear and 2) is immediate from the definitions of $\geq_{\mathrm{ds}}$ and $\geq_{\mathrm{shs}}$. 3) and 4) follow from 2) and Propositions 3.3 and 3.7 ∎

In particular, Proposition 6 reveals that $\geq_{\mathrm{shs}}$ includes $\geq_{\mathrm{sh}}$. We finish this section by stating a couple of results, that are direct consequences of Definition 7, Proposition 1 and Theorem 7.

Proposition 7: For all $H \in \mathcal{S}$, if $K_1 \geq_{\mathrm{sh}} K_2 \Rightarrow K_1 \geq_{\mathrm{c}}^{H} K_2$, then $K_1 \geq_{\mathrm{shs}} K_2 \Rightarrow K_1 \geq_{\mathrm{c}}^{H} K_2$

Proposition 8:

1) $K_1 \geq_{\mathrm{shs}} K_2 \Rightarrow K_1 \geq_{\mathrm{c}}^{H_1} K_2$
2) $K_1 \geq_{\mathrm{shs}} K_2 \Rightarrow K_1 \geq_{\mathrm{c}}^{H^\infty} K_2$

VII. Conclusions and future work

Two new preorders on channels have been introduced, based on a supermodularity property shared by some of the most common entropies in the literature. Several relationships between this ordering and existing ones have been established.

Two main open problems arise from this work: 1) Are $\geq_{\mathrm{ds}}$ and $\geq_{\mathrm{mc}}^{\mathcal{S}}$ equivalent? 2) Does $\geq_{\mathrm{shs}}$ imply $\geq_{\mathrm{c}}^{\mathcal{S}}$?

References

[1] M. S. Alvim, K. Chatzikokolakis, C. Palamidessi, and G. Smith. Measuring information leakage using generalized gain functions. In *Proc. IEEE 25th Computer Security Foundations Symposium (CSF)*, pages 265–279, 2012.

[2] A. Américo, MHR. Khouzani, and P. Malacaria. Deterministic Channel Design for Minimum Leakage. In *Proc. IEEE 32nd Computer Security Foundations Symposium (CSF)*, pages 428–441, 2019.

[3] S. Arimoto. Information measures and capacity of order α for discrete memoryless channels. *Topics in information theory*, 1977.

[4] F. Buscemi. Comparison of noisy channels and reverse data-processing theorems. In *Proc. 2017 IEEE Information Theory Workshop (ITW)*, pages 489–493, Nov 2017.

[5] D. Clark, S. Hunt, and P. Malacaria. Quantitative information flow, relations and polymorphic types. *J. Log. and Comput.*, 15(2):181–199, April 2005.

[6] J. Cohen, J. H. B. Kempermann, and G. Zbaganu. *Comparisons of Stochastic Matrices with Applications in Information Theory, Statistics, Economics and Population*. Springer Science & Business Media, 1998.

[7] T. M. Cover. Broadcast channels. *IEEE Transactions on Information Theory*, 18(1):2–14, 1972.

[8] A. A. El Gamal. The capacity of a class of broadcast channels. *IEEE Transactions on Information Theory*, 25(2):166–169, March 1979.

[9] M. Hayashi. Exponential decreasing rate of leaked information in universal random privacy amplification. *IEEE Transactions on Information Theory*, 57(6):3989–4001, 2011.

[10] M. Iwamoto and J. Shikata. Information theoretic security for encryption based on conditional rényi entropies. In C. Padró, editor, *Information Theoretic Security*, pages 103–121, Cham, 2014. Springer International Publishing.

[11] MHR. Khouzani and P. Malacaria. Generalised Entropies and Metric-Invariant Optimal Countermeasures for Information Leakage under Symmetric Constraints. *IEEE Transactions on Information Theory*, 2018.

[12] J. Korner and K. Marton. Comparison of two noisy channels. *Topics in Information Theory*, pages 411–423, 1977.

[13] A. W. Marshall, I. Olkin, and B. C. Arnold. *Inequalities: theory of majorization and its applications*, volume 143. Mathematics In Science And Engineering, Academic Press, 1979.

[14] J. L. Massey. Guessing and entropy. In *Proc. IEEE Int. Symposium on Information Theory (ISIT)*, page 204, June 1994.

[15] A. McIver, C. Morgan, G. Smith, B. Espinoza, and L. Meinicke. Abstract channels and their robust information-leakage ordering. In *Proc. 3rd Int. Conf. Principles of Security and Trust (POST)*, volume 8414 of *LNCS*, pages 83–102. Springer, 2014.

[16] A. Rényi. On Measures of Entropy and Information. In *Proc. 4th Berkeley Symposium on Mathematics, Statistics, and Probability*, pages 547–561, 1961.

[17] C. E. Shannon. A note on a partial ordering for communication channels. *Information and control*, 1(4):390–397, 1958.

[18] G. Smith. On the foundations of quantitative information flow. In *Proc. 12th Int. Conf. Foundations of Software Science and Computational Structures (FOSSACS)*, volume 5504 of *LNCS*, pages 288–302. Springer, 2009.

[19] D. M. Topkis. *Supermodularity and complementarity*. Princeton university press, 1998.

[20] Y. Zhang and C. Tepedelenliolu. Analytical and numerical characterizations of shannon ordering for discrete memoryless channels. *IEEE Transactions on Information Theory*, 60(1):72–83, Jan 2014.