entropy

MDPI

Article

On the α-q-Mutual Information and the α-q-Capacities

Velimir M. Ilić [1,*] and Ivan B. Djordjević [2]

[1] Mathematical Institute of the Serbian Academy of Sciences and Arts, Kneza Mihaila 36, 11000 Beograd, Serbia
[2] Department of Electrical and Computer Engineering, University of Arizona, 1230 E. Speedway Blvd., Tucson, AZ 85721, USA; ivan@email.arizona.edu
* Correspondence: velimir.ilic@gmail.com

Abstract: The measures of information transfer which correspond to non-additive entropies have intensively been studied in previous decades. The majority of the work includes the ones belonging to the Sharma–Mittal entropy class, such as the Rényi, the Tsallis, the Landsberg–Vedral and the Gaussian entropies. All of the considerations follow the same approach, mimicking some of the various and mutually equivalent definitions of Shannon information measures, and the information transfer is quantified by an appropriately defined measure of mutual information, while the maximal information transfer is considered as a generalized channel capacity. However, all of the previous approaches fail to satisfy at least one of the ineluctable properties which a measure of (maximal) information transfer should satisfy, leading to counterintuitive conclusions and predicting nonphysical behavior even in the case of very simple communication channels. This paper fills the gap by proposing two parameter measures named the α-q-mutual information and the α-q-capacity. In addition to standard Shannon approaches, special cases of these measures include the α-mutual information and the α-capacity, which are well established in the information theory literature as measures of additive Rényi information transfer, while the cases of the Tsallis, the Landsberg–Vedral and the Gaussian entropies can also be accessed by special choices of the parameters α and q. It is shown that, unlike the previous definition, the α-q-mutual information and the α-q-capacity satisfy the set of properties, which are stated as axioms, by which they reduce to zero in the case of totally destructive channels and to the (maximal) input Sharma–Mittal entropy in the case of perfect transmission, which is consistent with the maximum likelihood detection error. In addition, they are non-negative and less than or equal to the input and the output Sharma–Mittal entropies, in general. Thus, unlike the previous approaches, the proposed (maximal) information transfer measures do not manifest nonphysical behaviors such as sub-capacitance or super-capacitance, which could qualify them as appropriate measures of the Sharma–Mittal information transfer.

Keywords: rényi entropy; tsallis entropy; landsberg—vedral entropy; gaussian entropy; sharma—mittal entropy; α-mutual information; α-channel capacity

Citation: Ilić, V.M.; Djordjević, I.B. On the α-q-Mutual Information and the α-q-Capacities. *Entropy* **2021**, *23*, 702. https://doi.org/10.3390/e23060702

Academic Editors: Petr Jizba and Jan Korbel

Received: 12 February 2021
Accepted: 26 May 2021
Published: 1 June 2021

1. Introduction

In the past, extensive work has been written on defining the information measures which generalize the Shannon entropy [1], such as the one-parameter Rényi entropy [2], the Tsallis entropy [3], the Landsberg–Vedral entropy [4], the Gaussian entropy [5], and the two-parameter Sharma–Mittal entropy [5,6], which reduces to former ones for special choices of the parameters. The Sharma–Mittal entropy can axiomatically be founded as the unique q-additive measure [7,8] which satisfies generalized Shannon–Kihinchin axioms [9,10] and which has widely been explored in different research fields starting from statistics [11] and thermodynamics [12,13] to quantum mechanics [14,15], machine learning [16,17] and cosmology [18,19]. The Sharma–Mittal entropy has also been recognized in the field of information theory, where the measures of conditional Sharma–Mittal entropy [20], Sharma–Mittal divergences [21] and Sharma–Mittal entropy rate [22] have been established and analyzed.

Considerable research has also been done in the field of communication theory in order to analyze information transmission in the presence of noise if, instead of Shannon's entropy, the information is quantified with (instances of) Sharma–Mittal entropy and, in general, the information transfer is quantified by an appropriately defined measure of mutual information, while the maximal information transfer is considered as a generalized channel capacity. Thus, after Rényi's proposal for the additive generalization of Shannon entropy [2], several different definitions for Rényi information transfer were proposed by Sibson [23], Arimoto [24], Augustin [25], Csiszar [26], Lapidoth and Pfister [27] and Tomamichel and Hayashi [28]. These measures have been explored thoroughly and their operational characterization in coding theory, hypothesis testing, cryptography and quantum information theory was established, which qualifies them as a reasonable measure of Rényi information transfer [29]. Similar attempts have also been made in the case of non-additive entropies. Thus, starting from the work of Daroczy [30], who introduced a measure for generalized information transfer related to the Tsallis entropy, several attempts followed for the measures which correspond to non-additive particular instances of the Sharma–Mittal entropy, so the definitions for the Rényi information transfer were considered in [24,31], for the Tsallis information transfer in [32] and for the Landsber–Vedral information transfer in [4,33].

In this paper we provide a general treatment of the Sharma–Mittal entropy transfer and a detailed analysis of existing measures, showing that all of the definitions related to non-additive entropies fail to satisfy at least one of the ineluctable properties common to the Shannon case, which we state as axioms, by which the information transfer has to be non-negative, less than the input and output uncertainty, equal to the input uncertainty in the case of perfect transmission and equal to zero, in the case of a totally destructive channel. Thus, breaking some of these axioms implies unexpected and counterintuitive conclusions about the channels, such as achieving super-capacitance or sub-capacitance [4], which could be treated as nonphysical behavior. As an alternative, we propose the α-q-mutual information as a measure of Sharma–Mittal information transfer, maximized with the α-q-capacity. The α-q mutual information generalizes the α-mutual information by Arimoto [24], which is defined as a q-difference between the input Sharma–Mittal entropy and the appropriately defined conditional Sharma–Mittal entropy if the output is given, while the α-q-capacity represents a generalization of Arimoto's α-capacity in the case of $q = 1$. In addition, several other instances can be obtained by specifying the values of parameters α and q, which includes the information transfer measures for the Tsallis, the Landsber–Vedral and the Shannon entropy, as well as the case of the Gaussian entropy which was not considered before in the context of information transmission.

The paper is organized as follows. The basic properties and special instances of the Sharma–Mittal entropy are listed in Section 2. Section 3 reviews the basics of communication theory, introduces the basic communication channels and establishes the set of axioms which information transfer measures should satisfy. The information transfer measures which are defined by Arimoto are introduced in Section 4, and the alternative definitions for Rényi information transfer measures are discussed in Section 5. Finally, the α-q-mutual information and the α-q-capacities are proposed and their properties analyzed in Section 6 while the previously proposed measures of Sharma–Mittal entropy transfer are discussed in Section 7.

2. Sharma–Mittal Entropy

Let the sets of positive and nonnegative real numbers be denoted with $\mathbb{R}^+$ and $\mathbb{R}_0^+$, respectively, and let the mapping $\eta_q : \mathbb{R} \to \mathbb{R}$ be defined in

$$\eta_q(x) = \begin{cases} x, & \text{for} \quad q = 1 \\ \dfrac{2^{(1-q)\,x} - 1}{(1-q)\ln 2}, & \text{for} \quad q \neq 1 \end{cases} \tag{1}$$

so that its inverse is given in

$$\eta_q^{-1}(x) = \begin{cases} x, & \text{for} \quad q = 1 \\ \dfrac{1}{1-q}\log((1-q)x\ln 2 + 1), & \text{for} \quad q \neq 1 \end{cases}. \tag{2}$$

The mapping η_q and its inverse are increasing continuous (hence invertible) functions such that $\eta(0) = 0$. The q-logarithm is defined in

$$\mathrm{Log}_q(x) = \eta_q(\log x) = \begin{cases} \log x, & \text{for} \quad q = 1 \\ \dfrac{x^{(1-q)} - 1}{(1-q)\ln 2}, & \text{for} \quad q \neq 1 \end{cases}, \tag{3}$$

and its inverse, the q-exponential, is defined in

$$\mathrm{Exp}_q(y) = \begin{cases} 2^y, & \text{for} \quad q = 1 \\ (1 + (1-q)y\ln 2)^{\frac{1}{1-q}} & \text{for} \quad q \neq 1 \end{cases}, \tag{4}$$

for $1 + (1-q)y\ln 2 > 0$. Using η_q, we can define the pseudo-addition operation $\oplus_q$ [7,8]

$$x \oplus_q y = \eta_q\left(\eta_q^{-1}(x) + \eta_q^{-1}(y)\right) = x + y + (1-q)xy; \quad x, y \in \mathbb{R}, \tag{5}$$

and its inverse operation, the pseudo substraction

$$x \ominus_q y = \eta_q\left(\eta_q^{-1}(x) - \eta_q^{-1}(y)\right) = \frac{x-y}{1 + (1-q)y\ln 2}; \quad x, y \in \mathbb{R}. \tag{6}$$

The $\oplus_q$ can be rewritten in terms of the generalized logarithm by settings $x = \log u$ and $y = \log v$ so that

$$\mathrm{Log}_q(u \cdot v) = \mathrm{Log}_q(u) \oplus_q \mathrm{Log}_q(v); \quad u, v \in \mathbb{R}_+. \tag{7}$$

Let the set of all n-dimensional distributions be denoted with

$$\Delta_n \equiv \left\{ (p_1, \ldots, p_n) \middle| \; p_i \geq 0, \sum_{i=1}^{n} p_i = 1 \right\}; \quad n > 1. \tag{8}$$

Let the function $H_n : \Delta_n \to \mathbb{R}_0^+$ satisfy the following the Shannon–Khinchin axioms, for all $n \in \mathbb{N}$, $n > 1$.

GSK1 H_n is continuous in Δ_n;

GSK2 H_n takes its largest value for the uniform distribution, $U_n = (1/n, \ldots, 1/n) \in \Delta_n$, i.e., $H_n(P) \leq H_n(U_n)$, for any $P \in \Delta_n$;

GSK3 H_n is expandable: $H_{n+1}(p_1, p_2, \ldots, p_n, 0) = H_n(p_1, p_2, \ldots, p_n)$ for all $(p_1, \ldots, p_n) \in \Delta_n$;

GSK4 Let $P = (p_1, \ldots, p_n) \in \Delta_n$, $PQ = (r_{11}, r_{12}, \ldots, r_{nm}) \in \Delta_{nm}$, $n, m \in \mathbb{N}$, $n, m > 1$ such that $p_i = \sum_{j=1}^{m} r_{ij}$, and $Q_{|k} = (q_{1|k}, \ldots, q_{m|k}) \in \Delta_m$, where $q_{i|k} = r_{ik}/p_k$ and $\alpha \in \mathbb{R}_0^+$ are some fixed parameters. Then,

$$H_{nm}(PQ) = H_n(P) \oplus_q H_m(Q|P), \quad \text{where} \quad H_m(Q|P) = f^{-1}\left(\sum_{k=1}^{n} p_k^{(\alpha)} f(H_m(Q_{|k}))\right), \tag{9}$$

where f is an invertible continuous function and $P^{(\alpha)} = (p_1^{(\alpha)}, \ldots, p_n^{(\alpha)}) \in \Delta_n$ is the α-escort distribution of distribution $P \in \Delta_n$ defined in

$$p_k^{(\alpha)} = \frac{p_k^\alpha}{\sum_{i=1}^{n} p_i^\alpha}, \quad k = 1, \ldots, n, \quad \alpha > 0. \tag{10}$$

GSK5 $H_2\left(\frac{1}{2},\frac{1}{2}\right) = \mathrm{Log}_q(1)$.

As shown in [9], the unique function H_n, which satisfies [GSK1]-[GSK5], is Sharma–Mittal entropy [6].

In the following paragraphs we will assume that X and Y are discrete jointly distributed random variables taking values from sample spaces $\{x_1,\ldots,x_n\}$ and $\{y_1,\ldots,y_m\}$, and distributed in accordance to $P_X \in \Delta_n$ and $P_Y \in \Delta_m$, respectively. In addition, the joint distribution of X and Y will be denoted in $P_{X,Y} \in \Delta_{nm}$ and the conditional distribution of X given Y will be denoted in $P_{X|Y} = \frac{P_{X,Y}(x,y)}{P_Y(y)} \in \Delta_m$, provided that $P_Y(y) > 0$. We will identify the entropy of a random variable X with the entropy of its distribution P_X and the Sharma–Mittal entropy will be denoted with $H_{\alpha,q}(X) \equiv H_n(P_X)$.

Thus, for a random variable which is distributed to X, Sharma–Mittal entropy can be expressed in

$$H_{\alpha,q}(X) = \frac{1}{1-q}\left(\left(\sum_x P_X(x)^\alpha\right)^{\frac{1-q}{1-\alpha}} - 1\right), \tag{11}$$

and it can equivalently be expressed as the η_q transformation of Rényi entropy as in

$$H_{\alpha,q}(X) \equiv \eta_q(R_\alpha(X)). \tag{12}$$

Sharma–Mittal entropy, for $\alpha, q \in \mathbb{R}_0^+ \setminus 1$, being a continuous function of the parameters and the sums goes over the support of P_X. Thus, in the case of $q = 1$, $\alpha \neq 1$, Sharma–Mittal reduces to Rényi entropy of order α [2]

$$R_\alpha(X) \equiv H_{\alpha,1}(X) = \frac{1}{1-\alpha}\log\left(\sum_x P_X(x)^\alpha\right), \tag{13}$$

which further reduces to Shannon entropy for $\alpha = 1, q = 1$, [34]

$$S(X) \equiv H_{1,1}(X) = \sum_x P_X(x)\log P_X(x), \tag{14}$$

while in the case of $q \neq 1$, $\alpha = 1$ it reduces to Gaussian entropy [5]

$$G_q(X) \equiv H_{1,q}(X) = \frac{1}{(1-q)\ln 2}\left(\prod_{i=1}^n P_X(x)^{P_X(x)} - 1\right). \tag{15}$$

In addition, Tsallis entropy [3] is obtained for $\alpha = q \neq 1$,

$$T_q(X) \equiv \frac{1}{(1-q)\ln 2}\left(\sum_x P_X(x)^q - 1\right), \tag{16}$$

while in the case of for $q = 2-\alpha$ it reduces to the Landsberg–Vedral entropy [4]

$$L_\alpha(X) \equiv H_{\alpha,2-\alpha}(X) = \frac{1}{(\alpha-1)\ln 2}\left(\frac{1}{\sum_x P_X(x)^\alpha} - 1\right). \tag{17}$$

3. Sharma–Mittal Information Transfer Axioms

One of the main goals of information and communication theories is characterization and analysis of the information transfer between sender X and receiver Y, which communicate through a channel. The sender and receiver are described by probability distributions P_X and P_Y while the communication channel with the input X and the output Y is described by the transition matrix $P_{Y|X}$:

$$P_{Y|X}^{(i,j)} \equiv P_{Y|X}(y_j|x_i). \tag{18}$$

We assume that maximum likelihood detection is performed at the receiver, which is defined by the mapping $d : \{y_1, \ldots, y_m\} \to \{x_1, \ldots, x_n\}$ as follows:

$$d(y_j) = x_i \quad \Leftrightarrow \quad P_{Y|X}(y_j|x_i) > P_{Y|X}(y_j|x_k); \quad \text{for all } k \neq i, \tag{19}$$

assuming that the inequality in (19) is uniquely satisfied. Thus, if the input symbol x_i is sent and the output symbol y_j is received, the x_i will be detected if $x_i = d(y_j)$ and a detection error will be made otherwise, and we define the error function functions $\phi : \{x_1, \ldots, x_m\} \times \{y_1, \ldots, y_m\} \to \{0, 1\}$ as in

$$\phi(x_i, y_j) = \begin{cases} 1, & \text{if } x_i = d(y_j) \\ 0, & \text{otherwise,} \end{cases} \tag{20}$$

the detection error if a symbol x_i is sent

$$P_{err}(x_i) = \sum_{y_j} P_{Y|X}(y_j|x_i)\phi(x_i, y_j); \quad \text{for all} \quad x_i, \tag{21}$$

as well as the average detection error

$$\bar{P}_{err} = \sum_{x_i} P_X(x_i) P_{err}(x_i) = \sum_{x_i, y_j} P_{X,Y}(x, y)\phi(x_i, y_j). \tag{22}$$

Totally destructive channel: A channel is said to be totally destructive if

$$P_{Y|X}^{(i,j)} = P_{Y|X}(y_j|x_i) = P_Y(y_j) = \frac{1}{m}; \quad \text{for all} \quad x_i, \tag{23}$$

i.e., if the sender X and receiver Y are described by independent random variables,

$$X \perp\!\!\!\perp Y \quad \Leftrightarrow \quad P_{X,Y}(x, y) = P_X(x) P_Y(y), \tag{24}$$

where the relationship of independence is denoted in $\perp\!\!\!\perp$. In this case, $\phi_i(y_j) = 1$ for all y_j and the probability of error is $P_{err}(x_i) = 1$; for all x_i, as well as the average probability of error $\bar{P}_{err} = 1$, which means that a correct maximum likelihood detection is not possible.

Perfect communication channel: A channel is said to be perfect if for every x_i,

$$P_{Y|X}(y_j|x_i) > 0, \quad \text{for at least one } y_j \tag{25}$$

and for every y_j

$$P_{Y|X}(y_j|x_i) > 0, \quad \text{for exactly one } x_i. \tag{26}$$

Note that in this case $P_{Y|X}(y_j|x_i)$ can still take a zero value for some y_j and that $\phi_i(y_j) = 0$ for any non-zero $P_{Y|X}(y_j|x_i)$. Thus, the error probability is equal to zero $P_{err}(x_i) = 0$; for all x_i, as well as the average probability of error $\bar{P}_{err} = 0$, which means that perfect detection is possible by means of a maximum likelihood detector.

Noisy channel with non-overlapping outputs: A simple example of a perfect transmission channel is the noisy channel with non-overlapping outputs (NOC), which is schematically described in Figure 1. It is a 2-input $m = 2k$-output channel ($k \in \mathbb{N}$) defined by the transition matrix:

$$P_{Y|X} = \begin{bmatrix} P_{Y|X}(\cdot|x_1) \\ P_{Y|X}(\cdot|x_2) \end{bmatrix} = \begin{bmatrix} \frac{1}{k} & \ldots & \frac{1}{k} & 0 & \ldots & 0 \\ 0 & \ldots & 0 & \frac{1}{k} & \ldots & \frac{1}{k} \end{bmatrix} \tag{27}$$

(in this and in the following matrices, the symbol "$\cdots$" stands for the k-time repletion). In the case of $k = 1$ and $m = 2k = 2$, the channel reduces to the noiseless channel. Although the channel is noisy, the input can always be recovered from the output (if y_j is received and $j \leq k$, the input symbol x_1 is sent, otherwise x_2 is sent). Thus, it is expected that the

information which is passed through the channel is equal to the information that can be generated by the input. Note that for a channel input distributed in accordance with

$$P_X = \begin{bmatrix} P_X(x_1) \\ P_X(x_2) \end{bmatrix} = \begin{bmatrix} a \\ 1-a \end{bmatrix}; \quad 0 \leq a \leq 1, \tag{28}$$

the joint probability distribution $P_{X,Y}$ can be expressed as in:

$$P_{X,Y} = \begin{bmatrix} \frac{a}{k} & \dots & \frac{a}{k} & 0 & \dots & 0 \\ 0 & \dots & 0 & \frac{1-a}{k} & \dots & \frac{1-a}{k} \end{bmatrix} \tag{29}$$

and the output distribution P_Y, which can be obtained by the summations over columns, is

$$P_Y = [P_Y(y_1), \dots, P_Y(y_m)]^T = \left[\frac{a}{k}, \dots, \frac{a}{k}, \frac{1-a}{k}, \dots, \frac{1-a}{k}\right]^T. \tag{30}$$

Binary symmetric channels: The binary symmetric channel (BSC) is a two input two output channel described by the transition matrix

$$P_{Y|X} = \begin{bmatrix} P_{Y|X}(\cdot|x_1)^T \\ P_{Y|X}(\cdot|x_2)^T \end{bmatrix} = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}, \tag{31}$$

which is schematically described in Figure 2. Note that for $p = \frac{1}{2}$ BSC reduces to a totally destructive channel, while in the case of $p = 0$ it reduces to a perfect channel.

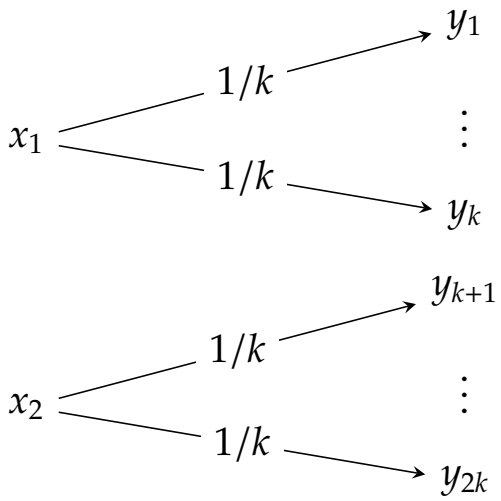

Figure 1. Noisy channel with non-overlapping outputs.

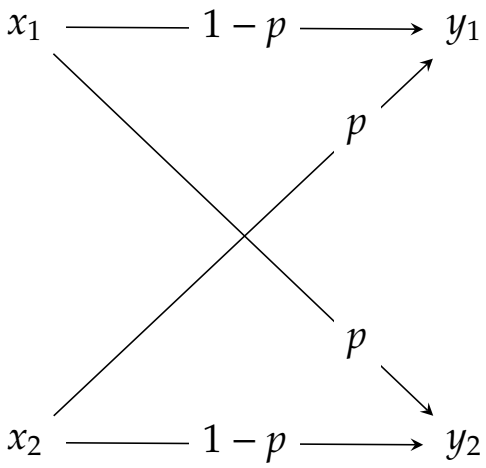

Figure 2. Binary symmetric channel.

Sharma–Mittal Information Transfer Axioms

In this paper, we search for information theoretical measures of information transfer between sender X and receiver Y, which communicate through a channel if the information is measured with Sharma–Mittal entropy. Thus, we are interested in the information transfer measure, $I_{\alpha,q}(X,Y)$, which is called the α-q-mutual information and its maximum,

$$C = \max_{P_X} I_{\alpha,q}(X,Y), \tag{32}$$

which is called the α-q-capacity and which requires the following set of axioms to be satisfied.

($\overline{A}_1$) The channel cannot convey negative information, i.e.,

$$C_{\alpha,q}(P_{Y|X}) \geq I_{\alpha,q}(X,Y) \geq 0. \tag{33}$$

($\overline{A}_2$) The information transfer is zero in the case of a totally destructive channel, i.e.,

$$P_{Y|X}(y|x) = \frac{1}{m}, \text{ for all } x,y \quad \Rightarrow \quad I_{\alpha,q}(X,Y) = C_{\alpha,q}(P_{Y|X}) = 0, \tag{34}$$

which is consistent with the conclusion that the average probability of error is one, $\bar{P}_{err} = 1$, in the case of a totally destructive channel.

($\overline{A}_3$) In the case of perfect transmission, the information transfer is equal to the input information, i.e.,

$$X = Y \quad \Rightarrow \quad I_{\alpha,q}(X,Y) = H_{\alpha,q}(X), \quad C_{\alpha,q}(P_{Y|X}) = \mathrm{Log}_q n, \tag{35}$$

which is consistent with the conclusion that the average probability of error is zero, $\bar{P}_{err} = 0$, in the case of a perfect transmission channel, so that all the information from the input is conveyed.

($\overline{A}_4$) The channel cannot transfer more information than it is possible to be sent, i.e.,

$$I_{\alpha,q}(X,Y) \leq C_{\alpha,q}(P_{Y|X}) \leq \mathrm{Log}_q\, n, \tag{36}$$

which means that a channel cannot add additional information.

($\overline{A}_5$) The channel cannot transfer more information than it is possible to be received, i.e.,

$$I_{\alpha,q}(X,Y) \leq C_{\alpha,q}(P_{Y|X}) \leq \mathrm{Log}_q\, m, \tag{37}$$

which means that a channel cannot add additional information.

($\overline{A}_6$) Consistency with the Shannon case:

$$\lim_{q\to 1,\alpha\to 1} I_{\alpha,q}(X,Y) = I(X,Y), \quad \text{and} \quad \lim_{q\to 1,\alpha\to 1} C_{\alpha,q}(P_{Y|X}) = C(P_{Y|X}) \tag{38}$$

Thus, the axioms ($\overline{A}_2$) and ($\overline{A}_3$) ensure that the information measures are consistent with the maximum likelihood detection (19)–(21). On the other hand, the axioms ($\overline{A}_1$), ($\overline{A}_4$) and ($\overline{A}_5$), prevent a situation in which a physical system conveys information in spite of going through a completely destructive channel, or in which the negative information transfer is observed, indicating that the channel adds or removes information by itself, which could be treated as nonphysical behavior without an intuitive explanation. Finally, the property ($\overline{A}_6$) ensure that the information transfer measures can be considered as generalizations of corresponding Shannon measures. For these reasons, we assume that the satisfaction of the properties ($\overline{A}_1$)–($\overline{A}_5$) is mandatory for any reasonable definition of Sharma–Mittal information transfer measures.

4. The α-Mutual Information and the α-Capacity

One of the first proposals for the Rényi mutual information goes back to Arimoto [24], who considered the following definition of mutual information:

$$I_\alpha(X,Y) = \frac{\alpha}{1-\alpha}\log\left(\sum_y\left(\sum_x P_X^{(\alpha)}(x)P_{Y|X}^{\alpha}(y|x)\right)^{\frac{1}{\alpha}}\right), \tag{39}$$

where the escort distribution $P_{X^{(\alpha)}}$ is defined as in (10), and he also invented an iterative algorithm for the computation of the α-capacity [35], which is defined from the α-mutual information:

$$C_\alpha(P_{Y|X}) = \max_{P_X} I_\alpha(X,Y). \tag{40}$$

Notably, Arimoto's mutual information can equivalently be represented using the conditional Rényi entropy

$$R_\alpha(X|Y) = \frac{\alpha}{\alpha-1}\log_2\sum_y P_Y(y)\left(\sum_x P_{X|Y=y}(x)^\alpha\right)^{\frac{1}{\alpha}}, \tag{41}$$

as in

$$I_\alpha(X,Y) \equiv R_\alpha(X) - R_\alpha(X|Y), \tag{42}$$

which can be interpreted as the input uncertainty reduction after the output symbols are received and, in the case of $\alpha \to 1$, the previous definition reduces to the Shannon case. In addition, this measure is directly related to the famous Gallager exponent

$$E_0(\rho, P_X) = -\log\left(\sum_y\left(\sum_x P_X(x)P_{Y|X}^{\frac{1}{1+\rho}}(y|x)\right)^{1+\rho}\right), \tag{43}$$

which has been widely used to establish the upper bound of error probability in channel coded communication systems [36] via the relationship [29]

$$I_\alpha(X,Y) = \frac{\alpha}{1-\alpha}E_0\left(\frac{1}{\alpha}-1, P_X^{(\alpha)}\right). \tag{44}$$

In addition, in the case of $\alpha \to 1$, it reduces to

$$I_1(X,Y) = \lim_{\alpha\to 1} I_\alpha(X,Y) = I(X,Y), \tag{45}$$

where

$$I(X,Y) = \sum_{x,y} P_{X,Y}(x,y) \log \frac{P_{X,Y}(x,y)}{P_X(x)P_Y(y)} \tag{46}$$

stands for Shannon's mutual information [37].

The α-mutual information $I_\alpha(X,Y)$ and the α-capacity $C_\alpha(P_{Y_X})$ satisfy the axioms $(\overline{A}_1)$–$(\overline{A}_6)$ for $q = 1$ and $\alpha > 0$, as stated by the following theorem, which further justifies their usage as the measures of (maximal) information transfer.

Theorem 1. *The mutual information measures I_α and C_α satisfy the following set of properties:*

(A_1) *The channel cannot convey negative information, i.e.,*

$$C_\alpha(P_{Y|X}) \geq I_\alpha(X,Y) \geq 0. \tag{47}$$

(A_2) *The (maximal) information transfer is zero in the case of a totally destructive channel, i.e.,*

$$P_{Y|X}(y|x) = \frac{1}{m}, \text{ for all } x, y \quad \Rightarrow \quad I_\alpha(X,Y) = C_\alpha(P_{Y|X}) = 0. \tag{48}$$

(A_3) *In the case of perfect transmission, the (maximal) information transfer is equal to the (maximal) input information, i.e.,*

$$X = Y \quad \Rightarrow \quad I_\alpha(X,Y) = R_\alpha(X), \quad C_\alpha(P_{Y|X}) = \log n. \tag{49}$$

(A_4) *The channel cannot transfer more information than it is possible to be sent, i.e.,*

$$I_\alpha(X,Y) \leq C_\alpha(P_{Y|X}) \leq \log\ n; \tag{50}$$

(A_5) *The channel cannot transfer more information than it is possible to be received, i.e.,*

$$I_\alpha(X,Y) \leq C_\alpha(P_{Y|X}) \leq \log\ m. \tag{51}$$

(A_6) *Consistency with the Shannon case:*

$$\lim_{\alpha\to 1} I_\alpha(X,Y) = I(X,Y), \quad \textit{and} \quad \lim_{\alpha\to 1} C_\alpha(P_{Y|X}) = C(P_{Y|X}) \tag{52}$$

Proof. As shown in [38], $R_\alpha(X|Y) \leq R_\alpha(X)$, and the nonnegativity property (A_1) follows from the definition of Arimoto's mutual information (42). In addition, if $X \perp\!\!\!\perp Y$, then $P_{Y|X}(y|x) = P_Y(y)$ so that the definition (61) implies the property (A_2). Furthermore, in the case of a perfect transmission channel, the mutual information (61) can be represented in

$$I_\alpha(X,Y) = \frac{\alpha}{\alpha-1} \log \frac{\sum_y \left(\sum_x P_X(x)^\alpha P_{Y|X}^\alpha(y|x)\right)^{\frac{1}{\alpha}}}{\left(\sum_x P_X^{(\alpha)}(x)\right)^{\frac{1}{\alpha}}} = \frac{\alpha}{\alpha-1} \log \frac{\sum_y \left(P_X(d(y))^\alpha P_{Y|X}^\alpha(y \mid d(y))\right)^{\frac{1}{\alpha}}}{\left(\sum_x P_X^{(\alpha)}(x)\right)^{\frac{1}{\alpha}}}, \tag{53}$$

and since

$$\begin{aligned} \sum_y \left(P_X(d(y))^\alpha P_{Y|X}^\alpha(y \mid d(y))\right)^{\frac{1}{\alpha}} &= \sum_y P_X(d(y)) P_{Y|X}(y \mid d(y)) = \\ \sum_x \sum_{y:d(y)=x} P_X(d(y)) P_{Y|X}(y \mid d(y)) &= \sum_x P_X(x) \sum_{y:d(y)=x} P_{Y|X}(y|x) = 1, \end{aligned} \tag{54}$$

we obtain $I_\alpha(X,Y) = R_\alpha(X)$, which proves the property (A_3). Moreover, from the definition as shown in [38], Arimoto's conditional entropy is positive and satisfies the weak chain rule $R_\alpha(X|Y) \geq R_\alpha(X) - \log m$, so that the properties (A_4) and (A_5) follow from the definition of Arimoto's mutual information (42). Finally, the property (A_6) follows directly from the

equation (45) and can be approved using L'Hôpital's rule, which completes the proof of the theorem. □

5. Alternative Definitions of the α-Mutual Information and the α-Channel Capacity

Since Rényi's proposal, there have been several lines of research to find an appropriate definition and characterization of information transfer measures related to Rényi entropy, which are established by the substitution of the Rényi divergence measure

$$D_\alpha(P||Q) = \frac{1}{\alpha - 1} \log \left(\sum_x P(x)^\alpha Q(x)^{1-\alpha} \right), \tag{55}$$

instead of the Kullback–Leibler one,

$$D(P||Q) = D_1(P||Q) = \sum_x P(x) \log \frac{P(x)}{Q(x)}, \tag{56}$$

in some of the various definitions which are equivalent in the case of Shannon information measures (46) [29]:

$$\begin{aligned} I(X,Y) &= \min_{Q_Y} \mathbb{E}\left[D_\alpha \left(P_{Y|X} \| Q_Y \right) \right] = \min_{Q_Y} \mathbb{E}\left[D_\alpha \left(P_{Y|X} \| Q_Y \right) \right] \\ &= \min_{Q_X} \min_{Q_Y} D_\alpha(P_{X,Y} \| Q_X Q_Y) = D_\alpha(P_{X,Y} \| P_X P_Y) = S(X) - S(X|Y) \end{aligned} \tag{57}$$

where $S(X|Y)$ stands for the Shannon conditional entropy,

$$S(X|Y) = \sum_{x,y} P_{X,Y}(x,y) \log P_{X|Y}(x|y). \tag{58}$$

All of these measures are consistent with the Shannon case in view of the property (A_6), but their direct usage as measures of Rényi information transfer leads to a breaking of some the properties (A_1)–(A_5), which justifies the usage of Arimoto's measures from the previous section as appropriate ones in the context of this research. In the following section, we review the alternative definitions.

5.1. Information Transfer Measures by Sibson

Alternative approaches based on Rényi divergence were proposed by Sibson [23] and considered later by several authors in the context of quantum secure communications [39–44], who introduced

$$J_\alpha^1(X;Y) = \min_{Q_Y} D_\alpha \left(P_{Y|X} P_X \| Q_Y P_X \right), \tag{59}$$

which can be represented as in [26]

$$J_\alpha^1(X,Y) = \frac{\alpha}{\alpha - 1} \log \left(\sum_y \left(\sum_x P_X(x) P_{Y|X}^\alpha(y|x) \right)^{\frac{1}{\alpha}} \right) \tag{60}$$

and, in the discrete setting, can be related to the Gallager exponent as in [29]:

$$J_\alpha^1(X,Y) = \frac{\alpha}{1-\alpha} E_0 \left(\frac{1}{\alpha} - 1, P_X \right), \tag{61}$$

which differs from Arimoto's definition (61) since in this case the escort distribution does not participate in the error exponent, but an ordinary one does. However, in the case of a perfect channel for which $X = Y$, the conditional distribution $P_{Y|X}^\alpha(y|x) = 1$ for $x = y$ and zero otherwise, so Sibson's measure (60) reduces to $R_{1/\alpha}(X)$, thus breaking the axiom (A_3). This disadvantage can be overcome by the reparametrization $\alpha \leftrightarrow 1/\alpha$ so that $J_{1/\alpha}^1(X,Y)$

is used as a measure of Rényi information transfer, and the properties of the resulting measure can be considered in a manner similar to the case of Arimoto.

5.2. Information Transfer Measures by Augustin and Csiszar

An alternative definition of Rényi mutual information was also presented by Augustin [25], and later Csiszar [26], who defined

$$J_\alpha^2(X;Y) = \min_{Q_Y} \mathbb{E}\Big[D_\alpha\Big(P_{Y|X}\|Q_Y\Big)\Big], \tag{62}$$

However, in the case of perfect transmission, for which $X = Y$, the measure reduces to Shannon entropy

$$J_\alpha^2(X;Y) = S(X), \tag{63}$$

which breaks the axiom (A_3).

5.3. Information Transfer Measures by Lapidoth, Pfister, Tomamichel and Hayashi

A similar obstacle to the case of the Augustin–Csiszar measure can be observed in the case of mutual information which was considered by Lapidoth and Pfister [27] and Tomamichel and Hayashi [28], who proposed

$$J_\alpha^3(X;Y) = \min_{Q_X} \min_{Q_Y} D_\alpha(P_{X,Y}\|Q_X Q_Y). \tag{64}$$

As shown in [27] (Lemma 11), if $X = Y$, then

$$J_\alpha^3(X;Y) = \begin{cases} \frac{\alpha}{1-\alpha} \lim_{\alpha\to\infty} R_\alpha(X) & \text{if } \alpha \in \left[0, \frac{1}{2}\right], \\ R_{\frac{\alpha}{2\alpha-1}}(X) & \text{if } \alpha > \frac{1}{2} \end{cases} \tag{65}$$

so the axiom (A_3) is broken in this case, as well.

Remark 1. *Despite the difference between the definitions of information transfer, in the discrete setting, the alternative definitions discussed above reach the same maximum over the set of input probability distributions, P_X, [26,29,45].*

5.4. Information Transfer Measures by Chapeau-Blondeau, Delahaies, Rousseau, Tridenski, Zamir, Ingber and Harremoes

Chapeau-Blondeau, Delahaies and Rousseau [31], and independently Tridenski, Zamir and Ingber [46] and Harremoes [47], defined the Rényi mutual information using the Rényi divergence (55), so that the mutual information defined using the Rényi divergence

$$J_\alpha^4(X,Y) = D_\alpha(P_{X,Y}\|P_X P_Y) \tag{66}$$

for $\alpha > 0$ and $\alpha \neq 1$, while in the case of $\alpha = 1$ it reduces to Shannon mutual information. However, the ordinal definition can correspond only to a Rényi entropy of order $2 - \alpha$ since in the case of $X = Y$ it reduces to $J_\alpha^4(X,Y) = R_{2-\alpha}(X)$ (see also [47]), which can be overcome by the reparametrization $\alpha = 2 - q$, similar to the case of Sibson's measure. This measure has been discussed in the past with various operational characterizations, and could also be considered as a measure of information transfer, although the satisfaction of all of the axioms (A_1)–(A_6) is not self-evident for general channels.

5.5. Information Transfer Measures by Jizba, Kleinert and Shefaat

Finally, we will mention the definition by Jizba, Kleinert and Shefaat [48],

$$J_\alpha^4(X,Y) \equiv R_\alpha(X) - \hat{R}_\alpha(X|Y), \tag{67}$$

which is defined in the same manner as in Arimoto's case (42), but with another choice of conditional Rényi entropy

$$\hat{R}_\alpha(X|Y) = \frac{1}{1-\alpha}\log\sum_x P_X^{(\alpha)}(x)2^{(1-\alpha)R_\alpha(X|Y=y)}, \tag{68}$$

which arises from the Generalized Shannon–Khinchin axiom [GSK4] if the pseudo-additivity in the equation (9) is restricted to an ordinary addition, in which case the GSK axioms uniquely determine Rényi entropy [49]. However, despite its wide applicability in the modeling of causality and financial time series, this mutual information can take negative values which breaks the axiom (A_1), which is assumed to be mandatory in this paper. For further discussion of the physicalism of negative mutual information in the domain of financial time series analysis, the reader is referred to [48].

6. The α-q Mutual Information and the α-q-Capacity

In the past several attempts have been done to define an appropriate channel capacity measure which corresponds to instances of the Sharma–Mittal entropy class. All of them follow a similar recipe by which the channel capacity is defined as in (32), as a maximum of appropriately defined mutual information $I_{\alpha,q}$. However, all of the classes consider only special cases of Sharma–Mittal entropy and all of them fail to satisfy at least one of the properties $(\overline{A}_1)$–$(\overline{A}_6)$ which an information transfer has to satisfy, as will be discussed Section 7.

In this section we propose a general measures of the α-q mutual information and the α-q-capacity by the requirement that the axioms $(\overline{A}_1)$–$(\overline{A}_6)$ are satisfied, which could qualify them as appropriate measures of information transfer, without nonphysical properties. The special instances of the α-q (maximal) information transfer measures are also discussed and the analytic expressions for a binary symmetric channel are provided.

6.1. The α-q Information Transfer Measures and Its Instances

The α-q-mutual information (42) is defined using the q-subtraction defined in (6), as follows:

$$I_{\alpha,q}(X,Y) = H_{\alpha,q}(X) \ominus_q H_{\alpha,q}(X|Y), \tag{69}$$

where we introduced the conditional Sharma–Mittal entropy $H_{\alpha,q}(Y|X)$ as in

$$H_{\alpha,q}(X|Y) = \eta_q(R_\alpha(X|Y)) = \frac{1}{(1-q)\ln 2}\left(\left(\sum_y P_Y(y)\left(\sum_x P_{X|Y=y}(x)^\alpha\right)^{\frac{1}{\alpha}}\right)^{\frac{\alpha(1-q)}{\alpha-1}} - 1\right), \tag{70}$$

$R_\alpha(X|Y)$ stands for Arimoto's definition of the conditional Rényi entropy (41). The expression (69) can also be obtained if the mapping η_q is applied to both sides of the equality (42), by which Arimoto's mutual information is defined, so we may establish the relationship

$$I_{\alpha,q}(X,Y) = \eta_q(I_\alpha(X,Y)) = \eta_q\left(\frac{\alpha}{1-\alpha}\log\left(\sum_y\left(\sum_x P_X^{(\alpha)}(x)P_{Y|X}^\alpha(y|x)\right)^{\frac{1}{\alpha}}\right)\right), \tag{71}$$

which can be represented using the Gallager error exponent (43) as in

$$I_{\alpha,q}(X,Y) = \eta_q\left(\frac{\alpha}{1-\alpha}E_0\left(\frac{1}{\alpha}-1, P_X^{(\alpha)}\right)\right) = \frac{1}{(1-q)\ln 2}\left(2^{\frac{\alpha(1-q)}{1-\alpha}E_0\left(\frac{1}{\alpha}-1,P_X^{(\alpha)}\right)} - 1\right). \tag{72}$$

Arimoto's α-q-capacity is now defined in

$$C_{\alpha,q} = \max_{P_X} I_{\alpha,q}(X,Y), \tag{73}$$

and using the fact that η_q is increasing, it can be related with the corresponding α-capacity as in

$$C_{\alpha,q} = \max_{P_X} I_{\alpha,q}(X,Y) = \max_{P_X} \eta_q(I_\alpha(X,Y)) = \eta_q\left(\max_{P_X} I_\alpha(X,Y)\right) = \eta_q\Big(C_\alpha(P_{Y|X})\Big). \quad (74)$$

Using the expressions (45) and (71), in the case of $\alpha = 1$, the α-q mutual information reduces to

$$\begin{aligned} I_{1,q} &= \frac{1}{(1-q)\ln 2}\left(\prod_{x,y} 2^{P_{X,Y}(x,y)\log\frac{P_{X,Y}(x,y)}{P_X(x)P_Y(y)}} - 1\right) \\ &= \frac{1}{(1-q)\ln 2}\left(\prod_{x,y}\left(\frac{P_{X,Y}(x,y)}{P_X(x)P_Y(y)}\right)^{P_{X,Y}(x,y)} - 1\right). \end{aligned} \quad (75)$$

The α-q-capacity is given in

$$C_{1,q} = \max_{P_X}\left(\frac{1}{(1-q)\ln 2}\left(\prod_{x,y}\left(\frac{P_{X,Y}(x,y)}{P_X(x)P_Y(y)}\right)^{P_{X,Y}(x,y)} - 1\right)\right) \quad (76)$$

and these measures can serve as (maximal) information transfer measures corresponding to Gaussian entropy, which was not considered before in the context of information transmission. Naturally, if in addition $q \to 1$, the measures reduce to Shannon's mutual information and Shannon capacity [37].

Additional special cases of the α-q (maximal) information transfer include the α-mutual information (42) and the α-capacity (40), which are obtained for $q = 1$; the measures which correspond to Tsallis entropy can be obtained for $q = \alpha$ and the ones which correspond to Landsberg–Vedral entropy for $q = 2 - \alpha$. These special instances are listed in Table 1.

As discussed in Section 7, previously considered information measures cover only particular special cases and break at least one of the axioms $(\overline{A}_1)$–$(\overline{A}_5)$, which leads to unexpected and counterintuitive conclusions about the channels, such as negative information transfer and achieving super-capacitance or sub-capacitance [4], which could be treated as a nonphysical behavior. On the other hand, apart from the generality, the α-q information transfer measures proposed in this paper overcame the disadvantages which could qualify them as appropriate measures, as stated in the following theorem.

Theorem 2. *The α-q information transfer measures $I_{\alpha,q}$ and $C_{\alpha,q}$ satisfy the set of the axioms $(\overline{A}_1)$–$(\overline{A}_6)$.*

Proof. The proof is the straightforward application of the mapping η_q to the equations in the α-mutual information properties (A_1)–(A_5), while the $(\overline{A}_6)$ follows from the above discussion. □

Remark 2. *Note that the symmetry $I_{\alpha,q}(X,Y) = I_{\alpha,q}(Y,X)$ does not hold in general in the case of the α-q mutual information nor in the case of the α mutual information [50,51] and if the mutual information is defined so that the symmetry is preserved, some of the axioms $(\overline{A}_1)$–$(\overline{A}_6)$ might be broken. In addition, the alternative definition of the mutual information, $I_{\alpha,q}(Y,X) = H_{\alpha,q}(Y) - H_{\alpha,q}(Y|X)$, which uses an ordinary substraction operator instead of $\ominus_q$ operation, can also be introduced, but in this case the property $(\overline{A}_5)$ might not hold in general, as discussed in Section 7.*

6.2. The α-q-Capacity of Binary Symmetric Channels

As shown by Cai and Verdú [45], the α-mutual information of Arimoto's type I_α is maximized for the uniform distribution $P_X = (1/2, 1/2)$, and Arimoto's α-capacity has the value

$$C_\alpha(BSC) = 1 - r_\alpha(p), \quad (77)$$

where the binary entropy function r_α is defined as

$$r_\alpha(p) = R_\alpha(p, 1-p) = \frac{1}{1-\alpha}\log(p^\alpha + (1-p)^\alpha), \tag{78}$$

for $\alpha > 0$, $\alpha \neq 1$, while in the limit of $\alpha \to 1$, the expression (78) reduces to the well-known result for the Shannon capacity (see Fano [52])

$$C_1(BSC) = \lim_{\alpha\to 1} C_\alpha(BSC) = 1 + p\log p + (1-p)\log(1-p). \tag{79}$$

The analytic expressions for the α-q-capacities of binary symmetric channel's can be obtained from the expressions (74) and (77), so that

$$C_{\alpha,q}(BSC) = \eta_q(C_\alpha(BSC)) = \frac{1}{(1-q)\ln 2}\left(2^{1-q}(p^\alpha + (1-p)^\alpha)^{-\frac{1-q}{1-\alpha}} - 1\right); \tag{80}$$

in the case of $q = 1$, it reduces to the case of Rényi entropy while, in the case of $\alpha = 1$, to the case of Gaussian entropy (77)

$$C_{1,q}(BSC) = \frac{1}{(1-q)\ln 2}\left(2p^p(1-p)^{1-p} - 1\right). \tag{81}$$

The analytic expressions for BSC α-q capacities for other instances can straightforwardly be obtained by specifying the values of the parameters, whose instances are listed in Table 1, while the plots of the BSC α-q-capacities, which correspond to the Gaussian and the Tsallis entropies, are shown in Figures 3 and 4.

The α-q-capacity (80) can equivalently be expressed in

$$C_{\alpha,q}(BSC) = \mathrm{Log}_q\, 2 \ominus_q h_{\alpha,q}(p), \tag{82}$$

where the Sharma–Mittal binary entropy function is defined in

$$h_{\alpha,q}(p) = H_{\alpha,q}(p, 1-p) = \frac{1}{1-q}\left((p^\alpha + (1-p)^\alpha)^{\frac{1-q}{1-\alpha}} - 1\right), \tag{83}$$

which reduces to the Rényi binary entropy function, in the case of $q = 1$,

$$h_{\alpha,1}(p) = \lim_{q\to 1} h_{\alpha,q}(p) = R_\alpha(p, 1-p) = \frac{1}{1-\alpha}\log(p^\alpha + (1-p)^\alpha)), \tag{84}$$

to the Tsallis binary entropy function, in the case of $\alpha = 1$,

$$h_{q,q}(p) = h_{q,q}(p) = T_q(p, 1-p) = \frac{1}{1-q}(p^q + (1-p)^q - 1), \tag{85}$$

to the Gaussian binary entropy function, in the case of $\alpha = 1$,

$$h_{1,q}(p) = \lim_{\alpha\to 1} h_{\alpha,q}(p) = G_q(p, 1-p) = \frac{1}{(1-q)\ln 2}\left(p^{-(1-q)p}(1-p)^{-(1-q)(1-p)} - 1\right), \tag{86}$$

and to the Shannon binary entropy function, in the case of $\alpha = q = 1$,

$$h_{1,1}(p) = \lim_{q,\alpha\to 1} h_{\alpha,q}(p) = S(p, 1-p) = -p\log p - (1-p)\log(1-p). \tag{87}$$

The expression (82) can be interpreted similarly as in the Shannon case. Thus, a BSC channel with input X and output Y can be modeled with an input–output relation $Y = X \oplus Z$ where $\oplus$ stands for modulo 2 sum and Z is channel noise taking values from $\{1, 0\}$, distributed in accordance with $(p, 1-p)$. If we measure the information which is lost per bit during transmission with the Sharma–Mittal entropy $H_{\alpha,q}(Z) = h_\alpha(p)$, then $C_{\alpha,q}$ stands for useful information left over for every bit of information received.

Table 1. Instances of the α-q-mutual information for different values of the parameters and corresponding expressions for the BSC α-q-capacities.

$H_{\alpha,q}$	$I_{\alpha,q}$	$C_{\alpha,q}$
S $\alpha = q = 1$	$\sum_{x,y} P_{X,Y}(x,y) \log \frac{P_{X,Y}(x,y)}{P_X(x)P_Y(y)}$	$1 + p \log p + (1-p)\log(1-p)$
R_α $q = 1$	$\frac{\alpha}{1-\alpha} E_0\left(\frac{1}{\alpha} - 1, P_X^{(\alpha)}\right)$	$1 - \frac{\log(p^\alpha + (1-p)^\alpha)}{1-\alpha}$
T_q $q = \alpha$	$\frac{1}{(1-q)\ln 2}\left(2^{qE_0\left(\frac{1}{q}-1, P_X^{(q)}\right)} - 1\right)$	$\frac{1}{(1-q)\ln 2}\left(2^{1-q}(p^q + (1-p)^q)^{-1} - 1\right)$
L_α $q = 2 - \alpha$	$\frac{1}{(\alpha-1)\ln 2}\left(2^{-\alpha E_0\left(\frac{1}{\alpha}-1, P_X^{(\alpha)}\right)} - 1\right)$	$\frac{1}{(1-\alpha)\ln 2}\left(2^{\alpha-1}(p^\alpha + (1-p)^\alpha) - 1\right)$
G_q $\alpha = 1$	$\frac{1}{(1-q)\ln 2}\left(\prod_{x,y}\left(\frac{P_{X,Y}(x,y)}{P_X(x)P_Y(y)}\right)^{P_{X,Y}(x,y)} - 1\right)$	$\frac{1}{(1-q)\ln 2}\left(2^{1-q}p^{(1-q)p}(1-p)^{(1-q)(1-p)} - 1\right)$
$E_0(\rho, P_X) = -\log\left(\sum_y \left(\sum_x P_X(x) P_{Y\|X}^{\frac{1}{1+\rho}}(y\|x)\right)^{1+\rho}\right)$		

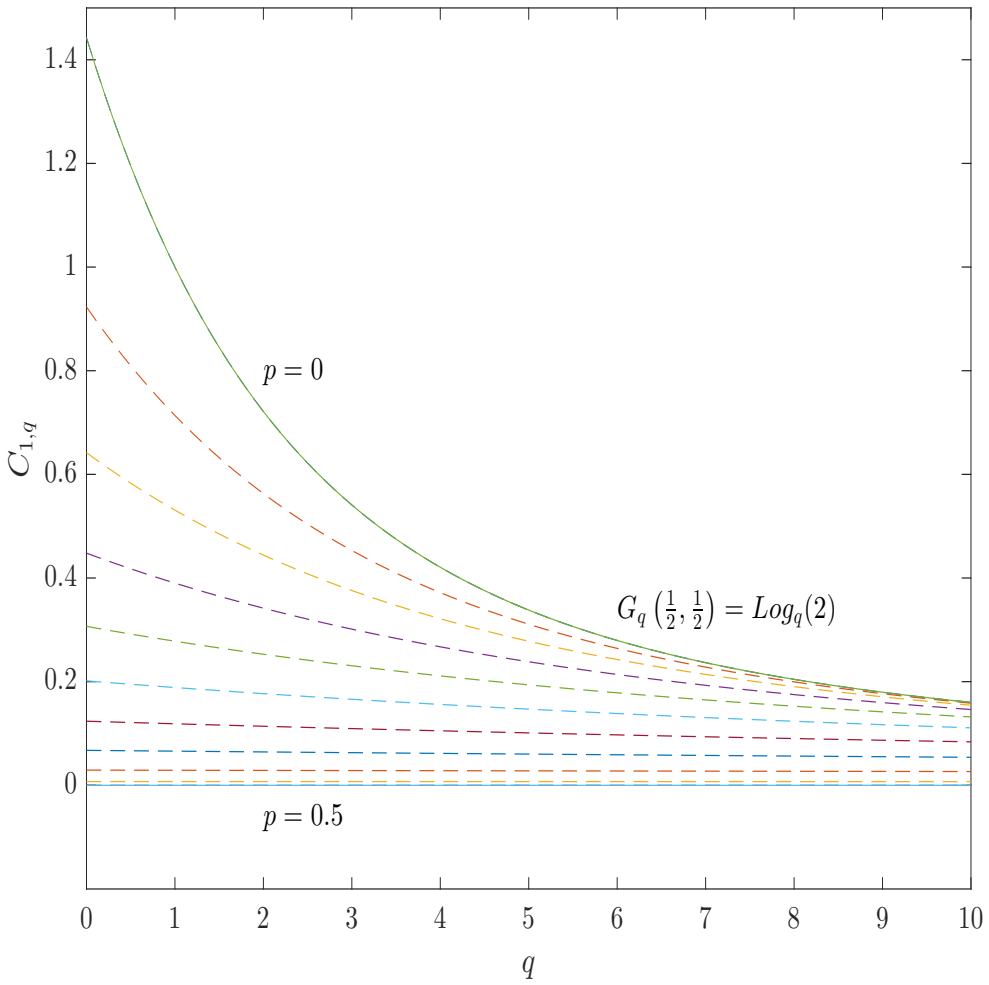

Figure 3. The α-q-capacity of BSC for the Gaussian entropy (the case of $\alpha = 1$) as a function of q for various values of the channel parameter p from 0.5 (totally destructive channel) to 0 (perfect transmission). All of the curves lies between 0 and $\mathrm{Log}_q 2$, which is the maximum value of the Gaussian entropy.

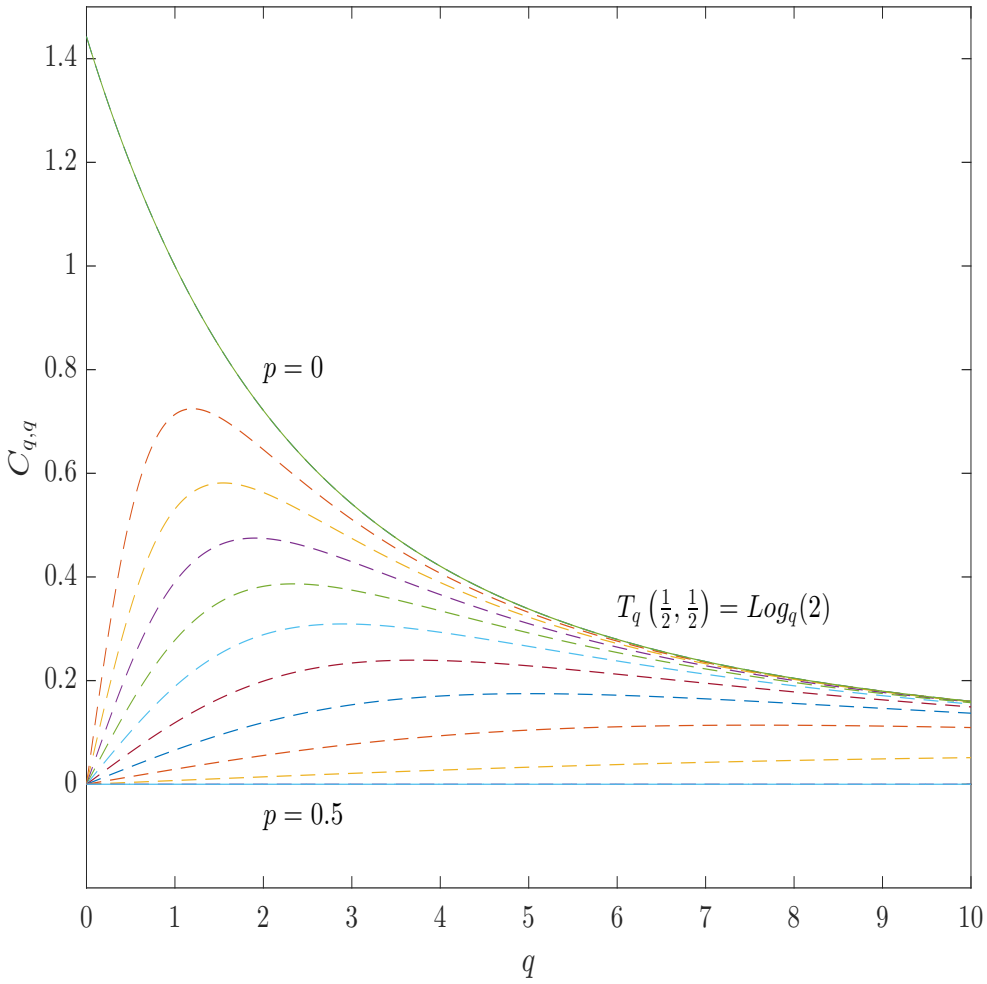

Figure 4. The α-q-capacity of BSC for the Tsallis entropy (the case of $\alpha = q$) as a function of q for various values of the channel parameter p from 0.5 (totally destructive channel) to 0 (perfect transmission). All of the curves lies between 0 and $\mathrm{Log}_q 2$, which is the maximum value of the Tsallis entropy.

7. An Overview of the Previous Approaches to Sharma–Mittal Information Transfer Measures

In this section, we review the previous attempts at a definition of Sharma–Mittal information transfer measures, which are defined from the basic requirement of consistency with the Shannon measure as given by the axiom $(\overline{A}_6)$. However, as we show in the following paragraphs, all of them break at least one of the axioms $(\overline{A}_1)$–$(\overline{A}_5)$, which are satisfied in the case of the α-q (maximal) information transfer measures (69) and (73), in accordance with the discussion in Section 6.

7.1. Daróczy's Capacity

The first considerations of generalized channel capacities and generalized mutual information for the q-entropy go back to Daróczy [30], who introduced conditional Tsallis entropy

$$\bar{T}_q(Y|X) = \sum_x P_X^q(x) T_q(Y|X = x), \tag{88}$$

where the row entropies are defined as in

$$T_q(Y|X = x) = \frac{1}{(1-q)\log(2)} \left(\sum_x P_{Y|X}(y|x)^q - 1 \right) \tag{89}$$

and the mutual information is defined as in

$$J^5_{\alpha,q}(X,Y) = T_q(Y) - \bar{T}_q(Y|X). \tag{90}$$

However, in the case of a totally destructive channel, $X \perp\!\!\!\perp Y$, $P_{Y|X}(y|x) = P_Y(y)$, $T_q(Y|X=x) = T_q(Y)$ and

$$T_q(Y|X) = T_q(Y)\sum_x P_X(x)^q \tag{91}$$

so that

$$J^5_{\alpha,q}(X,Y) = T_q(Y)\left(1-\sum_x P_X(x)^q\right) = \left(1-\sum_x P_X(x)^q\right)\mathrm{Log}_q m. \tag{92}$$

This expression is zero for an input probability distribution $P_X = (1,0,\ldots,0)$ and its permutations, but, in general, it is negative for $q<1$, positive for $q>1$ and 0 only for $q=1$, so the axiom $(\overline{A}_2)$ is broken (see Figure 5). As a result, the channel capacity, which is defined in accordance to (32), is zero for $q \le 1$ and positive for $q>1$, as illustrated in Figure 6 by the example of BSC for which the Daroczy's channel capacity can be computed as in [30,53]

$$C^5_q(BSC) = \frac{1-2^{1-q}}{q-1} - \frac{2^{-q}}{q-1}[1-(1-p)^q - p^q]. \tag{93}$$

In the same figure, we plotted the graph for the α-q channel capacities proposed in this paper, and all of them remain zero in the case of a totally destructive BSC, as expected.

7.2. Yamano Capacities

Similar problems to the ones mentioned above arise in the case of mutual information and corresponding capacity measures considered by Yamano [33], who addressed the information transmission characterized by Landsberg–Vedral entropy L_q, given in (17).

Thus, the first proposal is based on the mutual information of the form

$$J^6_q(X,Y) = L_q(X) + L_q(Y) - L_q(X,Y), \tag{94}$$

where the joint entropy is defined in

$$L_q(X,Y) = \frac{1}{q-1}\left(\frac{1}{\sum_{x,y}P_{X,Y}(x,y)^q} - 1\right). \tag{95}$$

However, in the case of a fully destructive channel, $P_Y(y) = 1/m$ and $P_{X,Y}(x,y) = P_X(x)/m$, so that

$$J^6_q(X,Y) = \frac{1}{q-1}\left(\frac{1}{\sum_x P_X(x)^q} - 1\right) + \frac{1}{q-1}\left(m^{q-1}-1\right) - \frac{1}{q-1}\left(m^{q-1}\frac{1}{\sum_x P_X(x)^q} - 1\right), \tag{96}$$

which can be simplified to

$$J^6_q(X,Y) = \frac{1-m^{q-1}}{q-1}\left(\frac{1}{\sum_x P_X(x)^q} - 1\right). \tag{97}$$

Similarly to the case of Daroczy's capacity, this expression is zero for an input probability distribution $P_X = (1,0,\ldots,0)$ and its permutations but, in general, it is negative for $q>1$, positive for $q<1$ and 0 only for $q=1$, so the axiom $(\overline{A}_2)$ is broken (see Figure 5). In Figure 6 we illustrated the Yamano channel capacity as a function of the parameter q, in the case of two input channels with $P_X = [a, 1-a]$, the channel capacity is zero for $q>1$ (which is obtained for $P_X = [1,0]$), and

$$C^6_q(BSC) = \frac{1}{q-1}\left(2^q - 1 - 2^{2q-2}\right), \tag{98}$$

for $q > 1$ (which is obtained for $P_X = [1/2, 1/2]$). In the same Figure, we plotted the graph for the α-q channel capacities proposed in this paper, and, as before, all of them remain zero in the case of a totally destructive BSC, as expected.

Further attempts were made in [33], where the mutual information is defined in an analogous manner to (66) and (66), with the generalized divergence measure introduced in [54]. Thus, the alternative measure for mutual information is defined in

$$J_q^7(X,Y) = \frac{1}{(1-q)\ln 2}\frac{1}{\sum_{x,y} P_{X,Y}^q(x,y)}\left[1 - \sum_{x,y} P_{X,Y}(x,y)\left(\frac{P_X(x)P_Y(y)}{P_{X,Y}(x,y)}\right)^{1-q}\right]. \tag{99}$$

However, in the case of the simplest perfect communication channel for which $X = Y$, the mutual information reduces to

$$J_q^7(X,Y) = \frac{1}{(1-q)\ln 2}\frac{1-\sum_x P_X(x)^{2-q}}{\sum_x P_X(x)^q} \neq L_q(X), \tag{100}$$

which breaks the axiom $(\overline{A}_3)$.

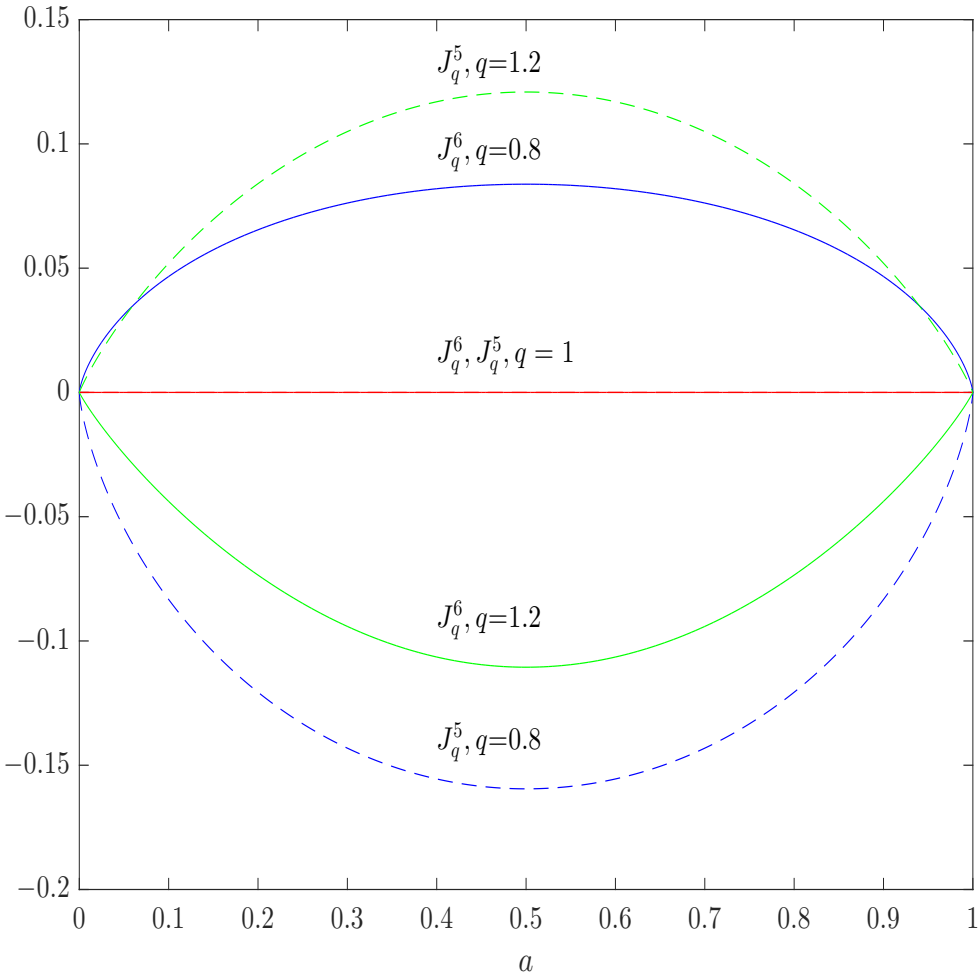

Figure 5. Daróczy's (solid lines) and Yamano's (dashed lines) mutual information in the case of a totally destructive BSC as functions of the input distribution parameter a, $P_X = [a, 1-a]^T$ for different values of q, obtaining negative values for $q < 1$ and $q > 1$, respectively, breaking the axioms $(\overline{A}_1)$ and $(\overline{A}_2)$. The α-q-mutual information is zero; for all q, and satisfies $(\overline{A}_1)$ and $(\overline{A}_2)$.

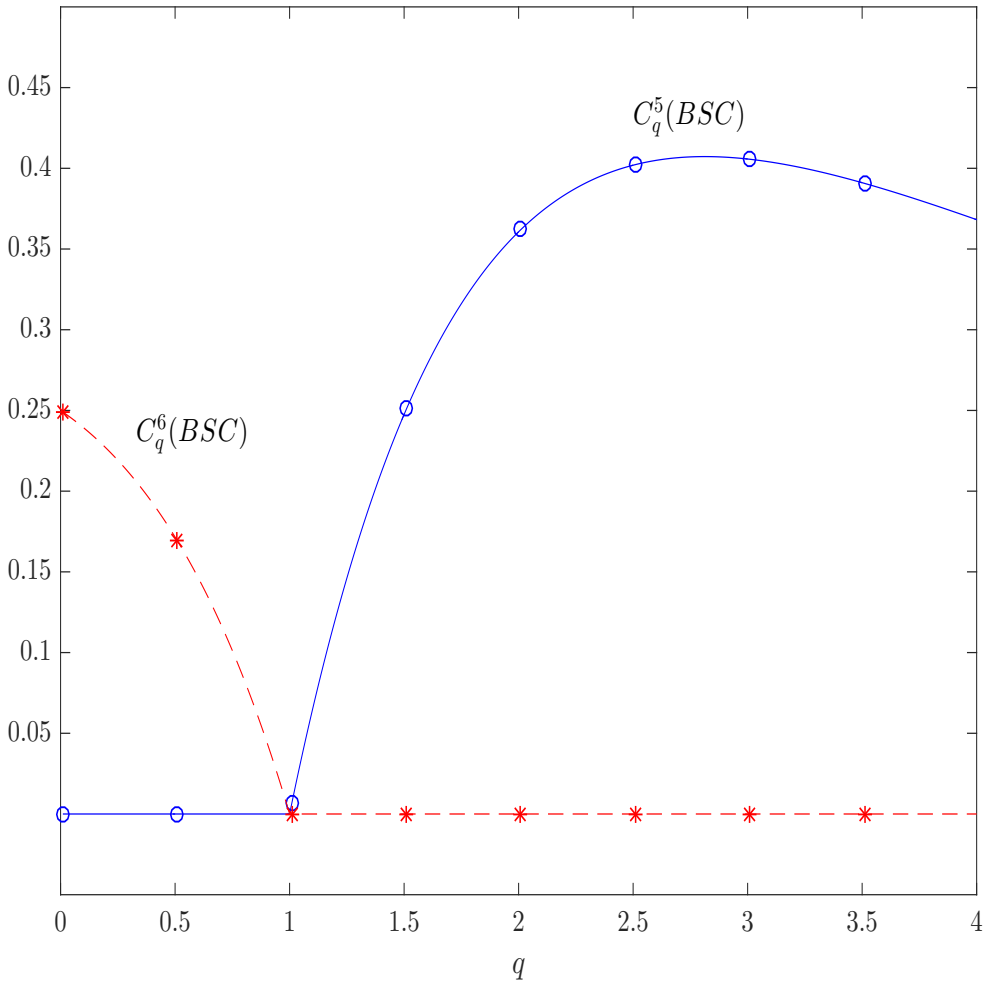

Figure 6. Daróczy's (solid lines) and Yamano's (dashed lines) capacities in the case of totally destructive BSC as functions of the parameter q. In the regions of $q < 1$ and $q > 1$, respectively, the corresponding negative mutual information is maximized for $P_X = [1, 0]^T$ (zero capacity) having the positive values outside the regions and breaking the axiom ($\overline{A}_2$). The α-q-capacity is zero; for all q, and satisfies ($\overline{A}_2$).

7.3. Landsber–Vedral capacities

To avoid these problems, Landsberg and Vedral [4] proposed the mutual information measure and related channel capacities for the Sharma–Mittal entropy class $H_{\alpha,q}$, particularly considering the choice of $q = \alpha$, which corresponds to Tsallis entropy, $q = 2 - \alpha$, and the case of $q = 1$, which corresponds to the Rényi entropy

$$J^8_{\alpha,q}(X, Y) = H_{\alpha,q}(Y) - \tilde{H_{\alpha,q}}(Y|X), \tag{101}$$

where the conditional entropy $\tilde{H_{\alpha,q}}^{LV}(Y|X)$ is defined as in

$$\tilde{H_{\alpha,q}}(Y|X) = \sum_x P_X(x) H_{\alpha,q}(Y|X = x) \tag{102}$$

and

$$H_{\alpha,q}(Y|X = x) = \frac{1}{1-q}\left(\left(\sum_y P_{Y|X}(y|x)^\alpha\right)^{\frac{1-q}{1-\alpha}} - 1\right). \tag{103}$$

Although this definition bears some similarities to the α-q mutual information proposed in formula (69), several key differences can be observed. First of all, it character-

izes the information transfer as the output uncertainty reduction after the input symbols are known, instead of input uncertainty reduction, after the output symbols are known (42). In addition, it uses the ordinary—operation instead of the $\ominus_q$ one. In addition, note that the definition of conditional entropy (102) generally differs from the definition proposed in (70).

The definition (101) resolves the issue of the axiom $(\overline{A}_2)$ which appears in the case of the Daroczy capacity, since in the case of a totally destructive channel $(X \perp\!\!\!\perp Y)$, $P_{Y|X}(y|x) = P_Y(y)$ and $L_q(Y|X = x) = L_q(Y)$ and $L_q(Y|X) = L_q(Y)$, so that $I^{Iv}_{\alpha,q}(X, Y) = 0$. However, the problems remain with the axiom $(\overline{A}_5)$, which can be observed in the case of a noisy channel with non-overlapping outputs if the number of channel inputs is lower than the number of channel outputs $n < m$. Indeed, in the case of a noisy channel with non-overlapping outputs given by the transition matrix (27), both of the row entropies $L_q(Y|X = x)$ have the same value, which is independent of x

$$H_{\alpha,q}(Y|X = x) = \frac{k^{1-q} - 1}{(q-1)\ln 2} = \mathrm{Log}_q k; \quad \text{for} \quad x = x_1, x_2, \tag{104}$$

and the maximal value of Landsberg–Vedral mutual information (101) is obtained only by maximizing $H_{\alpha,q}(Y)$ over P_X, which is achieved if X is uniformly distributed, since in this case Y is uniformly distributed, as well as ($a = \frac{1}{2}$ in (28)), so the maximal value of the output entropy is $H_{\alpha,q}(Y) = \mathrm{Log}_q(2k)$ and the mutual information is maximized for

$$C^8_{\alpha,q}(NOC) = \mathrm{Log}_q(2k) - \mathrm{Log}_q(k), \tag{105}$$

which is greater than $\mathrm{Log}_q(2)$ for $k \geq 2$, i.e., for $m \geq 4$ outputs, so the axiom $(\overline{A}_5)$ is broken, which is illustrated in Figure 7.

7.4. Chapeau-Blondeau–Delahaies–Rousseau Capacities

Following a similar approach to the one in Section 5.4, Chapeau-Blondeau, Delahaies and Rousseau considered the definition of mutual information which corresponds to the Tsallis entropy using Tsallis divergence,

$$D_{q,q}(P||Q) = \frac{1}{q-1}\left(\sum_x P(x)^q Q(x)^{1-q} - 1\right), \tag{106}$$

can be written in

$$\begin{aligned} J^9_q(X, Y) &= D_{q,q}(P_{X,Y}\|P_X P_Y) = \eta_q\big(D_q(P_{X,Y}\|P_X P_Y)\big) \\ &= \frac{1}{1-q}\left(1 - \sum_{x,y} P_{X,Y}(x, y)^q P_X(x)^{1-q} P_Y(y)^{1-q}\right). \end{aligned} \tag{107}$$

However, this definition is not directly applicable as a measure of information transfer to the Tsallis entropy with index q, since in the case of $X = Y$ it reduces to $J^9_q(X, Y) = T_{2-q}(X)$, and requires the reparametrization $q \leftrightarrow 2 - q$, similar to Section 5.4, while the satisfaction of the axioms $(\overline{A}_4)$ and $(\overline{A}_5)$ is not self evident.

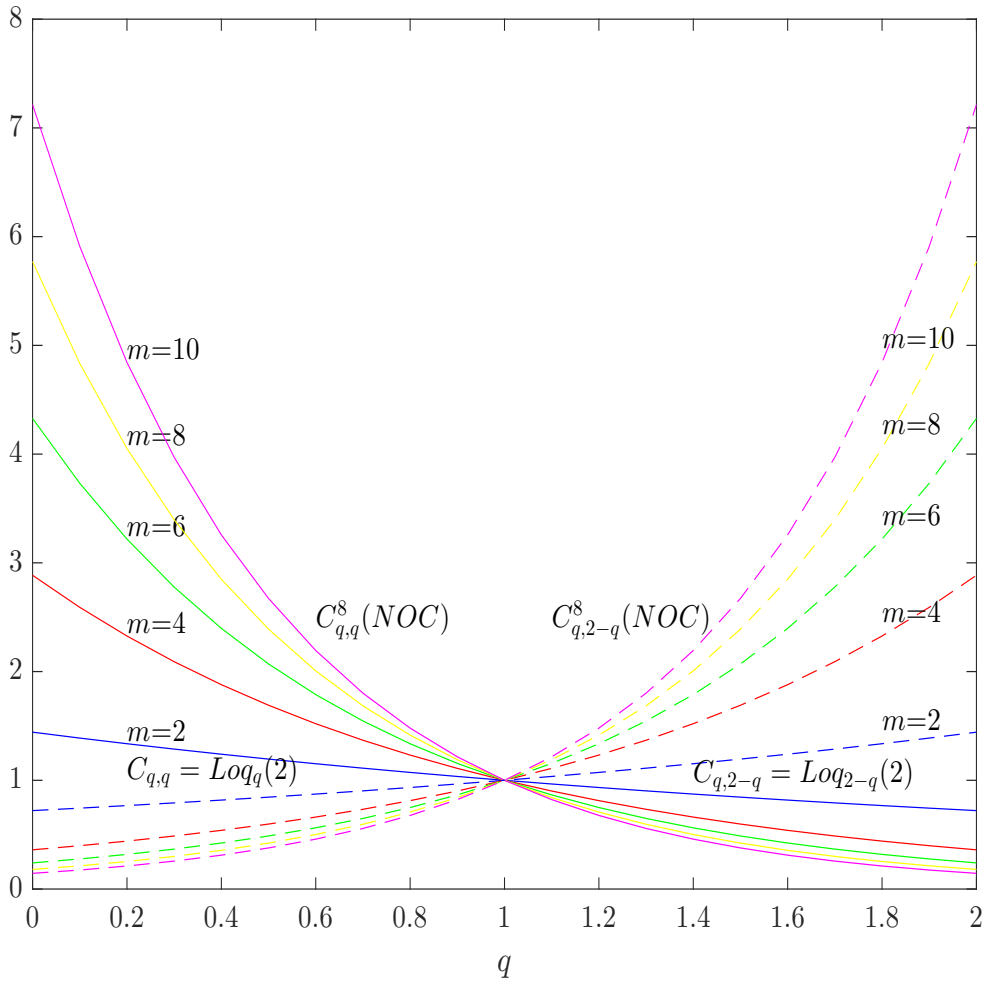

Figure 7. Landsberg–Vedral capacities for the Tsallis (solid lines) and the Landsberg–Vedral (dashed lines) entropies in the case of a (perfect) noisy channel with non-overlapping outputs with m outputs as functions of q, for different values of m. The axiom $(\overline{A}_4)$ is broken for all $m > 2$ and satisfied in the case of corresponding α-q-capacities, $C_{q,q}$ and $C_{q,2-q}$.

8. Conclusions and Future Work

A general treatment of the Sharma–Mittal entropy transfer was provided together with the analyses of existing information transfer measures for the non-additive Sharma–Mittal information transfer. It was shown that the existing definitions fail to satisfy at least one of the axioms common to the Shannon case, by which the information transfer has to be non-negative, less than the input and output uncertainty, equal to the input uncertainty in the case of perfect transmission and equal to zero in the case of a totally destructive channel. Thus, breaking some of these axioms implies unexpected and counterintuitive conclusions about the channels, such as achieving super-capacitance or sub-capacitance [4], which could be treated as nonphysical behavior. In this paper, alternative measures of the α-q mutual information and the α-q channel capacity were proposed so that all of the axioms which are broken in the case of the Sharma–Mittal information transfer measures considered before are satisfied, which could qualify them as physically consistent measures of information transfer.

Taking into account the previous research of non-extensive statistical mechanics [3], where the linear growth of the physical quantities has been recognized as a critical property in non-extensive [55] and non-exponentially growing systems [56], and taking into account the previous research from the field of information theory, where the Sharma–Mittal entropy has been considered an appropriate scaling measure which provides extensive information rates [21], the α-q mutual information and the α-q channel capacity seem to be

promising measures for the characterization of information transmission in the systems where the Shannon entropy rate diverges or disappears in an infinite time limit. In addition, as was shown in this paper, the proposed information transfer measures are compatible with the maximum likelihood detection, which indicates their potential for operational characterization of coding theory and hypothesis testing problems [26].

Author Contributions: Conceptualization, V.M.I. and I.B.D.; validation, V.M.I. and I.B.D.; formal analysis, V.M.I.; funding acquisition, I.B.D.; project administration, I.B.D.; writing—original draft preparation, V.M.I.; writing—review and editing, V.M.I. and I.B.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported in part by NSF under grants 1907918 and 1828132 and by Ministry of Science and Technological Development, Republic of Serbia, Grants Nos. ON 174026 and III 044006. The APC was funded by NSF under grants 1907918 and 1828132.

Data Availability Statement: Not applicable.

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

References

1. Ilić, V.M.; Stanković, M.S. A unified characterization of generalized information and certainty measures. *Phys. A Stat. Mech. Appl.* **2014**, *415*, 229–239. [CrossRef]
2. Renyi, A. *Probability Theory*; North-Holland Series in applied mathematics and mechanics; North-Holland Publishing Company: Amsterdam, The Netherlands, 1970.
3. Tsallis, C. Possible generalization of Boltzmann-Gibbs statistics. *J. Stat. Phys.* **1988**, *52*, 479–487. [CrossRef]
4. Landsberg, P.T.; Vedral, V. Distributions and channel capacities in generalized statistical mechanics. *Phys. Lett. A* **1998**, *247*, 211–217. [CrossRef]
5. Frank, T.; Daffertshofer, A. Exact time-dependent solutions of the Renyi Fokker-Planck equation and the Fokker-Planck equations related to the entropies proposed by Sharma and Mittal. *Phys. A Stat. Mech. Appl.* **2000**, *285*, 351–366. [CrossRef]
6. Sharma, B.; Mittal, D. New non-additive measures of entropy for discrete probability distributions. *J. Math. Sci.* **1975**, *10*, 28–40.
7. Tsallis, C. What are the numbers that experiments provide. *Quim. Nova* **1994**, *17*, 468–471.
8. Nivanen, L.; Le Méhauté, A.; Wang, Q.A. Generalized algebra within a nonextensive statistics. *Rep. Math. Phys.* **2003**, *52*, 437–444. [CrossRef]
9. Ilić, V.M.; Stanković, M.S. Generalized Shannon-Khinchin axioms and uniqueness theorem for pseudo-additive entropies. *Phys. A Stat. Mech. Appl.* **2014**, *411*, 138–145. [CrossRef]
10. Jizba, P.; Korbel, J. When Shannon and Khinchin meet Shore and Johnson: Equivalence of information theory and statistical inference axiomatics. *Phys. Rev. E* **2020**, *101*, 042126. [CrossRef]
11. Esteban, M.D.; Morales, D. A summary on entropy statistics. *Kybernetika* **1995**, *31*, 337–346.
12. Lenzi, E.; Scarfone, A. Extensive-like and intensive-like thermodynamical variables in generalized thermostatistics. *Phys. A Stat. Mech. Appl.* **2012**, *391*, 2543–2555. [CrossRef]
13. Frank, T.; Plastino, A. Generalized thermostatistics based on the Sharma-Mittal entropy and escort mean values. *Eur. Phys. J. B Condens. Matter Complex Syst.* **2002**, *30*, 543–549. [CrossRef]
14. Aktürk, O.Ü.; Aktürk, E.; Tomak, M. Can Sobolev inequality be written for Sharma-Mittal entropy? *Int. J. Theor. Phys.* **2008**, *47*, 3310–3320. [CrossRef]
15. Mazumdar, S.; Dutta, S.; Guha, P. Sharma–Mittal quantum discord. *Quantum Inf. Process.* **2019**, *18*, 1–26. [CrossRef]
16. Elhoseiny, M.; Elgammal, A. Generalized Twin Gaussian processes using Sharma–Mittal divergence. *Mach. Learn.* **2015**, *100*, 399–424. [CrossRef]
17. Koltcov, S.; Ignatenko, V.; Koltsova, O. Estimating Topic Modeling Performance with Sharma–Mittal Entropy. *Entropy* **2019**, *21*, 660. [CrossRef]
18. Jawad, A.; Bamba, K.; Younas, M.; Qummer, S.; Rani, S. Tsallis, Rényi and Sharma-Mittal holographic dark energy models in loop quantum cosmology. *Symmetry* **2018**, *10*, 635. [CrossRef]
19. Ghaffari, S.; Ziaie, A.; Moradpour, H.; Asghariyan, F.; Feleppa, F.; Tavayef, M. Black hole thermodynamics in Sharma–Mittal generalized entropy formalism. *Gen. Relativ. Gravit.* **2019**, *51*, 1–11. [CrossRef]
20. Américo, A.; Khouzani, M.; Malacaria, P. Conditional Entropy and Data Processing: An Axiomatic Approach Based on Core-Concavity. *IEEE Trans. Inf. Theory* **2020**, *66*, 5537–5547. [CrossRef]
21. Girardin, V.; Lhote, L. Rescaling entropy and divergence rates. *IEEE Trans. Inf. Theory* **2015**, *61*, 5868–5882. [CrossRef]
22. Ciuperca, G.; Girardin, V.; Lhote, L. Computation and estimation of generalized entropy rates for denumerable Markov chains. *IEEE Trans. Inf. Theory* **2011**, *57*, 4026–4034. [CrossRef]
23. Sibson, R. Information radius. *Z. Wahrscheinlichkeitstheorie Verwandte Geb.* **1969**, *14*, 149–160. [CrossRef]

24. Arimoto, S. Information Mesures and Capacity of Order α for Discrete Memoryless Channels. In *Topics in Information Theory*; Colloquia Mathematica Societatis János Bolyai; Csiszár, I., Elias, P., Eds.; North-Holland Pub. Co.: Amsterdam, The Netherlands, 1977; Volume 16, pp. 41–52.
25. Augustin, U. Noisy Channels. Ph.D. Thesis, Universität Erlangen-Nürnberg, Erlangen, Germany, 1978.
26. Csiszár, I. Generalized cutoff rates and Rényi's information measures. *IEEE Trans. Inf. Theory* **1995**, *41*, 26–34. [CrossRef]
27. Lapidoth, A.; Pfister, C. Two measures of dependence. *Entropy* **2019**, *21*, 778. [CrossRef]
28. Tomamichel, M.; Hayashi, M. Operational interpretation of Rényi information measures via composite hypothesis testing against product and Markov distributions. *IEEE Trans. Inf. Theory* **2017**, *64*, 1064–1082. [CrossRef]
29. Verdú, S. α-mutual information. In Proceedings of the 2015 Information Theory and Applications Workshop (ITA), San Diego, CA, USA, 1–6 February 2015; pp. 1–6.
30. Daróczy, Z. Generalized information functions. *Inf. Control* **1970**, *16*, 36–51. [CrossRef]
31. Chapeau-Blondeau, F.; Rousseau, D.; Delahaies, A. Renyi entropy measure of noise-aided information transmission in a binary channel. *Phys. Rev. E* **2010**, *81*, 051112. [CrossRef]
32. Chapeau-Blondeau, F.; Delahaies, A.; Rousseau, D. Tsallis entropy measure of noise-aided information transmission in a binary channel. *Phys. Lett. A* **2011**, *375*, 2211–2219. [CrossRef]
33. Yamano, T. A possible extension of Shannon's information theory. *Entropy* **2001**, *3*, 280–292. [CrossRef]
34. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [CrossRef]
35. Arimoto, S. Computation of random coding exponent functions. *Inf. Theory IEEE Trans.* **1976**, *22*, 665–671. [CrossRef]
36. Gallager, R. A simple derivation of the coding theorem and some applications. *IEEE Trans. Inf. Theory* **1965**, *11*, 3–18. [CrossRef]
37. Cover, T.M.; Thomas, J.A. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*; John Wiley & Sons, Inc: Hoboken, NJ, USA, 2006.
38. Fehr, S.; Berens, S. On the conditional Rényi entropy. *Inf. Theory IEEE Trans.* **2014**, *60*, 6801–6810. [CrossRef]
39. Wilde, M.M.; Winter, A.; Yang, D. Strong converse for the classical capacity of entanglement-breaking and Hadamard channels via a sandwiched Rényi relative entropy. *Commun. Math. Phys.* **2014**, *331*, 593–622. [CrossRef]
40. Gupta, M.K.; Wilde, M.M. Multiplicativity of completely bounded p-norms implies a strong converse for entanglement-assisted capacity. *Commun. Math. Phys.* **2015**, *334*, 867–887. [CrossRef]
41. Beigi, S. Sandwiched Rényi divergence satisfies data processing inequality. *J. Math. Phys.* **2013**, *54*, 122202. [CrossRef]
42. Hayashi, M.; Tomamichel, M. Correlation detection and an operational interpretation of the Rényi mutual information. *J. Math. Phys.* **2016**, *57*, 102201. [CrossRef]
43. Hayashi, M.; Tajima, H. Measurement-based formulation of quantum heat engines. *Phys. Rev. A* **2017**, *95*, 032132. [CrossRef]
44. Hayashi, M. Quantum Wiretap Channel With Non-Uniform Random Number and Its Exponent and Equivocation Rate of Leaked Information. *IEEE Trans. Inf. Theory* **2015**, *61*, 5595–5622. [CrossRef]
45. Cai, C.; Verdú, S. Conditional Rényi Divergence Saddlepoint and the Maximization of α-Mutual Information. *Entropy* **2019**, *21*, 969. [CrossRef]
46. Tridenski, S.; Zamir, R.; Ingber, A. The Ziv–Zakai–Rényi bound for joint source-channel coding. *IEEE Trans. Inf. Theory* **2015**, *61*, 4293–4315. [CrossRef]
47. Harremoës, P. Interpretations of Rényi entropies and divergences. *Phys. A Stat. Mech. Its Appl.* **2006**, *365*, 57–62. [CrossRef]
48. Jizba, P.; Kleinert, H.; Shefaat, M. Rényi's information transfer between financial time series. *Phys. A Stat. Mech. Appl.* **2012**, *391*, 2971–2989. [CrossRef]
49. Jizba, P.; Arimitsu, T. The world according to Rényi: Thermodynamics of multifractal systems. *Ann. Phys.* **2004**, *312*, 17–59. [CrossRef]
50. Iwamoto, M.; Shikata, J. Information theoretic security for encryption based on conditional Rényi entropies. In Proceedings of the International Conference on Information Theoretic Security, Singapore, 28–30 November 2013; pp. 103–121.
51. Ilić, V.; Djordjević, I.; Stanković, M. On a general definition of conditional Rényi entropies. *Proceedings* **2018**, *2*, 166. [CrossRef]
52. Fano, R.M. *Transmission of Information*; M.I.T. Press: Cambridge, MA, USA, 1961.
53. Ilic, V.M.; Djordjevic, I.B.; Küeppers, F. On the Daróczy-Tsallis capacities of discrete channels. *Entropy* **2015**, *20*, 2.
54. Yamano, T. Information theory based on nonadditive information content. *Phys. Rev. E* **2001**, *63*, 046105. [CrossRef]
55. Tsallis, C.; Gell-Mann, M.; Sato, Y. Asymptotically scale-invariant occupancy of phase space makes the entropy Sq extensive. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15377–15382. [CrossRef]
56. Korbel, J.; Hanel, R.; Thurner, S. Classification of complex systems by their sample-space scaling exponents. *New J. Phys.* **2018**, *20*, 093007. [CrossRef]