

On Two-Stage Guessing

Journal Article**Author(s):**

Graczyk, Robert; Sason, Igal

Publication date:

2021

Permanent link:

<https://doi.org/10.3929/ethz-b-000495006>

Rights / license:

[Creative Commons Attribution 4.0 International](#)

Originally published in:

Information 12(4), <https://doi.org/10.3390/info12040159>

Article

On Two-Stage Guessing

Robert Graczyk ¹ and Igal Sason ^{2,*}

¹ Signal and Information Processing Laboratory, ETH Zurich, 8092 Zurich, Switzerland; grackzyk@isi.ee.ethz.ch

² Andrew and Erna Viterbi Faculty of Electrical and Computer Engineering, Technion–Israel Institute of Technology, Haifa 3200003, Israel

* Correspondence: sason@ee.technion.ac.il; Tel.: +972-4-8294699

Abstract: Stationary memoryless sources produce two correlated random sequences X^n and Y^n . A guesser seeks to recover X^n in two stages, by first guessing Y^n and then X^n . The contributions of this work are twofold: (1) We characterize the least achievable exponential growth rate (in n) of any positive ρ -th moment of the total number of guesses when Y^n is obtained by applying a deterministic function f component-wise to X^n . We prove that, depending on f , the least exponential growth rate in the two-stage setup is lower than when guessing X^n directly. We further propose a simple Huffman code-based construction of a function f that is a viable candidate for the minimization of the least exponential growth rate in the two-stage guessing setup. (2) We characterize the least achievable exponential growth rate of the ρ -th moment of the total number of guesses required to recover X^n when Stage 1 need not end with a correct guess of Y^n and without assumptions on the stationary memoryless sources producing X^n and Y^n .

Keywords: guessing; majorization; method of types; Schur concavity; ranking function; Shannon entropy; Rényi entropy; Arimoto–Rényi conditional entropy



Citation: Graczyk, R.; Sason, I. On Two-Stage Guessing. *Information* **2021**, *12*, 159. <https://doi.org/10.3390/info12040159>

Academic Editor: Shraga I. Bross

Received: 26 February 2021

Accepted: 6 April 2021

Published: 9 April 2021

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



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

1. Introduction

Pioneered by Massey [1], McEliece and Yu [2], and Arikan [3], the guessing problem is concerned with recovering the realization of a finite-valued random variable X using a sequence of yes-no questions of the form “Is $X = x_1$?”, “Is $X = x_2$?”, etc., until correct. A commonly used performance metric for this problem is the ρ -th moment of the number of guesses until X is revealed (where ρ is a positive parameter).

When guessing a length- n i. i. d. sequence X^n (a tuple of n components that are drawn independently according to the law of X), the ρ -th moment of the number of guesses required to recover the realization of X^n grows exponentially with n , and the exponential growth rate is referred to as the guessing exponent. The least achievable guessing exponent was derived by Arikan [3], and it equals the order- $\frac{1}{1+\rho}$ Rényi entropy of X . Arikan’s result is based on the optimal deterministic guessing strategy, which proceeds in descending order of the probability mass function (PMF) of X^n .

In this paper, we propose and analyze a two-stage guessing strategy to recover the realization of an i. i. d. sequence X^n . In Stage 1, the guesser is allowed to produce guesses of an ancillary sequence Y^n that is jointly i. i. d. with X^n . In Stage 2, the guesser must recover X^n . We show the following:

1. When Y^n is generated by component-wise application of a mapping $f: \mathcal{X} \rightarrow \mathcal{Y}$ to X^n and the guesser is required to recover Y^n in Stage 1 before proceeding to Stage 2, then the least achievable guessing exponent (i.e., the exponential growth rate of the ρ -th moment of the total number of guesses in the two stages) equals

$$\rho \max \left\{ H_{\frac{1}{1+\rho}}(f(X)), H_{\frac{1}{1+\rho}}(X | f(X)) \right\}, \quad (1)$$

where the maximum is between the order- $\frac{1}{1+\rho}$ Rényi entropy of $f(X)$ and the conditional Arimoto–Rényi entropy of X given $f(X)$. We derive (1) in Section 3 and summarize our analysis in Theorem 1. We also propose a Huffman code-based construction of a function f that is a viable candidate for the minimization of (1) among all maps from \mathcal{X} to \mathcal{Y} (see Algorithm 2 and Theorem 2).

2. When X^n and Y^n are jointly i.i.d. according to the PMF P_{XY} and Stage 1 need not end with a correct guess of Y^n (i.e., the guesser may proceed to Stage 2 even if Y^n remains unknown), then the least achievable guessing exponent equals

$$\sup_{Q_{XY}} \left(\rho \min \{ H(Q_X), \max \{ H(Q_Y), H(Q_{X|Y}) \} \} - D(Q_{XY} \| P_{XY}) \right), \quad (2)$$

where the supremum is over all PMFs Q_{XY} defined on the same set as P_{XY} ; and $H(\cdot)$ and $D(\cdot \| \cdot)$ denote, respectively, the (conditional) Shannon entropy and the Kullback–Leibler divergence. We derive (2) in Section 4 and summarize our analysis in Theorem 3. Parts of Section 4 were presented in the conference paper [4].

Our interest in the two-stage guessing problem is due to its relation to information measures: Analogous to how the Rényi entropy can be defined operationally via guesswork, as opposed to its axiomatic definition, we view (1) and (2) as quantities that capture at what cost and to what extent knowledge of Y helps in recovering X . For example, minimizing (1) over descriptions $f(X)$ of X can be seen as isolating the most beneficial information of X in the sense that describing it in any more detail is too costly (the first term of the maximization in (1) exceeds the second), whereas a coarser description leaves too much uncertainty (the second term exceeds the first). Similarly, but with the joint law of (X, Y) fixed, (2) quantifies the least (partial) information of Y that benefits recovering X (because an optimal guessing strategy will proceed to Stage 2 when guessing Y no longer benefits guessing X). Note that while (1) and (2) are derived in this paper, studying their information-like properties is a subject of future research (see Section 5).

Besides its theoretic implications, the guessing problem is also applied practically in communications and cryptography. This includes sequential decoding [5,6], and measuring password strength [7], confidentiality of communication channels [8], and resilience against brute-force attacks [9]. It is also strongly related to task encoding and (lossless and lossy) compression (see, e.g., [10–15]).

Variations of the guessing problem include guessing under source uncertainty [16], distributed guessing [17,18], and guessing on the Gray–Wyner and the Slepian–Wolf network [19].

2. Preliminaries

We begin with some notation and preliminary material that are essential for the presentation in Section 3 ahead. The analysis in Section 4 relies on the method of types (see, e.g., Chapter 11 in [20]).

Throughout the paper, we use the following notation:

- For $m, n \in \mathbb{N}$ with $m < n$, let $[m : n] := \{m, \dots, n\}$;
- Let P be a PMF that is defined on a finite set \mathcal{X} . For $k \in [1 : |\mathcal{X}|]$, let $G_P(k)$ denote the sum of its k largest point masses, and let $p_{\max} := G_P(1)$. For $n \in \mathbb{N}$, denote by \mathcal{P}_n the set of all PMFs defined on $[1 : n]$.

The next definitions and properties are related to majorization and Rényi measures.

Definition 1 (Majorization). *Consider PMFs P and Q , defined on the same (finite or countably infinite) set \mathcal{X} . We say that Q majorizes P , denoted $P \prec Q$, if $G_P(k) \leq G_Q(k)$ for all $k \in [1 : |\mathcal{X}|]$. If P and Q are defined on finite sets of different cardinalities, then the PMF defined on the smaller set is zero-padded to match the cardinality of the larger set.*

By Definition 1, a unit mass majorizes any other distribution; on the other hand, the uniform distribution (on a finite set) is majorized by any other distribution of equal support.

Definition 2 (Schur-convexity/concavity). A function $f: \mathcal{P}_n \rightarrow \mathbb{R}$ is Schur-convex if, for every $P, Q \in \mathcal{P}_n$ with $P \prec Q$, we have $f(P) \leq f(Q)$. Likewise, f is Schur-concave if $-f$ is Schur-convex, i.e., if $P \prec Q$ implies that $f(P) \geq f(Q)$.

Definition 3 (Rényi entropy [21]). Let X be a random variable taking values on a finite or countably infinite set \mathcal{X} according to the PMF P_X . The order- α Rényi entropy $H_\alpha(X)$ of X is given by

$$H_\alpha(X) := \frac{1}{1-\alpha} \log \left(\sum_{x \in \mathcal{X}} P_X^\alpha(x) \right), \quad \alpha \in (0, 1) \cup (1, \infty), \tag{3}$$

where unless explicitly given, the base of $\log(\cdot)$ can be chosen arbitrarily, with $\exp(\cdot)$ denoting its inverse function. Via continuous extension,

$$H_0(X) := \log |\{x \in \mathcal{X} : P_X(x) > 0\}|, \tag{4}$$

$$H_1(X) := H(X) = - \sum_{x \in \mathcal{X}} P_X(x) \log P_X(x), \tag{5}$$

$$H_\infty(X) := \log \frac{1}{p_{\max}}, \tag{6}$$

where $H(X)$ is the (Shannon) entropy of X .

Proposition 1 (Schur-concavity of the Rényi entropy, Appendix F.3.a of [22]). The Rényi entropy of any order $\alpha > 0$ is Schur-concave (in particular, the Shannon entropy is Schur-concave).

Definition 4 (Arimoto-Rényi conditional entropy [23]). Let (X, Y) be a pair of random variables taking values on a product set $\mathcal{X} \times \mathcal{Y}$ according to the PMF P_{XY} . When \mathcal{X} is finite or countably infinite, the order- α Arimoto-Rényi conditional entropy $H_\alpha(X|Y)$ of X given Y is defined as follows:

- If $\alpha \in (0, 1) \cup (1, \infty)$,

$$H_\alpha(X|Y) := \frac{\alpha}{1-\alpha} \log \mathbb{E} \left[\left(\sum_{x \in \mathcal{X}} P_{X|Y}^\alpha(x|Y) \right)^{\frac{1}{\alpha}} \right]. \tag{7}$$

When \mathcal{Y} is finite, (7) can be simplified as follows:

$$H_\alpha(X|Y) = \frac{\alpha}{1-\alpha} \log \sum_{y \in \mathcal{Y}} \left(\sum_{x \in \mathcal{X}} P_{XY}^\alpha(x, y) \right)^{\frac{1}{\alpha}} \tag{8}$$

$$= \frac{\alpha}{1-\alpha} \log \sum_{y \in \mathcal{Y}} P_Y(y) \exp \left(\frac{1-\alpha}{\alpha} H_\alpha(X|Y = y) \right). \tag{9}$$

- If $\alpha \in \{0, 1, \infty\}$ and \mathcal{Y} is finite, then, via continuous extension,

$$H_0(X|Y) = \log \max_{y \in \mathcal{Y}} |\text{supp } P_{X|Y}(\cdot|y)| = \max_{y \in \mathcal{Y}} H_0(X|Y = y), \tag{10}$$

$$H_1(X|Y) = H(X|Y), \tag{11}$$

$$H_\infty(X|Y) = \log \frac{1}{\sum_{y \in \mathcal{Y}} \left\{ \max_{x \in \mathcal{X}} P_{X|Y}(x|Y) \cdot P_Y(y) \right\}}. \tag{12}$$

The properties of the Arimoto-Rényi conditional entropy were studied in [24,25].

Finally, the set $\mathcal{F}_{n,m}$ with $n, m \in \mathbb{N}$ denotes the set of all deterministic functions $f: [1 : n] \rightarrow [1 : m]$. If $m < n$, then a function $f \in \mathcal{F}_{n,m}$ is not one-to-one (i.e., it is a non-injective function).

3. Two-Stage Guessing: $Y_i = f(X_i)$

Let $X^n := (X_1, \dots, X_n)$ be a sequence of i.i.d. random variables taking values on a finite set \mathcal{X} . Assume without loss of generality that $\mathcal{X} = [1 : |\mathcal{X}|]$. Let $m \in [2 : |\mathcal{X}| - 1]$, $\mathcal{Y} := [1 : m]$, and let $f: \mathcal{X} \rightarrow \mathcal{Y}$ be a fixed deterministic function (i.e., $f \in \mathcal{F}_{|\mathcal{X}|,m}$). Consider guessing $X^n \in \mathcal{X}^n$ in two stages as follows (Algorithm 1).

Algorithm 1 (two-stage guessing algorithm):

- (a) Stage 1: $Y^n := (f(X_1), \dots, f(X_n)) \in \mathcal{Y}^n$ is guessed by asking questions of the form:
 - “Is $Y^n = \hat{y}_1$?”, “Is $Y^n = \hat{y}_2$?”, ...

until correct. Note that as $|\mathcal{Y}| = m < |\mathcal{X}|$, this stage cannot reveal X^n .

- (b) Stage 2: Based on Y^n , the sequence $X^n \in \mathcal{X}^n$ is guessed by asking questions of the form:

“Is $X^n = \hat{x}_1$?”, “Is $X^n = \hat{x}_2$?”, ...

until correct. If $Y^n = y^n$, the guesses $\hat{x}_k := (\hat{x}_{k,1}, \dots, \hat{x}_{k,n})$ are restricted to \mathcal{X} -sequences, which satisfy $f(\hat{x}_{k,i}) = y_i$ for all $i \in [1 : n]$.

The guesses $\hat{y}_1, \hat{y}_2, \dots$, in Stage 1 are in descending order of probability as measured by P_{Y^n} (i.e., \hat{y}_1 is the most probable sequence under P_{Y^n} ; \hat{y}_2 is the second most probable; and so on; ties are resolved arbitrarily). We denote the index of $y^n \in \mathcal{Y}^n$ in this guessing order by $g_{Y^n}(y^n)$. Note that because every sequence y^n is guessed exactly once, $g_{Y^n}(\cdot)$ is a bijection from \mathcal{Y}^n to $[1 : m^n]$; we refer to such bijections as ranking functions. The guesses $\hat{x}_1, \hat{x}_2, \dots$, in Stage 2 depend on Y^n and are in descending order of the posterior $P_{X^n|Y^n}(\cdot|Y^n)$. Following our notation from Stage 1, the index of $x^n \in \mathcal{X}^n$ in the guessing order induced by $Y^n = y^n$ is denoted $g_{X^n|Y^n}(x^n|y^n)$. Note that for every $y^n \in \mathcal{Y}^n$, the function $g_{X^n|Y^n}(\cdot|y^n)$ is a ranking function on \mathcal{X}^n . Using $g_{Y^n}(\cdot)$ and $g_{X^n|Y^n}(\cdot|\cdot)$, the total number of guesses $G_2(X^n)$ in Algorithm 1 can be expressed as

$$G_2(X^n) = g_{Y^n}(Y^n) + g_{X^n|Y^n}(X^n|Y^n), \tag{13}$$

where $g_{Y^n}(Y^n)$ and $g_{X^n|Y^n}(X^n|Y^n)$ are the number of guesses in Stages 1 and 2, respectively. Observe that guessing in descending order of probability minimizes the ρ -th moment of the number of guesses in both stages of Algorithm 1. By [3], for every $\rho > 0$, the guessing moments $\mathbb{E}[g_{Y^n}^\rho(Y^n)]$ and $\mathbb{E}[g_{X^n|Y^n}^\rho(X^n|Y^n)]$ can be (upper and lower) bounded in terms of $H_{\frac{1}{1+\rho}}(Y)$ and $H_{\frac{1}{1+\rho}}(X|Y)$ as follows:

$$\begin{aligned} (1 + n \ln m)^{-\rho} \exp\left(n\rho H_{\frac{1}{1+\rho}}(Y)\right) &\leq \mathbb{E}[g_{Y^n}^\rho(Y^n)] \\ &\leq \exp\left(n\rho H_{\frac{1}{1+\rho}}(Y)\right), \end{aligned} \tag{14a}$$

$$\begin{aligned} (1 + n \ln |\mathcal{X}|)^{-\rho} \exp\left(n\rho H_{\frac{1}{1+\rho}}(X|f(X))\right) &\leq \mathbb{E}[g_{X^n|Y^n}^\rho(X^n|Y^n)] \\ &\leq \exp\left(n\rho H_{\frac{1}{1+\rho}}(X|f(X))\right). \end{aligned} \tag{14b}$$

Combining (13) and (14), we next establish bounds on $\mathbb{E}[G_2(X^n)^\rho]$. In light of (13), we begin with bounds on the ρ -th power of a sum.

Lemma 1. Let $k \in \mathbb{N}$, and let $\{a_i\}_{i=1}^k$ be a non-negative sequence. For every $\rho > 0$,

$$s_1(k, \rho) \sum_{i=1}^k a_i^\rho \leq \left(\sum_{i=1}^k a_i \right)^\rho \leq s_2(k, \rho) \sum_{i=1}^k a_i^\rho, \tag{15}$$

where

$$s_1(k, \rho) := \begin{cases} 1, & \text{if } \rho \geq 1 \\ k^{\rho-1}, & \text{if } \rho \in (0, 1), \end{cases} \tag{16}$$

and

$$s_2(k, \rho) := \begin{cases} k^{\rho-1}, & \text{if } \rho \geq 1 \\ 1, & \text{if } \rho \in (0, 1). \end{cases} \tag{17}$$

- If $\rho \geq 1$, then the left and right inequalities in (15) hold with equality if, respectively, $k - 1$ of the a_i 's are equal to zero or $a_1 = \dots = a_k$;
- if $\rho \in (0, 1)$, then the left and right inequalities in (15) hold with equality if, respectively, $a_1 = \dots = a_k$ or $k - 1$ of the a_i 's are equal to zero.

Proof. See Appendix A. \square

Using the shorthand notation $k_1(\rho) := s_1(2, \rho)$, $k_2(\rho) := s_2(2, \rho)$, we apply Lemma 1 in conjunction with (13) and (14) (and the fact that $|\mathcal{X}| \geq m$) to bound $\mathbb{E}[G_2^\rho(X^n)]$ as follows:

$$k_1(\rho)(1 + n \ln |\mathcal{X}|)^{-\rho} \left[\exp\left(n\rho H_{\frac{1}{1+\rho}}(f(X))\right) + \exp\left(n\rho H_{\frac{1}{1+\rho}}(X | f(X))\right) \right] \leq \mathbb{E}[G_2^\rho(X^n)] \tag{18a}$$

$$\leq k_2(\rho) \left[\exp\left(n\rho H_{\frac{1}{1+\rho}}(f(X))\right) + \exp\left(n\rho H_{\frac{1}{1+\rho}}(X | f(X))\right) \right]. \tag{18b}$$

The bounds in (18) are asymptotically tight as n tends to infinity. To see this, note that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \ln \left(k_1(\rho)(1 + n \ln |\mathcal{X}|)^{-\rho} \right) = 0, \quad \lim_{n \rightarrow \infty} \frac{1}{n} \ln k_2(\rho) = 0, \tag{19}$$

and therefore, for all $\rho > 0$,

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{n} \ln \mathbb{E}[G_2^\rho(X^n)] \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \log \left(\exp\left(n\rho H_{\frac{1}{1+\rho}}(f(X))\right) + \exp\left(n\rho H_{\frac{1}{1+\rho}}(X | f(X))\right) \right). \end{aligned} \tag{20}$$

Since the sum of the two exponents on the right-hand side (RHS) of (20) is dominated by the larger exponential growth rate, it follows that

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{n} \log \left(\exp\left(n\rho H_{\frac{1}{1+\rho}}(f(X))\right) + \exp\left(n\rho H_{\frac{1}{1+\rho}}(X | f(X))\right) \right) \\ &= \rho \max \left\{ H_{\frac{1}{1+\rho}}(f(X)), H_{\frac{1}{1+\rho}}(X | f(X)) \right\}, \end{aligned} \tag{21}$$

and thus, by (20) and (21),

$$E_2(X; \rho, m, f) := \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}[G_2^\rho(X^n)] \tag{22}$$

$$= \rho \max \left\{ H_{\frac{1}{1+\rho}}(f(X)), H_{\frac{1}{1+\rho}}(X | f(X)) \right\}. \tag{23}$$

As a sanity check, note that if $m = |\mathcal{X}|$ and f is the identity function id (i.e., $\text{id}(x) = x$ for all $x \in \mathcal{X}$), then X^n is revealed in Stage 1, and (with Stage 2 obsolete) the ρ -th moment of the total number of guesses grows exponentially with rate

$$\rho H_{\frac{1}{1+\rho}}(\text{id}(X)) = \rho H_{\frac{1}{1+\rho}}(X).$$

This is in agreement with the RHS of (22), as

$$\begin{aligned} & \rho \max \left\{ H_{\frac{1}{1+\rho}}(\text{id}(X)), H_{\frac{1}{1+\rho}}(X | \text{id}(X)) \right\} \\ &= \rho \max \left\{ H_{\frac{1}{1+\rho}}(X), H_{\frac{1}{1+\rho}}(X | X) \right\} \end{aligned} \tag{24}$$

$$= \rho \max \left\{ H_{\frac{1}{1+\rho}}(X), 0 \right\} \tag{25}$$

$$= \rho H_{\frac{1}{1+\rho}}(X). \tag{26}$$

We summarize our results so far in Theorem 1 below.

Theorem 1. Let $X^n = (X_1, \dots, X_n)$ be a sequence of i.i.d. random variables, each drawn according to the PMF P_X of support $\mathcal{X} := [1 : |\mathcal{X}|]$. Let $m \in [2 : |\mathcal{X}| - 1]$, $f \in \mathcal{F}_{|\mathcal{X}|, m}$, and define $Y^n := (f(X_1), \dots, f(X_n))$. When guessing X^n according to Algorithm 1 (i.e., after first guessing Y^n in descending order of probability as measured by $P_{Y^n}(\cdot)$ and proceeding in descending order of probability as measured by $P_{X^n|Y^n}(\cdot|Y^n)$ for guessing X^n), the ρ -th moment of the total number of guesses $G_2(X^n)$ satisfies

- (a) the lower and upper bounds in (18) for all $n \in \mathbb{N}$ and $\rho > 0$;
- (b) the asymptotic characterization (22) for $\rho > 0$ and $n \rightarrow \infty$.

A Suboptimal and Simple Construction of f in Algorithm 1 and Bounds on $\mathbb{E}[G_2^\rho(X^n)]$

Having established in Theorem 1 that

$$\mathbb{E}[G_2^\rho(X^n)] \approx \exp(nE_2(X; \rho, m, f)), \quad \rho > 0, \tag{27}$$

we now seek to minimize the exponent $E_2(X; \rho, m, f)$ in the RHS of (23) (for given PMF P_X , $\rho > 0$, and $m \in [2 : |\mathcal{X}| - 1]$) over all $f \in \mathcal{F}_{|\mathcal{X}|, m}$.

We proceed by considering a sub-optimal and simple construction of f , which enables obtaining explicit bounds as a function of the PMF P_X and the value of m , while this construction also does not depend on ρ .

For a fixed $m \in [2 : |\mathcal{X}| - 1]$, a non-injective deterministic function $f_m^* : \mathcal{X} \rightarrow [1 : m]$ is constructed by relying on the Huffman algorithm for lossless compression of $X := X_1$. This construction also (almost) achieves the maximal mutual information $I(X; f(X))$ among all deterministic functions $f : \mathcal{X} \rightarrow [1 : m]$ (this issue is elaborated in the sequel). Heuristically, apart from its simplicity, the motivation of this sub-optimal construction can be justified since it is expected to reduce the guesswork in Stage 2 of Algorithm 1, where one wishes to guess X^n on the basis of the knowledge of Y^n with $Y_i = f(X_i)$ for all $i \in [1 : n]$. In this setting, it is shown that the upper and lower bounds on $\mathbb{E}[G_2^\rho(X^n)]$ are (almost) asymptotically tight in terms of their exponential growth rate in n . Furthermore, these exponential bounds demonstrate a reduction in the required number of guesses for X^n , as compared to the optimal one-stage guessing of X^n .

In the sequel, the following construction of a deterministic function $f_m^* : \mathcal{X} \rightarrow [1 : m]$ is analyzed; this construction was suggested in the proofs of Lemma 5 in [26] and Theorem 2 in [14].

Algorithm 2 (construction of f_m^*):

- (a) Let $i := 1, P_1 := P_X$.

- (b) If $|\text{supp}(P_i)| = m$, let $R := P_i$, and go to Step c. If not, let $P_{i+1} := P_i$ with its two least likely symbols merged as in the Huffman code construction. Let $i \leftarrow i + 1$, and go to Step b.
- (c) Construct $f_m^* \in \mathcal{F}_{|\mathcal{X}|,m}$ by setting $f_m^*(k) = j$ if $P_1(k)$ has been merged into $R(j)$.

We now define $Y^{*n} := (Y_1^*, \dots, Y_n^*)$ with

$$Y_i^* := f_m^*(X_i), \quad \forall i \in [1 : n]. \tag{28}$$

Observe that due to [26] (Corollary 3 and Lemma 5) and because $f_m^*(\cdot)$ operates component-wise on the i.i.d. vector X^n , the following lower bound on $\frac{1}{n} I(X^n; Y^{*n})$ applies:

$$\begin{aligned} \frac{1}{n} I(X^n; Y^{*n}) &= I(X; f_m^*(X)) \\ &= H(f_m^*(X)) \\ &\geq \max_{f \in \mathcal{F}_{|\mathcal{X}|,m}} H(f(X)) - \beta^* \\ &= \max_{f \in \mathcal{F}_{|\mathcal{X}|,m}} I(X; f(X)) - \beta^* \\ &= \max_{f \in \mathcal{F}_{|\mathcal{X}|,m}} \frac{1}{n} I(X^n; Y^n) - \beta^*, \end{aligned} \tag{29}$$

where $Y^n := (f(X_1), \dots, f(X_n))$ and

$$\beta^* := \log\left(\frac{2}{e \ln 2}\right) \approx 0.08607 \text{ bits.} \tag{30}$$

From the proof of [26] (Theorem 3), we further have the following multiplicative bound:

$$\frac{1}{n} I(X^n; Y^{*n}) \geq \frac{10}{11} \max_{f \in \mathcal{F}_{|\mathcal{X}|,m}} \frac{1}{n} I(X^n; Y^n). \tag{31}$$

Note that by [26] (Lemma 1), the maximization problem in the RHS of (29) is strongly NP-hard [27]. This means that, unless $P = NP$, there is no polynomial-time algorithm that, given an arbitrarily small $\varepsilon > 0$, produces a deterministic function $f^{(\varepsilon)} \in \mathcal{F}_{|\mathcal{X}|,m}$ satisfying

$$I(X; f^{(\varepsilon)}(X)) \geq (1 - \varepsilon) \max_{f \in \mathcal{F}_{|\mathcal{X}|,m}} I(X; f(X)). \tag{32}$$

We next examine the performance of our candidate function f_m^* when applied in Algorithm 1. To that end, we first bound $\mathbb{E}[g_{Y^n}^\rho(Y^{*n})]$ in terms of the Rényi entropy of a suitably defined random variable $\tilde{X}_m \in [1 : m]$ constructed in Algorithm 3 below. In the construction, we assume without loss of generality that $P_X(1) \geq \dots \geq P_X(|\mathcal{X}|)$ and denote the PMF of \tilde{X}_m by $Q := R_m(P_X)$.

Algorithm 3 (construction of the PMF $Q := R_m(P_X)$ of the random variable \tilde{X}_m):

- If $m = 1$, then $Q := R_1(P_X)$ is defined to be a point mass at one;
- If $m = |\mathcal{X}|$, then $Q := R_{|\mathcal{X}|}(P_X)$ is defined to be equal to P_X .

Furthermore, for $m \in [2 : |\mathcal{X}| - 1]$,

- (a) If $P_X(1) < \frac{1}{m}$, then Q is defined to be the equiprobable distribution on $[1 : m]$;
- (b) Otherwise, the PMF Q is defined as

$$Q(i) := \begin{cases} P_X(i), & \text{if } i \in [1 : m^*], \\ \frac{1}{m - m^*} \sum_{j=m^*+1}^{|\mathcal{X}|} P_X(j), & \text{if } i \in \{m^* + 1, \dots, m\}, \end{cases} \tag{33}$$

where m^* is the maximal integer $i \in [1 : m - 1]$, which satisfies

$$P_X(i) \geq \frac{1}{m-i} \sum_{j=i+1}^{|\mathcal{X}|} P_X(j). \tag{34}$$

Algorithm 3 was introduced in [26,28]. The link between the Rényi entropy of \tilde{X}_m and that of Y^* was given in Equation (34) of [14]:

$$H_\alpha(Y^*) \in [H_\alpha(\tilde{X}_m) - v(\alpha), H_\alpha(\tilde{X}_m)], \quad \forall \alpha > 0, \tag{35}$$

where

$$v(\alpha) := \begin{cases} \log\left(\frac{\alpha-1}{2^\alpha-2}\right) - \frac{\alpha}{\alpha-1} \log\left(\frac{\alpha}{2^\alpha-1}\right), & \alpha \neq 1 \\ \log\left(\frac{2}{e \ln 2}\right) \approx 0.08607 \text{ bits}, & \alpha = 1. \end{cases} \tag{36}$$

The function $v: (0, \infty) \rightarrow (0, \log 2)$ is depicted in Figure 1 on the following page. It is monotonically increasing, continuous, and it satisfies

$$\lim_{\alpha \downarrow 0} v(\alpha) = 0, \quad \lim_{\alpha \rightarrow \infty} v(\alpha) = \log 2 (= 1 \text{ bit}). \tag{37}$$

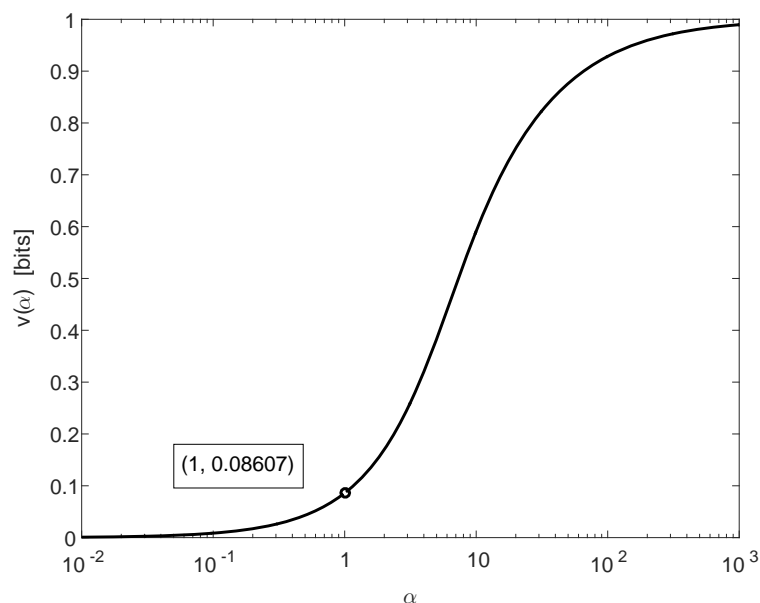


Figure 1. A plot of $v: (0, \infty) \rightarrow (0, \log 2)$, which is monotonically increasing, continuous, and satisfying (37).

Combining (14a) and (35) yields, for all $\rho > 0$,

$$\begin{aligned} (1 + n \ln m)^{-\rho} \exp\left(n \left[\rho H_{\frac{1}{1+\rho}}(\tilde{X}_m) - \rho v\left(\frac{1}{1+\rho}\right) \right]\right) \\ \leq \mathbb{E}[g_{Y^n}^\rho(Y^{*n})] \end{aligned} \tag{38a}$$

$$\leq \exp\left(n \rho H_{\frac{1}{1+\rho}}(\tilde{X}_m)\right), \tag{38b}$$

where, due to (30), the difference between the exponential growth rates (in n) of the lower and upper bounds in (38) is equal to $\rho v\left(\frac{1}{1+\rho}\right)$, and it can be verified to satisfy (see (30) and (36))

$$0 < \rho v\left(\frac{1}{1+\rho}\right) < \beta^* \approx 0.08607 \log 2, \quad \rho > 0, \tag{39}$$

where the leftmost and rightmost inequalities of (39) are asymptotically tight as we let $\rho \rightarrow 0^+$ and $\rho \rightarrow \infty$, respectively.

By inserting (38) into (18) and applying Inequality (39), it follows that for all $\rho > 0$,

$$\frac{k_1(\rho)}{(1+n \ln |\mathcal{X}|)^\rho} \left[\exp\left(n \left[\rho H_{\frac{1}{1+\rho}}(\tilde{X}_m) - \beta^* \right]\right) + \exp\left(n \rho H_{\frac{1}{1+\rho}}(X | f_m^*(X))\right) \right] \leq \mathbb{E}[G_2^\rho(X^n)] \tag{40a}$$

$$\leq k_2(\rho) \left[\exp\left(n \rho H_{\frac{1}{1+\rho}}(\tilde{X}_m)\right) + \exp\left(n \rho H_{\frac{1}{1+\rho}}(X | f_m^*(X))\right) \right]. \tag{40b}$$

Consequently, by letting n tend to infinity and relying on (22),

$$\max\left\{ \rho H_{\frac{1}{1+\rho}}(\tilde{X}_m) - \beta^*, \rho H_{\frac{1}{1+\rho}}(X | f_m^*(X)) \right\} \leq E_2(X; \rho, m, f_m^*) \tag{41a}$$

$$\leq \max\left\{ \rho H_{\frac{1}{1+\rho}}(\tilde{X}_m), \rho H_{\frac{1}{1+\rho}}(X | f_m^*(X)) \right\}. \tag{41b}$$

We next simplify the above bounds by evaluating the maxima in (41) as a function m and ρ . To that end, we use the following lemma.

Lemma 2. For $\alpha > 0$, let the two sequences $\{a_m(\alpha)\}$ and $\{b_m(\alpha)\}$ be given by

$$a_m(\alpha) := H_\alpha(\tilde{X}_m), \tag{42}$$

$$b_m(\alpha) := H_\alpha(X | f_m^*(X)), \tag{43}$$

with $m \in [1 : |\mathcal{X}|]$. Then,

- (a) The sequence $\{a_m(\alpha)\}$ is monotonically increasing (in m), and its first and last terms are zero and $H_\alpha(X)$, respectively.
- (b) The sequence $\{b_m(\alpha)\}$ is monotonically decreasing (in m), and its first and last terms are $H_\alpha(X)$ and zero, respectively.
- (c) If $\text{supp } P_X = \mathcal{X}$, then $\{a_m(\alpha)\}$ is strictly monotonically increasing, and $\{b_m(\alpha)\}$ is strictly monotonically decreasing. In particular, for all $m \in [2 : |\mathcal{X}| - 1]$, $a_m(\alpha)$ and $b_m(\alpha)$ are positive and strictly smaller than $H_\alpha(X)$.

Proof. See Appendix B. \square

Since symbols of probability zero (i.e., $x \in \mathcal{X}$ for which $P_X(x) = 0$) do not contribute to the expected number of guesses, assume without loss of generality that $\text{supp } P_X = \mathcal{X}$. In view of Lemma 2, we can therefore define

$$m_\rho^* = m_\rho^*(P_X) := \min\left\{ m \in [2 : |\mathcal{X}|] : a_m\left(\frac{1}{1+\rho}\right) \geq b_m\left(\frac{1}{1+\rho}\right) \right\}. \tag{44}$$

Using (44), we simplify (41) as follows:

- (a) If $m < m_\rho^*$, then

$$E_2(\rho, m) = \rho H_{\frac{1}{1+\rho}}(X | f_m^*(X)). \tag{45}$$

(b) Otherwise, if $m \geq m_\rho^*$, then

$$\rho H_{\frac{1}{1+\rho}}(\tilde{X}_m) - \beta^* \leq E_2(\rho, m) \leq \rho H_{\frac{1}{1+\rho}}(\tilde{X}_m). \tag{46}$$

Note that when guessing X^n directly, the ρ -th moment of the number of guesses grows exponentially with rate $\rho H_{\frac{1}{1+\rho}}(X)$ (cf. (14a)), due to Item c) in Lemma 2, since

$$H_{\frac{1}{1+\rho}}(\tilde{X}_m) = a_m \left(\frac{1}{1+\rho} \right) < a_{|\mathcal{X}|} \left(\frac{1}{1+\rho} \right) = H_{\frac{1}{1+\rho}}(X), \tag{47}$$

and also because conditioning (on a dependent chance variable) strictly reduces the Rényi entropy [24]. Equations (45) and (46) imply that for any $m \in [2 : |\mathcal{X}| - 1]$, guessing in two stages according to Algorithm 1 with $f = f_m^*$ reveals X^n sooner (in expectation) than guessing X^n directly.

We summarize our findings in this section in the second theorem below.

Theorem 2. For a given PMF P_X of support $\mathcal{X} = [1 : |\mathcal{X}|]$ and $m \in [2 : |\mathcal{X}| - 1]$, let the function $f_m^* \in \mathcal{F}_{|\mathcal{X}|,m}$ be constructed according to Algorithm 2, and let the random variable \tilde{X}_m be constructed according to the PMF in Algorithm 3. Let X^n be i.i.d. according to P_X , and let $Y^n := (f_m^*(X_1), \dots, f_m^*(X_n))$. Finally, for $\rho > 0$, let

$$E_1(\rho) := \rho H_{\frac{1}{1+\rho}}(X) \tag{48}$$

be the optimal exponential growth rate of the ρ -th moment single-stage guessing of X^n , and let

$$E_2(\rho, m) := E_2(X; \rho, m, f_m^*) \tag{49}$$

be given by (23) with $f := f_m^*$. Then, the following holds:

- (a) Cicalese et al., [26]: The maximization of the (normalized) mutual information $\frac{1}{n}I(X^n; Y^n)$ over all the deterministic functions $f: \mathcal{X} \rightarrow [1 : m]$ is a strongly NP-hard problem (for all n). However, the deterministic function $f := f_m^*$ almost achieves this maximization up to a small additive term, which is equal to $\beta^* := \log\left(\frac{2}{e \ln 2}\right) \approx 0.08607 \log 2$, and also up to a multiplicative term, which is equal to $\frac{10}{11}$ (see (29)–(31)).
- (b) The ρ -th moment of the number of guesses for X^n , which is required by the two-stage guessing in Algorithm 1, satisfies the non-asymptotic bounds in (40a) and (40b).
- (c) The asymptotic exponent $E_2(\rho, m)$ satisfies (45) and (46).
- (d) For all $\rho > 0$,

$$E_1(\rho) - E_2(\rho, m) \geq \rho \left[H_{\frac{1}{1+\rho}}(X) - \max \left\{ H_{\frac{1}{1+\rho}}(\tilde{X}_m), H_{\frac{1}{1+\rho}}(X | f_m^*(X)) \right\} \right] > 0, \tag{50}$$

so there is a reduction in the exponential growth rate (as a function of n) of the required number of guesses for X^n by Algorithm 1 (in comparison to the optimal one-stage guessing).

4. Two-Stage Guessing: Arbitrary (X, Y)

We next assume that X^n and Y^n are drawn jointly i.i.d. according to a given PMF P_{XY} , and we drop the requirement that Stage 1 need reveal Y^n prior to guessing X^n . Given $\rho > 0$, our goal in this section is to find the least exponential growth rate (in n) of the ρ -th moment of the total number of guesses required to recover X^n . Since Stage 1 may not reveal Y^n , we can no longer express the number of guesses using the ranking functions $g_{Y^n}(\cdot)$ and $g_{X^n|Y^n}(\cdot|\cdot)$ as in Section 3, and need new notation to capture the event that Y^n was not guessed in Stage 1. To that end, let \mathcal{G}_n be a subset of \mathcal{Y}^n , and let the ranking function

$$\tilde{g}_{Y^n}: \mathcal{G}_n \rightarrow [1 : |\mathcal{G}_n|] \tag{51}$$

denote the guessing order in Stage 1 with the understanding that if $Y^n \notin \mathcal{G}_n$, then

$$\tilde{g}_{Y^n}(Y^n) = |\mathcal{G}_n| \tag{52}$$

and the guesser moves on to Stage 2 knowing only that $Y^n \notin \mathcal{G}_n$. We denote the guessing order in Stage 2 by

$$\tilde{g}_{X^n|Y^n} : \mathcal{X}^n \times \mathcal{Y}^n \rightarrow [1 : |\mathcal{X}^n|], \tag{53}$$

where, for every $y^n \in \mathcal{Y}^n$, $\tilde{g}_{X^n|Y^n}(\cdot|y^n)$ is a ranking function on $[1 : |\mathcal{X}^n|]$ that satisfies

$$\tilde{g}_{X^n|Y^n}(\cdot|y^n) = \tilde{g}_{X^n|Y^n}(\cdot|\eta^n), \quad \forall y^n, \eta^n \notin \mathcal{G}_n. \tag{54}$$

Note that while \tilde{g}_{Y^n} and $\tilde{g}_{X^n|Y^n}$ depend on \mathcal{G}_n , we do not make this dependence explicit. In the remainder of this section, we prove the following variational characterization of the least exponential growth rate of the ρ -th moment of the total number of guesses in both stages $\tilde{g}_{Y^n}(Y^n) + \tilde{g}_{X^n|Y^n}(X^n|Y^n)$.

Theorem 3. *If (X^n, Y^n) are i.i.d. according to P_{XY} , then for all $\rho > 0$*

$$\begin{aligned} & \lim_{n \rightarrow \infty} \min_{\tilde{g}_{Y^n}, \tilde{g}_{X^n|Y^n}} \frac{1}{n} \log \mathbb{E} [(\tilde{g}_{Y^n}(Y^n) + \tilde{g}_{X^n|Y^n}(X^n|Y^n))^\rho] \\ & = \sup_{Q_{XY}} \left(\rho \min \{H(Q_X), \max \{H(Q_Y), H(Q_{X|Y})\}\} - D(Q_{XY} || P_{XY}) \right), \end{aligned} \tag{55}$$

where the supremum on the RHS of (55) is over all PMFs Q_{XY} on $\mathcal{X} \times \mathcal{Y}$ (and the limit exists).

Note that if P_{XY} is such that $Y = f(X)$, then the RHS of (55) is less than or equal to the RHS of (21). In other words, the guessing exponent of Theorem 3 is less than or equal to the guessing exponent of Theorem 1. This is due to the fact that guessing Y^n in Stage 1 before proceeding to Stage 2 (the strategy examined in Section 3) is just one of the admissible guessing strategies of Section 4 and not necessarily the optimal one.

We prove Theorem 3 in two parts: First, we show that the guesser can be assumed cognizant of the empirical joint type of (X^n, Y^n) ; by invoking the law of total expectation, averaging over denominator- n types Q_{XY} on $\mathcal{X} \times \mathcal{Y}$, we reduce the problem to evaluating the LHS of (55) under the assumption that (X^n, Y^n) is drawn uniformly at random from a type class $\mathcal{T}^{(n)}(Q_{XY})$ (instead of being i.i.d. P_{XY}). We conclude the proof by solving this reduced problem, showing in particular that when (X^n, Y^n) is drawn uniformly at random from a type class, the LHS of (55) can be achieved either by guessing Y^n in Stage 1 or skipping Stage 1 entirely.

We begin with the first part of the proof and show that the guesser can be assumed cognizant of the empirical joint type of (X^n, Y^n) ; we formalize and prove this claim in Corollary 1, which we derive from Lemma 3 below.

Lemma 3. *Let $\tilde{g}_{Y^n}^*$ and $\tilde{g}_{X^n|Y^n}^*$ be ranking functions that minimize the expectation in the LHS of (55), and likewise, let $\tilde{g}_{T;Y^n}^*$ and $\tilde{g}_{T;X^n|Y^n}^*$ be ranking functions cognizant of the empirical joint type $\Pi_{X^n Y^n}$ of (X^n, Y^n) that minimize the expectation in the LHS of (55) over all ranking functions depending on $\Pi_{X^n Y^n}$. Then, there exist positive constants a and k , which are independent of n , such that*

$$\mathbb{E} [(\tilde{g}_{Y^n}^*(Y^n) + \tilde{g}_{X^n|Y^n}^*(X^n|Y^n))^\rho] \leq \mathbb{E} [(\tilde{g}_{T;Y^n}^*(Y^n) + \tilde{g}_{T;X^n|Y^n}^*(X^n|Y^n))^\rho] kn^a. \tag{56}$$

Proof. See Appendix C. \square

Corollary 1. *If the limit*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}[(\tilde{g}_{T;Y^n}^*(Y^n) + \tilde{g}_{T;X^n|Y^n}^*(X^n|Y^n))^\rho], \quad \rho > 0, \tag{57}$$

exists, so does the limit in the LHS of (55), and the two are equal.

Proof. By (56),

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \min_{\tilde{g}_{Y^n}, \tilde{g}_{X^n|Y^n}} \frac{1}{n} \log \mathbb{E}[(\tilde{g}_{Y^n}(Y^n) + \tilde{g}_{X^n|Y^n}(X^n|Y^n))^\rho] \\ &= \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}[(\tilde{g}_{Y^n}^*(Y^n) + \tilde{g}_{X^n|Y^n}^*(X^n|Y^n))^\rho] \end{aligned} \tag{58}$$

$$\leq \limsup_{n \rightarrow \infty} \frac{1}{n} \log \left(\mathbb{E}[(\tilde{g}_{T;Y^n}^*(Y^n) + \tilde{g}_{T;X^n|Y^n}^*(X^n|Y^n))^\rho] kn^a \right) \tag{59}$$

$$= \limsup_{n \rightarrow \infty} \frac{1}{n} \left(\log \mathbb{E}[(\tilde{g}_{T;Y^n}^*(Y^n) + \tilde{g}_{T;X^n|Y^n}^*(X^n|Y^n))^\rho] + \log(k) + a \log(n) \right) \tag{60}$$

$$= \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}[(\tilde{g}_{T;Y^n}^*(Y^n) + \tilde{g}_{T;X^n|Y^n}^*(X^n|Y^n))^\rho] \tag{61}$$

$$= \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}[(\tilde{g}_{T;Y^n}^*(Y^n) + \tilde{g}_{T;X^n|Y^n}^*(X^n|Y^n))^\rho]. \tag{62}$$

The inverse inequality

$$\begin{aligned} & \liminf_{n \rightarrow \infty} \min_{\tilde{g}_{Y^n}, \tilde{g}_{X^n|Y^n}} \frac{1}{n} \log \mathbb{E}[(\tilde{g}_{Y^n}(Y^n) + \tilde{g}_{X^n|Y^n}(X^n|Y^n))^\rho] \\ & \geq \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}[(\tilde{g}_{T;Y^n}^*(Y^n) + \tilde{g}_{T;X^n|Y^n}^*(X^n|Y^n))^\rho] \end{aligned} \tag{63}$$

follows from the fact that an optimal guessing strategy depending on the empirical joint type $\Pi_{X^n Y^n}$ of (X^n, Y^n) cannot be outperformed by a guessing strategy ignorant of $\Pi_{X^n Y^n}$. \square

Corollary 1 states that the minimization in the LHS of (55) can be taken over guessing strategies cognizant of the empirical joint type of (X^n, Y^n) . As we show in the next lemma, this implies that evaluating the LHS of (55) can be further simplified by taking the expectation with (X^n, Y^n) drawn uniformly at random from a type class (instead of being i.i.d. P_{XY}).

Lemma 4. *Let $\mathbb{E}_{Q_{XY}}$ denote expectation with (X^n, Y^n) drawn uniformly at random from the type class $\mathcal{T}^{(n)}(Q_{XY})$. Then, the following limits exist and*

$$\begin{aligned} & \lim_{n \rightarrow \infty} \min_{\tilde{g}_{T;Y^n}, \tilde{g}_{T;X^n|Y^n}} \frac{1}{n} \log \mathbb{E}[(\tilde{g}_{T;Y^n}(Y^n) + \tilde{g}_{T;X^n|Y^n}(X^n|Y^n))^\rho] \\ &= \lim_{n \rightarrow \infty} \max_{Q_{XY}} \left(\min_{\tilde{g}_{T;Y^n}, \tilde{g}_{T;X^n|Y^n}} \frac{1}{n} \log \mathbb{E}_{Q_{XY}}[(\tilde{g}_{T;Y^n}(Y^n) + \tilde{g}_{T;X^n|Y^n}(X^n|Y^n))^\rho] \right. \\ & \quad \left. - D(Q_{XY} \| P_{XY}) \right), \end{aligned} \tag{64}$$

where the maximum in the RHS of (64) is taken over all denominator- n types on $\mathcal{X} \times \mathcal{Y}$.

Proof. Recall that $\Pi_{X^n Y^n}$ is the empirical joint type of (X^n, Y^n) , and let $\mathcal{T}_n(\mathcal{X} \times \mathcal{Y})$ denote the set of all denominator- n types on $\mathcal{X} \times \mathcal{Y}$. We prove Lemma 4 by applying the law of total expectation to the LHS of (64) (averaging over the events $\{\Pi_{X^n Y^n} = Q_{XY}\}$, $Q_{XY} \in \mathcal{T}_n(\mathcal{X} \times \mathcal{Y})$) and by approximating the probability of observing a given type using standard

tools from large deviations theory. We first show that the LHS of (64) is upper bounded by its RHS:

$$\begin{aligned} & \lim_{n \rightarrow \infty} \min_{\tilde{g}_{T;Y^n}, \tilde{g}_{T;X^n|Y^n}} \frac{1}{n} \log \mathbb{E} [(\tilde{g}_{T;Y^n}(Y^n) + \tilde{g}_{T;X^n|Y^n}(X^n|Y^n))^\rho] \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \log \left(\sum_{Q_{XY}} \min_{\tilde{g}_{T;Y^n}, \tilde{g}_{T;X^n|Y^n}} \mathbb{E}_{Q_{XY}} [(\tilde{g}_{T;Y^n}(Y^n) + \tilde{g}_{T;X^n|Y^n}(X^n|Y^n))^\rho] \right. \\ & \quad \left. \mathbb{P}[\Pi_{X^n Y^n} = Q_{XY}] \right) \end{aligned} \tag{65}$$

$$\leq \lim_{n \rightarrow \infty} \frac{1}{n} \log \left(\max_{Q_{XY}} \min_{\tilde{g}_{T;Y^n}, \tilde{g}_{T;X^n|Y^n}} \mathbb{E}_{Q_{XY}} [(\tilde{g}_{T;Y^n}(Y^n) + \tilde{g}_{T;X^n|Y^n}(X^n|Y^n))^\rho] \right. \\ \left. \mathbb{P}[\Pi_{X^n Y^n} = Q_{XY}] n^\alpha \right) \tag{66}$$

$$= \lim_{n \rightarrow \infty} \frac{1}{n} \log \left(\max_{Q_{XY}} \min_{\tilde{g}_{T;Y^n}, \tilde{g}_{T;X^n|Y^n}} \mathbb{E}_{Q_{XY}} [(\tilde{g}_{T;Y^n}(Y^n) + \tilde{g}_{T;X^n|Y^n}(X^n|Y^n))^\rho] \right. \\ \left. \mathbb{P}[\Pi_{X^n Y^n} = Q_{XY}] \right) \tag{67}$$

$$\leq \lim_{n \rightarrow \infty} \frac{1}{n} \log \left(\max_{Q_{XY}} \min_{\tilde{g}_{T;Y^n}, \tilde{g}_{T;X^n|Y^n}} \mathbb{E}_{Q_{XY}} [(\tilde{g}_{T;Y^n}(Y^n) + \tilde{g}_{T;X^n|Y^n}(X^n|Y^n))^\rho] \right. \\ \left. 2^{-nD(Q_{XY} \| P_{XY})} \right), \tag{68}$$

where (66) holds for sufficiently large α because the number of types grows polynomially in n (see Appendix C); and (68) follows from [20] (Theorem 11.1.4). We next show that the LHS of (64) is also lower bounded by its RHS:

$$\begin{aligned} & \lim_{n \rightarrow \infty} \min_{\tilde{g}_{T;Y^n}, \tilde{g}_{T;X^n|Y^n}} \frac{1}{n} \log \mathbb{E} [(\tilde{g}_{T;Y^n}(Y^n) + \tilde{g}_{T;X^n|Y^n}(X^n|Y^n))^\rho] \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \log \left(\sum_{Q_{XY}} \min_{\tilde{g}_{T;Y^n}, \tilde{g}_{T;X^n|Y^n}} \mathbb{E}_{Q_{XY}} [(\tilde{g}_{T;Y^n}(Y^n) + \tilde{g}_{T;X^n|Y^n}(X^n|Y^n))^\rho] \right. \\ & \quad \left. \mathbb{P}[\Pi_{X^n Y^n} = Q_{XY}] \right) \end{aligned} \tag{69}$$

$$\geq \lim_{n \rightarrow \infty} \frac{1}{n} \log \left(\max_{Q_{XY}} \min_{\tilde{g}_{T;Y^n}, \tilde{g}_{T;X^n|Y^n}} \mathbb{E}_{Q_{XY}} [(\tilde{g}_{T;Y^n}(Y^n) + \tilde{g}_{T;X^n|Y^n}(X^n|Y^n))^\rho] \right. \\ \left. \mathbb{P}[\Pi_{X^n Y^n} = Q_{XY}] \right) \tag{70}$$

$$\geq \lim_{n \rightarrow \infty} \frac{1}{n} \log \left(\max_{Q_{XY}} \min_{\tilde{g}_{T;Y^n}, \tilde{g}_{T;X^n|Y^n}} \mathbb{E}_{Q_{XY}} [(\tilde{g}_{T;Y^n}(Y^n) + \tilde{g}_{T;X^n|Y^n}(X^n|Y^n))^\rho] \right. \\ \left. 2^{-n[D(Q_{XY} \| P_{XY}) + \delta_n]} \right) \tag{71}$$

$$\geq \lim_{n \rightarrow \infty} \frac{1}{n} \log \left(\max_{Q_{XY}} \min_{\tilde{g}_{T;Y^n}, \tilde{g}_{T;X^n|Y^n}} \mathbb{E}_{Q_{XY}} [(\tilde{g}_{T;Y^n}(Y^n) + \tilde{g}_{T;X^n|Y^n}(X^n|Y^n))^\rho] \right. \\ \left. 2^{-nD(Q_{XY} \| P_{XY})} \right), \tag{72}$$

where in (71),

$$\delta_n := \frac{(|\mathcal{X}||\mathcal{Y}| - 1) \log(n + 1)}{n} \tag{73}$$

tends to zero as we let n tend to infinity, and the inequality in (71) follows again from [20] (Theorem 11.1.4). Together, (68) and (71) imply the equality in (64). \square

We now have established the first part of the proof of Theorem 3: In Corollary 1, we showed that the ranking functions \tilde{g}_{Y^n} and $\tilde{g}_{X^n|Y^n}$ can be assumed cognizant of the empirical joint type of (X^n, Y^n) , and in Lemma 4, we showed that under this assumption, the minimization of the ρ -th moment of the total number of guesses can be carried out with (X^n, Y^n) drawn uniformly at random from a type class.

Below, we give the second part of the proof: We show that if the pair (X^n, Y^n) is drawn uniformly at random from $\mathcal{T}^{(n)}(Q_{XY})$, then

$$\begin{aligned} & \lim_{n \rightarrow \infty} \min_{\tilde{g}_{Y^n}, \tilde{g}_{X^n|Y^n}} \frac{1}{n} \log \mathbb{E} [(\tilde{g}_{Y^n}(Y^n) + \tilde{g}_{X^n|Y^n}(X^n|Y^n))^\rho] \\ &= \rho \min \{H(Q_X), \max\{H(Q_Y), H(Q_{X|Y})\}\}, \quad \rho > 0. \end{aligned} \tag{74}$$

Note that Corollary 1 and Lemma 4 in conjunction with (74) conclude the proof of Theorem 3:

$$\begin{aligned} & \lim_{n \rightarrow \infty} \min_{\tilde{g}_{Y^n}, \tilde{g}_{X^n|Y^n}} \frac{1}{n} \log \mathbb{E} [(\tilde{g}_{Y^n}(Y^n) + \tilde{g}_{X^n|Y^n}(X^n|Y^n))^\rho] \\ &= \lim_{n \rightarrow \infty} \min_{\tilde{g}_{T;Y^n}, \tilde{g}_{T;X^n|Y^n}} \frac{1}{n} \log \mathbb{E} [(\tilde{g}_{T;Y^n}(Y^n) + \tilde{g}_{T;X^n|Y^n}(X^n|Y^n))^\rho] \end{aligned} \tag{75}$$

$$\begin{aligned} &= \lim_{n \rightarrow \infty} \max_{Q_{XY}} \left(\min_{\tilde{g}_{T;Y^n}, \tilde{g}_{T;X^n|Y^n}} \frac{1}{n} \log \mathbb{E}_{Q_{XY}} [(\tilde{g}_{T;Y^n}(Y^n) + \tilde{g}_{T;X^n|Y^n}(X^n|Y^n))^\rho] \right. \\ & \quad \left. - D(Q_{XY} \| P_{XY}) \right) \end{aligned} \tag{76}$$

$$= \sup_{Q_{XY}} \left(\rho \min \{H(Q_X), \max\{H(Q_Y), H(Q_{X|Y})\}\} - D(Q_{XY} \| P_{XY}) \right), \tag{77}$$

where the supremum in the RHS of (77) is taken over all PMFs Q_{XY} on $\mathcal{X} \times \mathcal{Y}$, and the step follows from (76) because the set of types is dense in the set of all PMFs (in the same sense that \mathbb{Q} is dense in \mathbb{R}).

It thus remains to prove (74). We begin with the direct part: Note that when (X^n, Y^n) is drawn uniformly at random from $\mathcal{T}^{(n)}(Q_{XY})$, the ρ -th moment of the total number of guesses grows exponentially with rate $\rho H(Q_X)$ if we skip Stage 1 and guess X^n directly and with rate $\rho \max\{H(Q_Y), H(Q_{X|Y})\}$ if we guess Y^n in Stage 1 before moving on to guessing X^n . To prove the second claim, we argue by case distinction on ρ . Assuming first that $\rho \leq 1$,

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E} [(\tilde{g}_{T;Y^n}(Y^n) + \tilde{g}_{T;X^n|Y^n}(X^n|Y^n))^\rho] \\ & \leq \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E} [\tilde{g}_{T;Y^n}(Y^n)^\rho + \tilde{g}_{T;X^n|Y^n}(X^n|Y^n)^\rho] \end{aligned} \tag{78}$$

$$= \lim_{n \rightarrow \infty} \frac{1}{n} \log \left(\mathbb{E} [\tilde{g}_{T;Y^n}(Y^n)^\rho] + \mathbb{E} [\tilde{g}_{T;X^n|Y^n}(X^n|Y^n)^\rho] \right) \tag{79}$$

$$\leq \lim_{n \rightarrow \infty} \frac{1}{n} \log \left(2^{n\rho H(Q_Y)} + 2^{n\rho H(Q_{X|Y})} \right) \tag{80}$$

$$= \rho \max \{H(Q_Y), H(Q_{X|Y})\}, \tag{81}$$

where (78) holds because $\rho \leq 1$ (see Lemma 1); (80) holds since, by the latter assumption, Y^n is revealed at the end of Stage 1, and thus, the guesser (cognizant of Q_{XY}) will only guess elements from the conditional type class $\mathcal{T}^{(n)}(Q_{X|Y}|Y^n)$; and (81) follows from the fact that the exponential growth rate of a sum of two exponents is dominated by the larger one. We wrap up the argument by showing that the LHS of (78) is also lower bounded by the RHS of (81):

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E} [(\tilde{g}_{T;Y^n}(Y^n) + \tilde{g}_{T;X^n|Y^n}(X^n|Y^n))^\rho] \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E} \left[2^\rho \left(\frac{1}{2} \tilde{g}_{T;Y^n}(Y^n) + \frac{1}{2} \tilde{g}_{T;X^n|Y^n}(X^n|Y^n) \right)^\rho \right] \end{aligned} \tag{82}$$

$$\geq \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E} [2^{\rho-1} (\tilde{g}_{T;Y^n}(Y^n)^\rho + \tilde{g}_{T;X^n|Y^n}(X^n|Y^n)^\rho)] \tag{83}$$

$$= \lim_{n \rightarrow \infty} \frac{1}{n} \log \left(\mathbb{E} [\tilde{g}_{T;Y^n}(Y^n)^\rho] + \mathbb{E} [\tilde{g}_{T;X^n|Y^n}(X^n|Y^n)^\rho] \right) \tag{84}$$

$$\geq \lim_{n \rightarrow \infty} \frac{1}{n} \log \left(\frac{1}{1+\rho} \left(2^{n\rho H(Q_Y)} + 2^{n\rho H(Q_{X|Y})} \right) 2^{-n\rho\delta_n} \right) \tag{85}$$

$$= \rho \max \{ H(Q_Y), H(Q_{X|Y}) \}, \tag{86}$$

where (83) follows from Jensen’s inequality; and (85) follows from [29] (Proposition 6.6) and the lower bound on the size of a (conditional) type class [20] (Theorem 11.1.3). The case $\rho > 1$ can be proven analogously and is hence omitted (with the term $2^{\rho-1}$ in the RHS of (83) replaced by one). Note that by applying the better of the two proposed guessing strategies (i.e., depending on Q_{XY} , either guess Y^n in Stage 1 or skip it) the ρ -th moment of the total number of guesses grows exponentially with rate $\rho \min \{ H(Q_X), \max \{ H(Q_Y), H(Q_{X|Y}) \} \}$. This concludes the direct part of the proof of (74). We remind the reader that while we have constructed a guessing strategy under the assumption that the empirical joint type $\Pi_{X^n Y^n}$ of (X^n, Y^n) is known, Lemma 3 implies the existence of a guessing strategy of equal asymptotic performance that does not depend on $\Pi_{X^n Y^n}$. Moreover, Lemma 3 is constructive in the sense that the type-independent guessing strategy can be explicitly derived from the type-cognizant one (cf. the proof of Proposition 6.6 in [29]).

We next establish the converse of (74) by showing that when (X^n, Y^n) is drawn uniformly at random from $\mathcal{T}^{(n)}(Q_{XY})$,

$$\begin{aligned} & \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E} [(\tilde{g}_{T;Y^n}(Y^n) + \tilde{g}_{T;X^n|Y^n}(X^n|Y^n))^\rho] \\ & \geq \rho \min \{ H(Q_X), \max \{ H(Q_Y), H(Q_{X|Y}) \} \} \end{aligned} \tag{87}$$

for all two-stage guessing strategies. To see why (87) holds, consider an arbitrary guessing strategy, and let the sequence n_1, n_2, \dots be such that

$$\begin{aligned} & \lim_{k \rightarrow \infty} \frac{1}{n_k} \log \mathbb{E} [(\tilde{g}_{T;Y^{n_k}}(Y^{n_k}) + \tilde{g}_{T;X^{n_k}|Y^{n_k}}(X^{n_k}|Y^{n_k}))^\rho] \\ & = \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E} [(\tilde{g}_{T;Y^n}(Y^n) + \tilde{g}_{T;X^n|Y^n}(X^n|Y^n))^\rho] \end{aligned} \tag{88}$$

and such that the limit

$$\alpha := \lim_{k \rightarrow \infty} \frac{|\mathcal{G}_{n_k}|}{|\mathcal{T}^{(n_k)}(Q_Y)|} \tag{89}$$

exists. Using Lemma 5 below, we show that the LHS of (88) (and thus, also the LHS of (87)) is lower bounded by $\rho H(Q_X)$ if $\alpha = 0$ and by $\rho \max \{ H(Q_Y), H(Q_{X|Y}) \}$ if $\alpha > 0$. This establishes the converse, because the lower of the two bounds must apply in any case.

Lemma 5. *If the pair (X^n, Y^n) is drawn uniformly at random from a type class $\mathcal{T}^{(n)}(Q_{XY})$ and*

$$\limsup_{n \rightarrow \infty} \frac{|\mathcal{G}_n|}{|\mathcal{T}^{(n)}(Q_Y)|} = 0, \tag{90}$$

then

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E} [\tilde{g}_{X^n|Y^n} (X^n|Y^n)^\rho] = \rho H(Q_X), \tag{91}$$

where Q_X and Q_Y denote the X - and Y -marginal of Q_{XY} .

Proof. Note that since

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log |\mathcal{T}^{(n)}(Q_X)| = H(Q_X), \tag{92}$$

the RHS of (91) is trivially upper bounded by $\rho H(Q_X)$. It thus suffices to show that (90) yields the lower bound

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E} [\tilde{g}_{X^n|Y^n} (X^n|Y^n)^\rho] \geq \rho H(Q_X). \tag{93}$$

To show that (90) yields (93), we define the indicator variable

$$E_n := \begin{cases} 0, & \text{if } Y^n \in \mathcal{G}_n \\ 1, & \text{else,} \end{cases} \tag{94}$$

and observe that due to (90) and the fact that Y^n is drawn uniformly at random from $\mathcal{T}^{(n)}(Q_Y)$,

$$\lim_{n \rightarrow \infty} \mathbb{P}[E_n = 1] = 1. \tag{95}$$

Consequently, $H(E_n)$ tends to zero as n tends to infinity, and because

$$H(X^n) - H(X^n | E_n) = I(X^n; E_n) \leq H(E_n), \tag{96}$$

we get

$$\lim_{n \rightarrow \infty} (H(X^n) - H(X^n | E_n)) = 0. \tag{97}$$

This and (95) imply that

$$\lim_{n \rightarrow \infty} \frac{1}{n} H(X^n) = \lim_{n \rightarrow \infty} \frac{1}{n} H(X^n | E_n = 1). \tag{98}$$

To conclude the proof of Lemma 5, we proceed as follows:

$$\begin{aligned} & \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E} [\tilde{g}_{X^n|Y^n} (X^n|Y^n)^\rho] \\ & \geq \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E} [\tilde{g}_{X^n|Y^n} (X^n|Y^n)^\rho | E_n = 1] \end{aligned} \tag{99}$$

$$\geq \liminf_{n \rightarrow \infty} \frac{1}{n} \rho H_{1/(1+\rho)}(X^n | E_n = 1) \tag{100}$$

$$\geq \liminf_{n \rightarrow \infty} \frac{1}{n} \rho H(X^n | E_n = 1) \tag{101}$$

$$= \liminf_{n \rightarrow \infty} \frac{1}{n} \rho H(X^n) \tag{102}$$

$$= \rho H(Q_X), \tag{103}$$

where (99) holds due to (95) and the law of total expectation; (100) follows from [3] (Theorem 1); (101) holds because the Rényi entropy is monotonically decreasing in its order and because $\rho > 0$; and (102) is due to (98). \square

We now conclude the proof of the converse part of (74). Assume first that α (as defined in (89)) equals zero. By (88) and Lemma 5,

$$\begin{aligned} & \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E} [(\tilde{g}_{T;Y^n}(Y^n) + \tilde{g}_{T;X^n|Y^n}(X^n|Y^n))^\rho] \\ &= \lim_{k \rightarrow \infty} \frac{1}{n_k} \log \mathbb{E} [(\tilde{g}_{T;Y^{n_k}}(Y^{n_k}) + \tilde{g}_{T;X^{n_k}|Y^{n_k}}(X^{n_k}|Y^{n_k}))^\rho] \end{aligned} \tag{104}$$

$$\geq \liminf_{k \rightarrow \infty} \frac{1}{n_k} \log \mathbb{E} [\tilde{g}_{T;Y^{n_k}}(X^{n_k}|Y^{n_k})^\rho] \tag{105}$$

$$= \rho H(Q_X), \tag{106}$$

establishing the first contribution to the RHS of (87). Next, let $\alpha > 0$. Applying [29] (Proposition 6.6) in conjunction with (89) and the fact that Y^{n_k} is drawn uniformly at random from $\mathcal{T}^{(n_k)}(Q_Y)$,

$$\mathbb{E} [\tilde{g}_{T;Y^{n_k}}(Y^{n_k})^\rho] \geq \frac{\alpha}{2} \cdot \frac{2^{n_k \rho H(Q_Y)} 2^{-n_k \delta_{n_k}}}{1 + \rho} \tag{107}$$

for all sufficiently large k . Using (107) and proceeding analogously as in (82) to (85), we now establish the second contribution to the RHS of (87):

$$\begin{aligned} & \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E} [(\tilde{g}_{T;Y^n}(Y^n) + \tilde{g}_{T;X^n|Y^n}(X^n|Y^n))^\rho] \\ &= \lim_{k \rightarrow \infty} \frac{1}{n_k} \log \mathbb{E} [(\tilde{g}_{T;Y^{n_k}}(Y^{n_k}) + \tilde{g}_{T;X^{n_k}|Y^{n_k}}(X^{n_k}|Y^{n_k}))^\rho] \end{aligned} \tag{108}$$

$$\geq \liminf_{k \rightarrow \infty} \frac{1}{n_k} \log \left(\mathbb{E} [\tilde{g}_{T;Y^{n_k}}(Y^{n_k})^\rho] + \mathbb{E} [\tilde{g}_{T;X^{n_k}|Y^{n_k}}(X^{n_k}|Y^{n_k})^\rho] \right) \tag{109}$$

$$\geq \liminf_{k \rightarrow \infty} \frac{1}{n_k} \log \left(\frac{\alpha}{2} \cdot \frac{2^{n_k \rho H(Q_Y)} 2^{-n_k \delta_{n_k}}}{1 + \rho} + \mathbb{E} [\tilde{g}_{T;X^{n_k}|Y^{n_k}}(X^{n_k}|Y^{n_k})^\rho] \right) \tag{110}$$

$$\geq \liminf_{k \rightarrow \infty} \frac{1}{n_k} \log \left(\frac{\alpha}{2} \cdot \frac{2^{n_k \rho H(Q_Y)} 2^{-n_k \delta_{n_k}}}{1 + \rho} + \frac{2^{n_k \rho H(Q_{X|Y})} 2^{-n_k \delta_{n_k}}}{1 + \rho} \right) \tag{111}$$

$$= \rho \max \{ H(Q_Y), H(Q_{X|Y}) \}, \tag{112}$$

where (110) is due to (107); and in (111), we granted the guesser access to Y^n at the beginning of Stage 2.

5. Summary and Outlook

We proposed a new variation on the Massey–Arikan guessing problem where, instead of guessing X^n directly, the guesser is allowed to first produce guesses of a correlated ancillary sequence Y^n . We characterized the least achievable exponential growth rate (in n) of the ρ -th moment of the total number of guesses in the two stages when X^n is i.i.d. according to P_X , $Y_i = f(X_i)$ for all $i \in [1 : n]$, and the guesser must recover Y^n in Stage 1 before proceeding to Stage 2 (Section 3, Theorems 1 and 2); and when the pair (X^n, Y^n) is jointly i.i.d. according to P_{XY} and Stage 1 need not reveal Y^n (Section 4, Theorem 3). Future directions of this work include:

1. The generalization of our results to a larger class of sources (e.g., Markov sources);
2. A study of the information-like properties of the guessing exponents (1) and (2);
3. Finding the optimal block-wise description $Y^n = f(X^n)$ and its associated two-stage guessing exponent;
4. The generalization of the cryptographic problems [8,9] to a setting where the adversary may also produce guesses of leaked side information.

Author Contributions: Both authors contributed equally to this work (in its various stages which include the conceptualization, analysis, writing and proofreading). Both authors read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing not applicable.

Acknowledgments: The authors are indebted to Amos Lapidoth for his contribution to Section 4 (see [4]). The constructive comments in the review process, which helped to improve the presentation, are gratefully acknowledged.

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

Appendix A. Proof of Lemma 1

If $\rho \geq 1$, then

$$\begin{aligned} \left(\sum_{i=1}^k a_i\right)^\rho &= k^\rho \left(\frac{1}{k} \sum_{i=1}^k a_i\right)^\rho \\ &\leq k^\rho \cdot \frac{1}{k} \sum_{i=1}^k a_i^\rho \\ &= k^{\rho-1} \sum_{i=1}^k a_i^\rho, \end{aligned} \tag{A1}$$

where (A1) holds by Jensen’s inequality and since the mapping $x \mapsto x^\rho$ for $x \geq 0$ is convex. If at least one of the non-negative a_i ’s is positive (if all a_i ’s are zero, it is trivial), then

$$\begin{aligned} \left(\sum_{i=1}^k a_i\right)^\rho &= \left(\sum_{j=1}^k a_j\right)^\rho \sum_{i=1}^k \left(\frac{a_i}{\sum_{j=1}^k a_j}\right)^\rho \\ &\geq \left(\sum_{j=1}^k a_j\right)^\rho \sum_{i=1}^k \left(\frac{a_i}{\sum_{j=1}^k a_j}\right)^\rho \\ &= \sum_{i=1}^k a_i^\rho, \end{aligned} \tag{A2}$$

$$= \sum_{i=1}^k a_i^\rho, \tag{A3}$$

where (A2) holds since $0 \leq \frac{a_i}{\sum_{j=1}^k a_j} \leq 1$ for all $i \in [1 : n]$, and $\rho \geq 1$. If $\rho \in (0,1)$, then inequalities (A1) and (A2) are reversed. The conditions for equalities in (15) are easily verified.

Appendix B. Proof of Lemma 2

From (33), $R_1(P_X)$ is a unit probability mass at one and $R_{|\mathcal{X}|}(P_X) \equiv P_X$, so (42) gives that

$$a_1(\alpha) = 0, \quad a_{|\mathcal{X}|}(\alpha) = H_\alpha(X). \tag{B1}$$

In view of Lemma 5 in [14], it follows that for all $m \in [1 : |\mathcal{X}| - 1]$,

$$a_{m+1}(\alpha) = \max_{Q \in \mathcal{P}_{m+1} : P_X \prec Q} H_\alpha(Q) \tag{B2}$$

$$\geq \max_{Q \in \mathcal{P}_m : P_X \prec Q} H_\alpha(Q) \tag{B3}$$

$$= a_m(\alpha), \tag{B4}$$

where (B2) and (B4) are due to [14] (38) and (42); (B3) holds since $\mathcal{P}_m \subset \mathcal{P}_{m+1}$.

We next prove Item b. Consider the sequence of functions $\{f_m^*\}_{m=1}^{|\mathcal{X}|}$, defined over the set \mathcal{X} . By construction (see Algorithm 2), $f_{|\mathcal{X}|}^*$ is the identity function since all the respective $|\mathcal{X}|$ nodes in the Huffman algorithm stay un-changed in this case. We also have $f_1^*(x) = 1$ for all $x \in \mathcal{X}$ (in the latter case, by Algorithm 2, all nodes are merged by the Huffman algorithm into a single node). Hence, from (43),

$$b_1(\alpha) = H_\alpha(X), \quad b_{|\mathcal{X}|}(\alpha) = 0. \tag{B5}$$

Consider the construction of the function f_m^* by Algorithm 2. Since the transition from $m + 1$ to m nodes is obtained by merging two nodes without affecting the other $m - 1$ nodes, it follows by the data processing theorem for the Arimoto–Rényi conditional entropy (see [24] (Theorem 2 and Corollary 1)) that, for all $m \in [1 : |\mathcal{X}| - 1]$,

$$b_{m+1}(\alpha) = H_\alpha(X | f_{m+1}^*(X)) \leq H_\alpha(X | f_m^*(X)) = b_m(\alpha). \tag{B6}$$

We finally prove Item c. Suppose that P_X is supported on the set \mathcal{X} . Under this assumption, it follows from the strict Schur concavity of the Rényi entropy that the inequality in (B3) is strict, and therefore, (B2)–(B4) imply that $a_m(\alpha) < a_{m+1}(\alpha)$ for all $m \in [1 : |\mathcal{X}| - 1]$. In particular, Item a implies that $0 < a_m(\alpha) < H_\alpha(X)$ holds for every $m \in [2 : |\mathcal{X}| - 1]$. Furthermore, the conditioning on $f_{m+1}^*(X)$ enables distinguishing between the two labels of \mathcal{X} , which correspond to the pair of nodes that are being merged (by the Huffman algorithm) in the transition from $f_{m+1}^*(X)$ to $f_m^*(X)$. Hence, the inequality in (B6) turns out to be strict under the assumption that P_X is supported on the set \mathcal{X} . In particular, under that assumption, it follows from Item b that $0 < b_m(\alpha) < H_\alpha(X)$ holds for every $m \in [2 : |\mathcal{X}| - 1]$.

Appendix C. Proof of Lemma 3

We prove Lemma 3 as a consequence of Corollary C1 below and the fact that the number of denominator- n types on a finite set grows polynomially in n ([20], Theorem 11.1.1).

Corollary C1. (Moser [29], (6.47) and Corollary 6.10) *Let the random triple (U, V, W) take values in the finite set $\mathcal{U} \times \mathcal{V} \times \mathcal{W}$, and let $(\tilde{g}_U^*(\cdot), \tilde{g}_{U|V}^*(\cdot|\cdot))$ and $(\tilde{g}_{U|W}^*(\cdot|\cdot), \tilde{g}_{U|V,W}^*(\cdot|\cdot, \cdot))$ be ranking functions that, for a given $\rho > 0$, minimize*

$$\mathbb{E}[(\tilde{g}_U(U) + \tilde{g}_{U|V}(U|V))^\rho] \tag{C1}$$

over all two-stage guessing strategies (with no access to W) and

$$\mathbb{E}[(\tilde{g}_{U|W}(U|W) + \tilde{g}_{U|V,W}(U|V, W))^\rho] \tag{C2}$$

over all two-stage guessing strategies cognizant of W . Then,

$$\mathbb{E}[(\tilde{g}_U^*(U) + \tilde{g}_{U|V}^*(U|V))^\rho] \leq \mathbb{E}[(\tilde{g}_{U|W}^*(U|W) + \tilde{g}_{U|V,W}^*(U|V, W))^\rho] |\mathcal{W}|^\rho. \tag{C3}$$

Lemma 3 follows from Corollary C1 with $U \leftarrow X^n, \mathcal{U} \leftarrow \mathcal{X}^n; V \leftarrow Y^n, \mathcal{V} \leftarrow \mathcal{Y}^n; W \leftarrow \Pi_{X^n Y^n}; \mathcal{W} \leftarrow \mathcal{T}_n(\mathcal{X} \times \mathcal{Y})$, and by noticing that for all $n \in \mathbb{N}$,

$$|\mathcal{T}_n(\mathcal{X} \times \mathcal{Y})|^\rho \leq (n + 1)^{\rho(|\mathcal{X} \times \mathcal{Y}| - 1)} \leq k n^a, \tag{C4}$$

where $a := \rho (|\mathcal{X} \times \mathcal{Y}| - 1)$ and $k := 2^a$ are positive constants independent of n .

References

1. Massey, J.L. Guessing and entropy. In Proceedings of the 1994 IEEE International Symposium on Information Theory, Trondheim, Norway, 27 June–1 July 1994; p. 204. [\[CrossRef\]](#)
2. McEliece, R.J.; Yu, Z. An inequality on entropy. In Proceedings of the 1995 IEEE International Symposium on Information Theory, Whistler, BC, Canada, 17–22 September 1995; p. 329. [\[CrossRef\]](#)
3. Arikan, E. An inequality on guessing and its application to sequential decoding. *IEEE Trans. Inf. Theory* **1996**, *42*, 99–105. [\[CrossRef\]](#) [\[CrossRef\]](#)
4. Graczyk, R.; Lapidoth, A. Two-stage guessing. In Proceedings of the 2019 IEEE International Symposium on Information Theory, Paris, France, 7–12 July 2019; pp. 475–479. [\[CrossRef\]](#)
5. Arikan, E.; Merhav, N. Joint source-channel coding and guessing with application to sequential decoding. *IEEE Trans. Information Theory* **1998**, *44*, 1756–1769. [\[CrossRef\]](#) [\[CrossRef\]](#)
6. Boztaş, S. Comments on “An inequality on guessing and its application to sequential decoding”. *IEEE Trans. Inf. Theory* **1997**, *43*, 2062–2063. [\[CrossRef\]](#) [\[CrossRef\]](#)
7. Cachin, C. Entropy Measures and Unconditional Security in Cryptography. Ph.D. Thesis, ETH Zurich, Zurich, Switzerland, 1997. [\[CrossRef\]](#)
8. Merhav, N.; Arikan, E. The Shannon cipher system with a guessing wiretapper. *IEEE Trans. Inf. Theory* **1999**, *45*, 1860–1866. [\[CrossRef\]](#) [\[CrossRef\]](#)
9. Bracher, A.; Hof, E.; Lapidoth, A. Guessing attacks on distributed-storage systems. *IEEE Trans. Inf. Theory* **2019**, *65*, 6975–6998. [\[CrossRef\]](#) [\[CrossRef\]](#)
10. Arikan, E.; Merhav, N. Guessing subject to distortion. *IEEE Trans. Inf. Theory* **1998**, *44*, 1041–1056. [\[CrossRef\]](#) [\[CrossRef\]](#)
11. Bracher, A.; Lapidoth, A.; Pfister, C. Distributed task encoding. In Proceedings of the 2017 IEEE International Symposium on Information Theory, Aachen, Germany, 25–30 June 2017; pp. 1993–1997. [\[CrossRef\]](#)
12. Bracher, A.; Lapidoth, A.; Pfister, C. Guessing with distributed encoders. *Entropy* **2019**, *21*, 298. [\[CrossRef\]](#) [\[CrossRef\]](#) [\[PubMed\]](#)
13. Christiansen, M.M.; Duffy, K.R. Guesswork, large deviations, and Shannon entropy. *IEEE Trans. Inf. Theory* **2013**, *59*, 796–802. [\[CrossRef\]](#) [\[CrossRef\]](#)
14. Sason, I. Tight bounds on the Rényi entropy via majorization with applications to guessing and compression. *Entropy* **2018**, *20*, 896. [\[CrossRef\]](#) [\[CrossRef\]](#) [\[PubMed\]](#)
15. Sason, I.; Verdú, S. Improved bounds on lossless source coding and guessing moments via Rényi measures. *IEEE Trans. Inf. Theory* **2018**, *64*, 4323–4346. [\[CrossRef\]](#) [\[CrossRef\]](#)
16. Sundaresan, R. Guessing under source uncertainty. *IEEE Trans. Inf. Theory* **2007**, *53*, 269–287. [\[CrossRef\]](#) [\[CrossRef\]](#)
17. Merhav, N.; Cohen, A. Universal randomized guessing with application to asynchronous decentralized brute-force attacks. *IEEE Trans. Inf. Theory* **2020**, *66*, 114–129. [\[CrossRef\]](#) [\[CrossRef\]](#)
18. Salamatian, S.; Beirami, A.; Cohen, A.; Médard, M. Centralized versus decentralized multi-agent guesswork. In Proceedings of the 2017 IEEE International Symposium on Information Theory, Aachen, Germany, 25–30 June 2017; pp. 2263–2267. [\[CrossRef\]](#)
19. Graczyk, R.; Lapidoth, A. Gray-Wyner and Slepian-Wolf guessing. In Proceedings of the 2020 IEEE International Symposium on Information Theory, Los Angeles, CA, USA, 21–26 June 2020; pp. 2207–2211. [\[CrossRef\]](#)
20. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2006. [\[CrossRef\]](#)
21. Rényi, A. On measures of entropy and information. In Proceedings of the 4th Berkeley Symposium on Probability Theory and Mathematical Statistics, Berkeley, CA, USA, 8–9 August 1961; pp. 547–561. [\[CrossRef\]](#)
22. Marshall, A.W.; Olkin, I.; Arnold, B.C. *Inequalities: Theory of Majorization and Its Applications*, 2nd ed.; Springer: New York, NY, USA, 2011. [\[CrossRef\]](#)
23. Arimoto, S. Information measures and capacity of order α for discrete memoryless channels. In Proceedings of the 2nd Colloquium on Information Theory, Keszthely, Hungary, 25–30 August 1975; Csiszár, I., Elias, P., Eds.; Colloquia Mathematica Societatis Janós Bolyai: Amsterdam, The Netherlands, 1977; Volume 16, pp. 41–52.
24. Fehr, S.; Berens, S. On the conditional Rényi entropy. *IEEE Trans. Inf. Theory* **2014**, *60*, 6801–6810. [\[CrossRef\]](#) [\[CrossRef\]](#)
25. Sason, I.; Verdú, S. Arimoto–Rényi conditional entropy and Bayesian M -ary hypothesis testing. *IEEE Trans. Inf. Theory* **2018**, *64*, 4–25. [\[CrossRef\]](#) [\[CrossRef\]](#)
26. Cicalese, F.; Gargano, L.; Vaccaro, U. Bounds on the entropy of a function of a random variable and their applications. *IEEE Trans. Inf. Theory* **2018**, *64*, 2220–2230. [\[CrossRef\]](#) [\[CrossRef\]](#)
27. Garey, M.R.; Johnson, D.S. *Computers and Intractability: A Guide to the Theory of NP-Completeness*; W. H. Freedman and Company: New York, NY, USA, 1979.
28. Cicalese, F.; Gargano, L.; Vaccaro, U. An information theoretic approach to probability mass function truncation. In Proceedings of the 2019 IEEE International Symposium on Information Theory, Paris, France, 7–12 July 2019; pp. 702–706. [\[CrossRef\]](#)
29. Moser, S.M. *Advanced Topics in Information Theory (Lecture Notes)*, 4th edition; Signal and Information Processing Laboratory: ETH Zürich, Switzerland; Institute of Communications Engineering, National Chiao Tung University: Hsinchu, Taiwan, 2020. [\[CrossRef\]](#)