entropy

MDPI

Article

Guessing with Distributed Encoders

Annina Bracher [1], Amos Lapidoth [2] and Christoph Pfister [2,*]

1 P&C Solutions, Swiss Re, 8022 Zurich, Switzerland; annina.bracher@gmail.com
2 Signal and Information Processing Laboratory, ETH Zurich, 8092 Zurich, Switzerland; lapidoth@isi.ee.ethz.ch
* Correspondence: pfister@isi.ee.ethz.ch

Received: 10 December 2018; Accepted: 14 March 2019; Published: 19 March 2019

Abstract: Two correlated sources emit a pair of sequences, each of which is observed by a different encoder. Each encoder produces a rate-limited description of the sequence it observes, and the two descriptions are presented to a guessing device that repeatedly produces sequence pairs until correct. The number of guesses until correct is random, and it is required that it have a moment (of some prespecified order) that tends to one as the length of the sequences tends to infinity. The description rate pairs that allow this are characterized in terms of the Rényi entropy and the Arimoto–Rényi conditional entropy of the joint law of the sources. This solves the guessing analog of the Slepian–Wolf distributed source-coding problem. The achievability is based on random binning, which is analyzed using a technique by Rosenthal.

Keywords: Arimoto–Rényi conditional entropy; distributed source coding; guessing; Rényi entropy

1. Introduction

In the Massey–Arıkan guessing problem [1,2], a random variable X is drawn from a finite set $\mathcal{X}$ according to some probability mass function (PMF) P_X, and it has to be determined by making guesses of the form "Is X equal to x?" until the guess is correct. The guessing order is determined by a guessing function G, which is a bijective function from $\mathcal{X}$ to $\{1,\ldots,|\mathcal{X}|\}$. Guessing according to G proceeds as follows: the first guess is the element $\hat{x}_1 \in \mathcal{X}$ satisfying $G(\hat{x}_1) = 1$; the second guess is the element $\hat{x}_2 \in \mathcal{X}$ satisfying $G(\hat{x}_2) = 2$, and so on. Consequently, $G(X)$ is the number of guesses needed to guess X. Arıkan [2] showed that for any $\rho > 0$, the ρth moment of the number of guesses required by an optimal guesser G^* to guess X is bounded by:

$$\frac{1}{(1+\ln|\mathcal{X}|)^\rho}\, 2^{\rho H_{1/(1+\rho)}(X)} \leq \mathrm{E}[G^*(X)^\rho] \leq 2^{\rho H_{1/(1+\rho)}(X)}, \tag{1}$$

where $\ln(\cdot)$ denotes the natural logarithm, and $H_{1/(1+\rho)}(X)$ denotes the Rényi entropy of order $\frac{1}{1+\rho}$, which is defined in Section 3 ahead (refinements of (1) were recently derived in [3]).

Guessing with an encoder is depicted in Figure 1. Here, prior to guessing X, the guesser is provided some side information about X in the form of $f(X)$, where $f\colon \mathcal{X} \to \{1,\ldots,\mathsf{M}\}$ is a function taking on at most M different values ("labels"). Accordingly, a guessing function $G(\cdot|\cdot)$ is a function from $\mathcal{X} \times \{1,\ldots,\mathsf{M}\}$ to $\{1,\ldots,|\mathcal{X}|\}$ such that for every label $m \in \{1,\ldots,\mathsf{M}\}$, $G(\cdot|m)\colon \mathcal{X} \to \{1,\ldots,|\mathcal{X}|\}$ is bijective. If, among all encoders, f^* minimizes the ρth moment of the number of guesses required by an optimal guesser to guess X after observing $f(X)$, then [4] (Corollary 7):

$$\frac{1}{(1+\ln|\mathcal{X}|)^\rho}\, 2^{\rho[H_{1/(1+\rho)}(X)-\log\mathsf{M}]} \leq \mathrm{E}[G^*(X|f^*(X))^\rho] \leq 1 + 2^{\rho[H_{1/(1+\rho)}(X)-\log\mathsf{M}+1]}. \tag{2}$$

Thus, in guessing a sequence of independent and identically distributed (IID) random variables, a description rate of approximately $H_{1/(1+\rho)}(X)$ bits per symbol is needed to drive the ρth moment of the number of guesses to one as the sequence length tends to infinity [4,5] (see Section 2 for more related work).

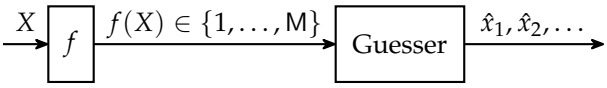

Figure 1. Guessing with an encoder f.

In this paper, we generalize the single-encoder setting from Figure 1 to the setting with distributed encoders depicted in Figure 2, which is the analog of Slepian–Wolf coding [6] for guessing: A source generates a sequence of pairs $\{(X_i, Y_i)\}_{i=1}^{n}$ over a finite alphabet $\mathcal{X} \times \mathcal{Y}$. The sequence X^n is described by one of $\lfloor 2^{n\mathsf{R_X}} \rfloor$ labels and the sequence Y^n by one of $\lfloor 2^{n\mathsf{R_Y}} \rfloor$ labels using functions:

$$f_n \colon \mathcal{X}^n \to \{1, \ldots, \lfloor 2^{n\mathsf{R_X}} \rfloor\}, \tag{3}$$

$$g_n \colon \mathcal{Y}^n \to \{1, \ldots, \lfloor 2^{n\mathsf{R_Y}} \rfloor\}, \tag{4}$$

where $\mathsf{R_X} \geq 0$ and $\mathsf{R_Y} \geq 0$. Based on $f_n(X^n)$ and $g_n(Y^n)$, a guesser repeatedly produces guesses of the form $(\hat{x}^n, \hat{y}^n)$ until $(\hat{x}^n, \hat{y}^n) = (X^n, Y^n)$.

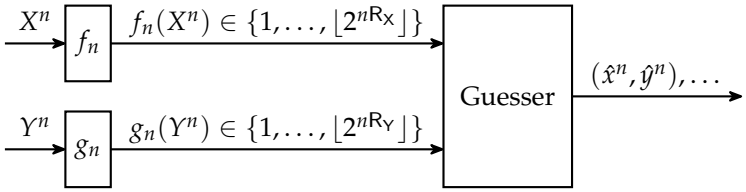

Figure 2. Guessing with distributed encoders f_n and g_n.

For a fixed $\rho > 0$, a rate pair $(\mathsf{R_X}, \mathsf{R_Y}) \in \mathbb{R}^2_{\geq 0}$ is called achievable if there exists a sequence of encoders and guessing functions $\{(f_n, g_n, G_n)\}_{n=1}^{\infty}$ such that the ρth moment of the number of guesses tends to one as n tends to infinity, i.e.,

$$\lim_{n\to\infty} \mathrm{E}\big[G_n\big(X^n, Y^n | f_n(X^n), g_n(Y^n)\big)^{\rho}\big] = 1. \tag{5}$$

Our main contribution is Theorem 1, which characterizes the achievable rate pairs. For a fixed $\rho > 0$, let the region $\mathcal{R}(\rho)$ comprise all rate pairs $(\mathsf{R_X}, \mathsf{R_Y}) \in \mathbb{R}^2_{\geq 0}$ satisfying the following inequalities simultaneously:

$$\mathsf{R_X} \geq \limsup_{n\to\infty} \frac{H_{\tilde{\rho}}(X^n|Y^n)}{n}, \tag{6}$$

$$\mathsf{R_Y} \geq \limsup_{n\to\infty} \frac{H_{\tilde{\rho}}(Y^n|X^n)}{n}, \tag{7}$$

$$\mathsf{R_X} + \mathsf{R_Y} \geq \limsup_{n\to\infty} \frac{H_{\tilde{\rho}}(X^n, Y^n)}{n}, \tag{8}$$

where the Rényi entropy $H_\alpha(\cdot)$ and the Arimoto–Rényi conditional entropy $H_\alpha(\cdot|\cdot)$ of order α are both defined in Section 3 ahead, and throughout the paper,

$$\tilde{\rho} \triangleq \frac{1}{1+\rho}. \tag{9}$$

Theorem 1. *For any $\rho > 0$, all rate pairs in the interior of $\mathcal{R}(\rho)$ are achievable, while those outside $\mathcal{R}(\rho)$ are not. If $\{(X_i, Y_i)\}_{i=1}^{\infty}$ are IID according to P_{XY}, then (6)–(8) reduce to:*

$$\mathsf{R}_\mathsf{X} \geq H_{\tilde{\rho}}(X|Y), \tag{10}$$

$$\mathsf{R}_\mathsf{Y} \geq H_{\tilde{\rho}}(Y|X), \tag{11}$$

$$\mathsf{R}_\mathsf{X} + \mathsf{R}_\mathsf{Y} \geq H_{\tilde{\rho}}(X,Y). \tag{12}$$

Proof. The converse follows from Corollary 1 in Section 4; the achievability follows from Corollary 2 in Section 5; and the reduction of (6)–(8) to (10)–(12) in the IID case follows from (19) and (20) ahead. □

The rate region defined by (10)–(12) resembles the rate region of Slepian–Wolf coding [6] (Theorem 15.4.1); the difference is that the Shannon entropy and conditional entropy are replaced by their Rényi counterparts. The rate regions are related as follows:

Remark 1. *For memoryless sources and $\rho > 0$, the region $\mathcal{R}(\rho)$ is contained in the Slepian–Wolf region. Typically, the containment is strict.*

Proof. The containment follows from the monotonicity of the Arimoto–Rényi conditional entropy: (9) implies that $\tilde{\rho} \in (0,1)$, so, by [7] (Proposition 5), $H_{\tilde{\rho}}(X|Y) \geq H(X|Y)$, $H_{\tilde{\rho}}(Y|X) \geq H(Y|X)$, and $H_{\tilde{\rho}}(X,Y) \geq H(X,Y)$. As for the strict containment, first note that the Slepian–Wolf region contains at least one rate pair $(\mathsf{R}_\mathsf{X}, \mathsf{R}_\mathsf{Y})$ satisfying $\mathsf{R}_\mathsf{X} + \mathsf{R}_\mathsf{Y} = H(X,Y)$. Consequently, if $H_{\tilde{\rho}}(X,Y) > H(X,Y)$, then the containment is strict. Because $H_{\tilde{\rho}}(X,Y) > H(X,Y)$ unless (X,Y) is distributed uniformly over its support [8], the containment is typically strict.

The claim can also be shown operationally: The probability of error is equal to the probability that more than one guess is needed, and for every $\rho > 0$,

$$\Pr\big[G_n\big(X^n, Y^n | f_n(X^n), g_n(Y^n)\big) \geq 2\big] = \Pr\big[G_n\big(X^n, Y^n | f_n(X^n), g_n(Y^n)\big)^\rho - 1 \geq 2^\rho - 1\big] \tag{13}$$

$$\leq \frac{\mathrm{E}\big[G_n\big(X^n, Y^n | f_n(X^n), g_n(Y^n)\big)^\rho\big] - 1}{2^\rho - 1}, \tag{14}$$

where (14) follows from Markov's inequality. Thus, the probability of error tends to zero if the ρth moment of the number of guesses tends to one. □

Despite the resemblance between (10)–(12) and the Slepian–Wolf region, there is an important difference: while Slepian–Wolf coding allows separate encoding with the same sum rate as with joint encoding, this is not necessarily true in our setting:

Remark 2. *Although the sum rate constraint (12) is the same as in single-source guessing [5], separate encoding of X^n and Y^n may require a larger sum rate than joint encoding of X^n and Y^n.*

Proof. If $H_{\tilde{\rho}}(X|Y) + H_{\tilde{\rho}}(Y|X) > H_{\tilde{\rho}}(X,Y)$, then (10) and (11) together impose a stronger constraint on the sum rate than (12). For example, if:

$P_{XY}(x,y)$	$y=0$	$y=1$
$x=0$	0.65	0.17
$x=1$	0.17	0.01

and $\rho = 1$, then $H_{1/2}(X|Y) + H_{1/2}(Y|X) \approx 1.61$ bits, so separate (distributed) encoding requires a sum rate exceeding 1.61 bits as opposed to joint encoding, which is possible with $H_{1/2}(X,Y) \approx 1.58$ bits (in Slepian–Wolf coding, this cannot happen because $H(X,Y) - H(X|Y) - H(Y|X) = I(X;Y) \geq 0$). □

The guessing problem is related to the task-encoding problem, where based on $f_n(X^n)$ and $g_n(Y^n)$, the decoder outputs a list that is guaranteed to contain (X^n, Y^n), and the ρth moment of the list size is required to tend to one as n tends to infinity. While, in the single-source setting, the guessing

problem and the task-encoding problem have the same asymptotics [4], this is not the case in the distributed setting:

Remark 3. *For memoryless sources, the task-encoding region from [9] is strictly smaller than the guessing region $\mathcal{R}(\rho)$ unless X and Y are independent.*

Proof. In the IID case, the task-encoding region is the set of all rate pairs $(\mathsf{R}_\mathsf{X}, \mathsf{R}_\mathsf{Y}) \in \mathbb{R}^2_{\geq 0}$ satisfying the following inequalities [9] (Theorem 1):

$$\mathsf{R}_\mathsf{X} \geq H_{\tilde{\rho}}(X), \tag{15}$$

$$\mathsf{R}_\mathsf{Y} \geq H_{\tilde{\rho}}(Y), \tag{16}$$

$$\mathsf{R}_\mathsf{X} + \mathsf{R}_\mathsf{Y} \geq H_{\tilde{\rho}}(X,Y) + K_{\tilde{\rho}}(X;Y), \tag{17}$$

where $K_\alpha(X;Y)$ is a Rényi measure of dependence studied in [10] (when α is one, $K_\alpha(X;Y)$ is the mutual information). The claim now follows from the following observations: By [7] (Theorem 2), $H_{\tilde{\rho}}(X) \geq H_{\tilde{\rho}}(X|Y)$ with equality if and only if X and Y are independent; similarly, $H_{\tilde{\rho}}(Y) \geq H_{\tilde{\rho}}(Y|X)$ with equality if and only if X and Y are independent; and by [10] (Theorem 2), $K_{\tilde{\rho}}(X;Y) \geq 0$ with equality if and only if X and Y are independent. □

The rest of this paper is structured as follows: in Section 2, we review other guessing settings; in Section 3, we recall the Rényi information measures and prove some auxiliary lemmas; in Section 4, we prove the converse theorem; and in Section 5, we prove the achievability theorem, which is based on random binning and, in the case $\rho > 1$, is analyzed using a technique by Rosenthal [11].

2. Related Work

Tighter versions of (1) can be found in [3,12]. The large deviation behavior of guessing was studied in [13,14]. The relation between guessing and variable-length lossless source coding was explored in [3,15,16].

Mismatched guessing, where the assumed distribution of X does not match its actual distribution, was studied in [17], along with guessing under source uncertainty, where the PMF of X belongs to some known set, and a guesser was sought with good worst-case performance over that set. Guessing subject to distortion, where instead of guessing X, it suffices to guess an $\hat{X}$ that is close to X according to some distortion measure, was treated in [18].

If the guesser observes some side information Y, then the ρth moment of the number of guesses required by an optimal guesser is bounded by [2]:

$$\frac{1}{(1+\ln|\mathcal{X}|)^\rho}\, 2^{\rho H_{\tilde{\rho}}(X|Y)} \leq \mathrm{E}[G^*(X|Y)^\rho] \leq 2^{\rho H_{\tilde{\rho}}(X|Y)}, \tag{18}$$

where $H_{\tilde{\rho}}(X|Y)$ denotes the Arimoto–Rényi conditional entropy of order $\tilde{\rho} = \frac{1}{1+\rho}$, which is defined in Section 3 ahead (refinements of (18) were recently derived in [3]). Guessing is related to the cutoff rate of a discrete memoryless channel, which is the supremum over all rates for which the ρth moment of the number of guesses needed by the decoder to guess the message can be driven to one as the block length tends to infinity. In [2,19], the cutoff rate was expressed in terms of Gallager's E_0 function [20]. Joint source-channel guessing was considered in [21].

Guessing with an encoder, i.e., the situation where the side information can be chosen, was studied in [4], where it was also shown that guessing and task encoding [22] have the same asymptotics. With distributed encoders, however, task encoding [9] and guessing no longer have the same asymptotics; see Remark 3. Lower and upper bounds for guessing with a helper, i.e., an encoder that does not observe X, but has access to a random variable that is correlated with X, can be found in [5].

3. Preliminaries

Throughout the paper, $\log(\cdot)$ denotes the base-two logarithm. When clear from the context, we often omit sets and subscripts; for example, we write $\sum_x$ for $\sum_{x\in\mathcal{X}}$ and $P(x)$ for $P_X(x)$. The Rényi entropy [23] of order α is defined for positive α other than one as:

$$H_\alpha(X) \triangleq \frac{1}{1-\alpha}\log\sum_x P(x)^\alpha. \tag{19}$$

In the limit as α tends to one, the Shannon entropy is recovered, i.e., $\lim_{\alpha\to 1} H_\alpha(X) = H(X)$. The Arimoto–Rényi conditional entropy [24] of order α is defined for positive α other than one as:

$$H_\alpha(X|Y) \triangleq \frac{\alpha}{1-\alpha}\log\sum_y\left[\sum_x P(x,y)^\alpha\right]^{\frac{1}{\alpha}}. \tag{20}$$

In the limit as α tends to one, the Shannon conditional entropy is recovered, i.e., $\lim_{\alpha\to 1} H_\alpha(X|Y) = H(X|Y)$. The properties of the Arimoto–Rényi conditional entropy were studied in [7,24,25].

In the rest of this section, we recall some properties of the Arimoto–Rényi conditional entropy that will be used in Section 4 (Lemmas 1–3), and we prove auxiliary results for Section 5 (Lemmas 4–7).

Lemma 1 ([7], Theorem 2)**.** *Let $\alpha > 0$, and let P_{XYZ} be a PMF over the finite set $\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}$. Then,*

$$H_\alpha(X|Y,Z) \le H_\alpha(X|Z) \tag{21}$$

with equality if and only if X —○— Z —○— Y form a Markov chain.

Lemma 2 ([7], Proposition 4)**.** *Let $\alpha > 0$, and let P_{XYZ} be a PMF over the finite set $\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}$. Then,*

$$H_\alpha(X,Y|Z) \ge H_\alpha(X|Z) \tag{22}$$

with equality if and only if Y is uniquely determined by X and Z.

Lemma 3 ([7], Theorem 3)**.** *Let $\alpha > 0$, and let P_{XYZ} be a PMF over the finite set $\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}$. Then,*

$$H_\alpha(X|Y,Z) \ge H_\alpha(X|Z) - \log|\mathcal{Y}|. \tag{23}$$

Lemma 4 ([20], Problem 4.15(f))**.** *Let $\mathcal{Y}$ be a finite set, and let $f\colon \mathcal{Y}\to\mathbb{R}_{\ge 0}$. Then, for all $p\in(0,1]$,*

$$\left[\sum_y f(y)\right]^p \le \sum_y f(y)^p. \tag{24}$$

Proof. If $\sum_y f(y) = 0$, then (24) holds because the left-hand side (LHS) and the right-hand side (RHS) are both zero. If $\sum_y f(y) > 0$, then:

$$\sum_y f(y)^p = \left[\sum_{y'} f(y')\right]^p \sum_y\left[\frac{f(y)}{\sum_{y'} f(y')}\right]^p \tag{25}$$

$$\ge \left[\sum_{y'} f(y')\right]^p \sum_y \frac{f(y)}{\sum_{y'} f(y')} \tag{26}$$

$$= \left[\sum_{y'} f(y')\right]^p, \tag{27}$$

where (26) holds because $p \in (0,1]$ and $f(y)/\sum_{y'} f(y') \in [0,1]$ for every $y \in \mathcal{Y}$. □

Lemma 5. *Let a, b, and c be nonnegative integers. Then, for all $p > 0$,*

$$(1+a+b+c)^p \leq 1 + 4^p(a^p + b^p + c^p) \tag{28}$$

(the restriction to integers cannot be omitted; for example, (28) does not hold if $a = b = c = 0.1$ and $p = 2$).

Proof. If $p \in (0,1]$, then (28) follows from Lemma 4 because $4^p \geq 1$. If $p > 1$, then the cases with $a+b+c \in \{0,1,2\}$ can be checked individually. For $a+b+c \geq 3$,

$$(1+a+b+c)^p = \left[\frac{3}{a+b+c} + 3\right]^p \cdot \left[\frac{a+b+c}{3}\right]^p \tag{29}$$

$$\leq 4^p \cdot \left[\frac{a+b+c}{3}\right]^p \tag{30}$$

$$\leq 4^p \cdot \frac{a^p + b^p + c^p}{3} \tag{31}$$

$$\leq 1 + 4^p(a^p + b^p + c^p), \tag{32}$$

where (30) holds because $a+b+c \geq 3$, and (31) follows from Jensen's inequality because $z \mapsto z^p$ is convex on $\mathbb{R}_{\geq 0}$ since $p > 1$. □

Lemma 6. *Let a, b, c, and d be nonnegative real numbers. Then, for all $p > 0$,*

$$(a+b+c+d)^p \leq 4^p(a^p + b^p + c^p + d^p). \tag{33}$$

Proof. If $p \in (0,1]$, then (33) follows from Lemma 4 because $4^p \geq 1$. If $p > 1$, then:

$$(a+b+c+d)^p = 4^p \cdot \left[\frac{a+b+c+d}{4}\right]^p \tag{34}$$

$$\leq 4^p \cdot \frac{a^p + b^p + c^p + d^p}{4} \tag{35}$$

$$\leq 4^p(a^p + b^p + c^p + d^p), \tag{36}$$

where (35) follows from Jensen's inequality because $z \mapsto z^p$ is convex on $\mathbb{R}_{\geq 0}$ since $p > 1$. □

Lemma 7 (Rosenthal)**.** *Let $p > 1$, and let $X_1, \ldots, X_n$ be independent random variables that are either zero or one. Then, $X \triangleq \sum_{i=1}^n X_i$ satisfies:*

$$\mathrm{E}[X^p] \leq 2^{p^2} \max\{\mathrm{E}[X], \mathrm{E}[X]^p\}. \tag{37}$$

Proof. This is a special case of [11] (Lemma 1). For convenience, we also provide a self-contained proof:

$$\mathrm{E}[X^p] = \mathrm{E}\left[\sum_{i\in\{1,\ldots,n\}} X_i \cdot \left\{\sum_{j\in\{1,\ldots,n\}} X_j\right\}^{p-1}\right] \tag{38}$$

$$= \mathrm{E}\left[\sum_{i\in\{1,\ldots,n\}} X_i \cdot \left\{1 + \sum_{j\in\{1,\ldots,n\}\setminus\{i\}} X_j\right\}^{p-1}\right] \tag{39}$$

$$= \sum_{i\in\{1,\ldots,n\}} \mathrm{E}\left[X_i \cdot \left\{1 + \sum_{j\in\{1,\ldots,n\}\setminus\{i\}} X_j\right\}^{p-1}\right] \tag{40}$$

$$= \sum_{i\in\{1,\ldots,n\}} \mathrm{E}[X_i] \cdot \mathrm{E}\left[\left\{1 + \sum_{j\in\{1,\ldots,n\}\setminus\{i\}} X_j\right\}^{p-1}\right] \tag{41}$$

$$\leq \sum_{i\in\{1,\ldots,n\}} \mathrm{E}[X_i] \cdot \mathrm{E}\left[\left\{1 + \sum_{j\in\{1,\ldots,n\}} X_j\right\}^{p-1}\right] \tag{42}$$

$$= \mathrm{E}[X] \cdot \mathrm{E}[(1+X)^{p-1}] \tag{43}$$

$$\leq \mathrm{E}[X] \cdot 2^{p-1} \cdot (1 + \mathrm{E}[X^{p-1}]) \tag{44}$$

$$= 2^{p-1}(\mathrm{E}[X] + \mathrm{E}[X]\mathrm{E}[X^{p-1}]) \tag{45}$$

$$\leq 2^{p-1}\left(\mathrm{E}[X] + \mathrm{E}[X]\mathrm{E}[X^p]^{\frac{p-1}{p}}\right) \tag{46}$$

$$\leq 2^p \max\left\{\mathrm{E}[X],\, \mathrm{E}[X]\mathrm{E}[X^p]^{\frac{p-1}{p}}\right\}, \tag{47}$$

where (39) holds because each X_i is either zero or one; (41) holds because $X_1, \ldots, X_n$ are independent; (42) holds because $z \mapsto z^{p-1}$ is increasing on $\mathbb{R}_{\geq 0}$ for $p > 1$; (44) holds because for real numbers $a \geq 0$, $b \geq 0$, and $r > 0$, we have $(a+b)^r \leq (2\max\{a,b\})^r = 2^r \max\{a^r, b^r\} \leq 2^r(a^r + b^r)$; and (46) follows from Jensen's inequality because $z \mapsto z^{(p-1)/p}$ is concave on $\mathbb{R}_{\geq 0}$ for $p > 1$.

We now consider two cases depending on which term on the RHS of (47) achieves the maximum: If the maximum is achieved by $\mathrm{E}[X]$, then $\mathrm{E}[X^p] \leq 2^p \mathrm{E}[X]$, which implies (37) because $2^p \leq 2^{p^2}$ since $p > 1$. If the maximum is achieved by $\mathrm{E}[X]\mathrm{E}[X^p]^{(p-1)/p}$, then:

$$\mathrm{E}[X^p] \leq 2^p \mathrm{E}[X]\mathrm{E}[X^p]^{\frac{p-1}{p}}. \tag{48}$$

Rearranging (48), we obtain:

$$\mathrm{E}[X^p] \leq 2^{p^2} \mathrm{E}[X]^p, \tag{49}$$

so (37) holds also in this case. □

4. Converse

In this section, we prove a nonasymptotic and an asymptotic converse result (Theorem 2 and Corollary 1, respectively).

Theorem 2. *Let $U \multimap X \multimap Y \multimap V$ form a Markov chain over the finite set $\mathcal{U} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V}$, and let $\tau \triangleq 1 + \ln|\mathcal{X} \times \mathcal{Y}|$. Then, for every $\rho > 0$ and for every guesser, the ρth moment of the number of guesses it takes to guess the pair (X, Y) based on the side information (U, V) satisfies:*

$$\mathrm{E}[G(X,Y|U,V)^\rho] \geq \max\Big\{ \begin{array}{l} 2^{\rho(H_{\tilde{\rho}}(X|Y) - \log|\mathcal{U}| - \log\tau)}, \\ 2^{\rho(H_{\tilde{\rho}}(Y|X) - \log|\mathcal{V}| - \log\tau)}, \\ 2^{\rho(H_{\tilde{\rho}}(X,Y) - \log|\mathcal{U}\times\mathcal{V}| - \log\tau)} \end{array} \Big\}. \tag{50}$$

Proof. We view (50) as three lower bounds corresponding to the three terms in the maximization on its RHS. The lower bound involving $H_{\tilde{\rho}}(X,Y)$ holds because:

$$\mathrm{E}[G(X,Y|U,V)^\rho] \geq 2^{\rho(H_{\tilde{\rho}}(X,Y|U,V) - \log\tau)} \tag{51}$$

$$\geq 2^{\rho(H_{\tilde{\rho}}(X,Y) - \log|\mathcal{U}\times\mathcal{V}| - \log\tau)}, \tag{52}$$

where (51) follows from (18) and (52) follows from Lemma 3. The lower bound involving $H_{\tilde{\rho}}(X|Y)$ holds because:

$$\begin{aligned}
\mathrm{E}[G(X,Y|U,V)^\rho] &\geq 2^{\rho(H_{\tilde{\rho}}(X,Y|U,V)-\log\tau)} && (53)\\
&\geq 2^{\rho(H_{\tilde{\rho}}(X,Y|U,V,Y)-\log\tau)} && (54)\\
&= 2^{\rho(H_{\tilde{\rho}}(X|U,V,Y)-\log\tau)} && (55)\\
&= 2^{\rho(H_{\tilde{\rho}}(X|U,Y)-\log\tau)} && (56)\\
&\geq 2^{\rho(H_{\tilde{\rho}}(X|Y)-\log|\mathcal{U}|-\log\tau)}, && (57)
\end{aligned}$$

where (53) follows from (18); (54) follows from Lemma 1; (55) follows from Lemma 2; (56) follows from Lemma 1 because $X \multimap (U,Y) \multimap V$ form a Markov chain; and (57) follows from Lemma 3. The lower bound involving $H_{\tilde{\rho}}(Y|X)$ is analogous to the one with $H_{\tilde{\rho}}(X|Y)$. □

Corollary 1. *For any $\rho > 0$, rate pairs outside $\mathcal{R}(\rho)$ are not achievable.*

Proof. We first show that (8) is necessary for a rate pair $(\mathsf{R_X}, \mathsf{R_Y}) \in \mathbb{R}^2_{\geq 0}$ to be achievable. Indeed, if (8) does not hold, then there exists an $\epsilon > 0$ such that for infinitely many n,

$$\frac{H_{\tilde{\rho}}(X^n, Y^n)}{n} \geq \mathsf{R_X} + \mathsf{R_Y} + \epsilon. \tag{58}$$

Using Theorem 2 with $\mathcal{X}' \triangleq \mathcal{X}^n$, $\mathcal{Y}' \triangleq \mathcal{Y}^n$, $\mathcal{U} \triangleq \{1, \ldots, \lfloor 2^{n\mathsf{R_X}} \rfloor\}$, $\mathcal{V} \triangleq \{1, \ldots, \lfloor 2^{n\mathsf{R_Y}} \rfloor\}$, $P_{X'Y'} \triangleq P_{X^nY^n}$, $U \triangleq f_n(X^n)$, $V \triangleq g_n(Y^n)$, and $\tau_n = 1 + n \ln|\mathcal{X} \times \mathcal{Y}|$ leads to:

$$\begin{aligned}
\mathrm{E}[G(X^n,Y^n|U,V)^\rho] &\geq 2^{\rho(H_{\tilde{\rho}}(X^n,Y^n)-\log|\mathcal{U}\times\mathcal{V}|-\log\tau_n)} && (59)\\
&\geq 2^{\rho n(\frac{1}{n}H_{\tilde{\rho}}(X^n,Y^n)-\mathsf{R_X}-\mathsf{R_Y}-\frac{1}{n}\log\tau_n)}. && (60)
\end{aligned}$$

It follows from (60), (58), and the fact that $\frac{1}{n}\log\tau_n$ tends to zero as n tends to infinity that the LHS of (59) cannot tend to one as n tends to infinity, so $(\mathsf{R_X}, \mathsf{R_Y})$ is not achievable if (8) does not hold. The necessity of (6) and (7) can be shown in the same way. □

5. Achievability

In this section, we prove a nonasymptotic and an asymptotic achievability result (Theorem 3 and Corollary 2, respectively).

Theorem 3. *Let $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{U}$, and $\mathcal{V}$ be finite nonempty sets; let P_{XY} be a PMF; let $\rho > 0$; and let $\epsilon > 0$ be such that:*

$$\begin{aligned}
\log|\mathcal{U}| &\geq H_{\tilde{\rho}}(X|Y) + \epsilon, && (61)\\
\log|\mathcal{V}| &\geq H_{\tilde{\rho}}(Y|X) + \epsilon, && (62)\\
\log|\mathcal{U}\times\mathcal{V}| &\geq H_{\tilde{\rho}}(X,Y) + \epsilon. && (63)
\end{aligned}$$

Then, there exist functions $f\colon \mathcal{X} \to \mathcal{U}$ and $g\colon \mathcal{Y} \to \mathcal{V}$ and a guesser such that the ρth moment of the number of guesses needed to guess the pair (X,Y) based on the side information $(f(X), g(Y))$ satisfies:

$$\mathrm{E}\big[G(X,Y|f(X),g(Y))^\rho\big] \leq \begin{cases} 1 + 4^{\rho+1} \cdot 2^{-\rho\epsilon} & \text{if } \rho \in (0,1], \\ 1 + 4^{(\rho+1)^2} \cdot 2^{-\epsilon} & \text{if } \rho > 1. \end{cases} \tag{64}$$

Proof. Our achievability result relies on random binning: we map each $x \in \mathcal{X}$ uniformly at random to some $u \in \mathcal{U}$ and each $y \in \mathcal{Y}$ uniformly at random to some $v \in \mathcal{V}$. We then show that the ρth moment of the number of guesses averaged over all such mappings $f\colon \mathcal{X} \to \mathcal{U}$ and $g\colon \mathcal{Y} \to \mathcal{V}$ is upper bounded by the RHS of (64). From this, we conclude that there exist f and g that satisfy (64).

Let the guessing function G correspond to guessing in decreasing order of probability [2] (ties can be resolved arbitrarily). Let f and g be distributed as described above, and denote by $\mathrm{E}_{f,g}[\cdot]$ the expectation with respect to f and g. Then,

$$\begin{aligned}
\mathrm{E}_{f,g}\left[\mathrm{E}[G(X,Y|f(X),g(Y))^\rho]\right] &= \sum_{x,y} P(x,y)\,\mathrm{E}_{f,g}[G(x,y|f(x),g(y))^\rho] && (65)\\
&\le \sum_{x,y} P(x,y)\,\mathrm{E}_{f,g}\left[\left\{\sum_{x',y'} \psi(x',y')\phi_f(x')\phi_g(y')\right\}^\rho\right] && (66)\\
&= \sum_{x,y} P(x,y)\,\mathrm{E}_{f,g}[(1+\beta_1+\beta_2+\beta_3)^\rho] && (67)\\
&\le 1 + 4^\rho \sum_{x,y} P(x,y)(\mathrm{E}_{f,g}[\beta_1^\rho] + \mathrm{E}_{f,g}[\beta_2^\rho] + \mathrm{E}_{f,g}[\beta_3^\rho]) && (68)
\end{aligned}$$

with:

$$\begin{aligned}
\psi(x',y') = \psi(x,y,x',y') &\triangleq \mathbb{1}\{P(x',y') \ge P(x,y)\}, && (69)\\
\phi_f(x') = \phi_f(x,x') &\triangleq \mathbb{1}\{f(x') = f(x)\}, && (70)\\
\phi_g(y') = \phi_g(y,y') &\triangleq \mathbb{1}\{g(y') = g(y)\}, && (71)\\
\beta_1 = \beta_1(x,y,f) &\triangleq \sum_{x'\ne x} \psi(x',y)\phi_f(x'), && (72)\\
\beta_2 = \beta_2(x,y,g) &\triangleq \sum_{y'\ne y} \psi(x,y')\phi_g(y'), && (73)\\
\beta_3 = \beta_3(x,y,f,g) &\triangleq \sum_{x'\ne x, y'\ne y} \psi(x',y')\phi_f(x')\phi_g(y'), && (74)
\end{aligned}$$

where $\mathbb{1}\{\cdot\}$ is the indicator function that is one if the condition comprising its argument is true and zero otherwise; (65) holds because (f,g) and (X,Y) are independent; (66) holds because the number of guesses is upper bounded by the number of (x',y') that are at least as likely as (x,y) and that are mapped to the same labels (u,v) as (x,y); (67) follows from splitting the sum depending on whether $x' = x$ or not and whether $y' = y$ or not and from the fact that $\psi(x,y) = \phi_f(x) = \phi_g(y) = 1$; and (68) follows from Lemma 5 because β_1, β_2, and β_3 are nonnegative integers. As indicated in (69)–(74), the dependence of ψ, ϕ_f, ϕ_g, β_1, β_2, and β_3 on x, y, f, and g is implicit in our notation.

We first treat the case $\rho \in (0,1]$. We bound the terms on the RHS of (68) as follows:

$$\begin{aligned}
\sum_{x,y} P(x,y)\,\mathrm{E}_{f,g}[\beta_1^\rho] &\le \sum_{x,y} P(x,y)\,\mathrm{E}_{f,g}[\beta_1]^\rho && (75)\\
&= \sum_{x,y} P(x,y)\left[\sum_{x'\ne x} \psi(x',y)\,\frac{1}{|\mathcal{U}|}\right]^\rho && (76)\\
&\le \sum_{x,y} P(x,y)\left[\sum_{x'}\left[\frac{P(x',y)}{P(x,y)}\right]^{\tilde\rho}\frac{1}{|\mathcal{U}|}\right]^\rho && (77)\\
&= \frac{1}{|\mathcal{U}|^\rho}\sum_{x,y} P(x,y)^{\tilde\rho}\left[\sum_{x'} P(x',y)^{\tilde\rho}\right]^\rho && (78)\\
&= \frac{1}{|\mathcal{U}|^\rho}\sum_{y}\left[\sum_{x} P(x,y)^{\tilde\rho}\right]\left[\sum_{x'} P(x',y)^{\tilde\rho}\right]^\rho && (79)\\
&= \frac{1}{|\mathcal{U}|^\rho}\sum_{y}\left[\sum_{x} P(x,y)^{\tilde\rho}\right]^{1+\rho} && (80)\\
&= 2^{\rho(H_{\tilde\rho}(X|Y) - \log|\mathcal{U}|)} && (81)
\end{aligned}$$

$$\leq 2^{-\rho\epsilon}, \tag{82}$$

where (75) follows from Jensen's inequality because $z \mapsto z^\rho$ is concave on $\mathbb{R}_{\geq 0}$ since $\rho \in (0,1]$; (76) holds because the expectation operator is linear and because $\mathrm{E}_{f,g}[\phi_f(x')] = 1/|\mathcal{U}|$ since $x' \neq x$; in (77), we extended the inner summation and used that $\psi(x',y) \leq [P(x',y)/P(x,y)]^{\tilde{\rho}}$; and (82) follows from (61). In the same way, we obtain:

$$\sum_{x,y} P(x,y)\,\mathrm{E}_{f,g}[\beta_2^\rho] \leq 2^{-\rho\epsilon}. \tag{83}$$

Similarly,

$$\begin{aligned}
\sum_{x,y} P(x,y)\,\mathrm{E}_{f,g}[\beta_3^\rho] &\leq \sum_{x,y} P(x,y)\,\mathrm{E}_{f,g}[\beta_3]^\rho && (84)\\
&= \sum_{x,y} P(x,y) \left[\sum_{x' \neq x, y' \neq y} \psi(x',y') \frac{1}{|\mathcal{U}\times\mathcal{V}|} \right]^\rho && (85)\\
&\leq \sum_{x,y} P(x,y) \left[\sum_{x',y'} \left[\frac{P(x',y')}{P(x,y)} \right]^{\tilde{\rho}} \frac{1}{|\mathcal{U}\times\mathcal{V}|} \right]^\rho && (86)\\
&= \frac{1}{|\mathcal{U}\times\mathcal{V}|^\rho} \sum_{x,y} P(x,y)^{\tilde{\rho}} \left[\sum_{x',y'} P(x',y')^{\tilde{\rho}} \right]^\rho && (87)\\
&= \frac{1}{|\mathcal{U}\times\mathcal{V}|^\rho} \left[\sum_{x,y} P(x,y)^{\tilde{\rho}} \right]^{1+\rho} && (88)\\
&= 2^{\rho(H_{\tilde{\rho}}(X,Y) - \log|\mathcal{U}\times\mathcal{V}|)} && (89)\\
&\leq 2^{-\rho\epsilon}. && (90)
\end{aligned}$$

From (68), (82), (83), and (90), we obtain:

$$\begin{aligned}
\mathrm{E}_{f,g}\left[\mathrm{E}[G(X,Y|f(X),g(Y))^\rho]\right] &\leq 1 + 3\cdot 4^\rho \cdot 2^{-\rho\epsilon} && (91)\\
&\leq 1 + 4^{\rho+1} \cdot 2^{-\rho\epsilon} && (92)
\end{aligned}$$

and hence infer the existence of $f\colon \mathcal{X} \to \mathcal{U}$ and $g\colon \mathcal{Y} \to \mathcal{V}$ satisfying (64).

We now consider (68) when $\rho > 1$. Unlike in the case $\rho \in (0,1]$, we cannot use Jensen's inequality as we did in (75). Instead, for fixed $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, we upper-bound the first expectation on the RHS of (68) by:

$$\begin{aligned}
\mathrm{E}_{f,g}[\beta_1^\rho] &\leq 2^{\rho^2} \max\left\{\mathrm{E}_{f,g}[\beta_1],\, \mathrm{E}_{f,g}[\beta_1]^\rho\right\} && (93)\\
&\leq 2^{\rho^2} \left(\mathrm{E}_{f,g}[\beta_1]^\rho + \mathrm{E}_{f,g}[\beta_1]\right), && (94)
\end{aligned}$$

where (93) follows from Lemma 7 because $\rho > 1$ and because β_1 is a sum of independent random variables taking values in $\{0,1\}$. By the same steps as in (76)–(82),

$$\sum_{x,y} P(x,y)\,\mathrm{E}_{f,g}[\beta_1]^\rho \leq 2^{-\rho\epsilon}. \tag{95}$$

As to the expectation of the other term on the RHS of (94),

$$\begin{aligned}
\sum_{x,y} P(x,y)\,\mathrm{E}_{f,g}[\beta_1] &\leq \left[\sum_{x,y} P(x,y)\,\mathrm{E}_{f,g}[\beta_1]^\rho \right]^{\frac{1}{\rho}} && (96)\\
&\leq 2^{-\epsilon}, && (97)
\end{aligned}$$

where (96) follows from Jensen's inequality because $z \mapsto z^{\frac{1}{\rho}}$ is concave on $\mathbb{R}_{\geq 0}$ since $\rho > 1$, and (97) follows from (95). From (94), (95), and (97), we obtain:

$$\sum_{x,y} P(x,y)\,\mathrm{E}_{f,g}[\beta_1^\rho] \leq 2^{\rho^2}(2^{-\rho\epsilon} + 2^{-\epsilon}) \tag{98}$$

$$\leq 2^{\rho^2+1} \cdot 2^{-\epsilon}, \tag{99}$$

where (99) holds because $2^{-\rho\epsilon} \leq 2^{-\epsilon}$ since $\rho > 1$ and $\epsilon > 0$. In the same way, we obtain for the second expectation on the RHS of (68):

$$\sum_{x,y} P(x,y)\,\mathrm{E}_{f,g}[\beta_2^\rho] \leq 2^{\rho^2+1} \cdot 2^{-\epsilon}. \tag{100}$$

Bounding $\mathrm{E}_{f,g}[\beta_3^\rho]$, i.e., the third expectation on the RHS of (68), is more involved because β_3 is not a sum of independent random variables. Our approach builds on the ideas used by Rosenthal [11] (Proof of Lemma 1); compare (47) and (48) with (108) and (123) ahead. For fixed $x \in \mathcal{X}$ and $y \in \mathcal{Y}$,

$$\mathrm{E}_{f,g}[\beta_3^\rho] = \mathrm{E}_{f,g}\left[\sum_{x' \neq x, y' \neq y} \psi(x',y')\phi_f(x')\phi_g(y') \cdot \left\{\sum_{\tilde{x} \neq x, \tilde{y} \neq y} \psi(\tilde{x},\tilde{y})\phi_f(\tilde{x})\phi_g(\tilde{y})\right\}^{\rho-1}\right] \tag{101}$$

$$= \mathrm{E}_{f,g}\left[\sum_{x' \neq x, y' \neq y} \psi(x',y')\phi_f(x')\phi_g(y') \cdot (1+\gamma_1+\gamma_2+\gamma_3)^{\rho-1}\right] \tag{102}$$

$$= \sum_{x' \neq x, y' \neq y} \mathrm{E}_{f,g}\left[\psi(x',y')\phi_f(x')\phi_g(y') \cdot (1+\gamma_1+\gamma_2+\gamma_3)^{\rho-1}\right] \tag{103}$$

$$= \sum_{x' \neq x, y' \neq y} \mathrm{E}_{f,g}\left[\psi(x',y')\phi_f(x')\phi_g(y')\right] \cdot \mathrm{E}_{f,g}\left[(1+\gamma_1+\gamma_2+\gamma_3)^{\rho-1}\right] \tag{104}$$

$$\leq \sum_{x' \neq x, y' \neq y} \mathrm{E}_{f,g}\left[\psi(x',y')\phi_f(x')\phi_g(y')\right] \cdot \mathrm{E}_{f,g}\left[(1+\delta_1+\delta_2+\beta_3)^{\rho-1}\right] \tag{105}$$

$$\leq \sum_{x' \neq x, y' \neq y} \mathrm{E}_{f,g}\left[\psi(x',y')\phi_f(x')\phi_g(y')\right] \cdot 4^{\rho-1} \cdot \mathrm{E}_{f,g}\left[1+\delta_1^{\rho-1}+\delta_2^{\rho-1}+\beta_3^{\rho-1}\right] \tag{106}$$

$$= 4^{\rho-1}\Bigg\{\mathrm{E}_{f,g}[\beta_3] + \sum_{y' \neq y} \frac{1}{|\mathcal{V}|}\mathrm{E}_{f,g}[\delta_1]\mathrm{E}_{f,g}[\delta_1^{\rho-1}] + \sum_{x' \neq x} \frac{1}{|\mathcal{U}|}\mathrm{E}_{f,g}[\delta_2]\mathrm{E}_{f,g}[\delta_2^{\rho-1}] + \mathrm{E}_{f,g}[\beta_3]\mathrm{E}_{f,g}[\beta_3^{\rho-1}]\Bigg\} \tag{107}$$

$$\leq 4^{\rho} \max\Bigg\{\mathrm{E}_{f,g}[\beta_3], \sum_{y' \neq y} \frac{1}{|\mathcal{V}|}\mathrm{E}_{f,g}[\delta_1]\mathrm{E}_{f,g}[\delta_1^{\rho-1}], \sum_{x' \neq x} \frac{1}{|\mathcal{U}|}\mathrm{E}_{f,g}[\delta_2]\mathrm{E}_{f,g}[\delta_2^{\rho-1}], \mathrm{E}_{f,g}[\beta_3]\mathrm{E}_{f,g}[\beta_3^{\rho-1}]\Bigg\} \tag{108}$$

with:

$$\gamma_1 = \gamma_1(x,y,x',y',f) \triangleq \sum_{\tilde{x} \notin \{x,x'\}} \psi(\tilde{x},y')\phi_f(\tilde{x}), \tag{109}$$

$$\gamma_2 = \gamma_2(x,y,x',y',g) \triangleq \sum_{\tilde{y} \notin \{y,y'\}} \psi(x',\tilde{y})\phi_g(\tilde{y}), \tag{110}$$

$$\gamma_3 = \gamma_3(x,y,x',y',f,g) \triangleq \sum_{\tilde{x} \notin \{x,x'\}, \tilde{y} \notin \{y,y'\}} \psi(\tilde{x},\tilde{y})\phi_f(\tilde{x})\phi_g(\tilde{y}), \tag{111}$$

$$\delta_1 = \delta_1(x,y,y',f) \triangleq \sum_{\tilde{x} \neq x} \psi(\tilde{x},y')\phi_f(\tilde{x}), \tag{112}$$

$$\delta_2 = \delta_2(x, y, x', g) \triangleq \sum_{\tilde{y} \neq y} \psi(x', \tilde{y}) \phi_g(\tilde{y}), \tag{113}$$

where (102) follows from splitting the sum in braces depending on whether $\tilde{x} = x'$ or not and whether $\tilde{y} = y'$ or not and from assuming $\psi(x', y') = \phi_f(x') = \phi_g(y') = 1$ within the braces, which does not change the value of the expression because it is multiplied by $\psi(x', y')\phi_f(x')\phi_g(y')$; (104) holds because $(\phi_f(x'), \phi_g(y'))$ and $(\gamma_1, \gamma_2, \gamma_3)$ are independent since $\tilde{x} \neq x'$ and $\tilde{y} \neq y'$; (105) holds because $\rho - 1 > 0$, $\gamma_1 \leq \delta_1$, $\gamma_2 \leq \delta_2$, and $\gamma_3 \leq \beta_3$; (106) follows from Lemma 6; and (107) follows from identifying $\mathrm{E}_{f,g}[\beta_3]$, $\mathrm{E}_{f,g}[\delta_1]$, and $\mathrm{E}_{f,g}[\delta_2]$ because $\phi_f(x')$ and $\phi_g(y')$ are independent, $\mathrm{E}_{f,g}[\phi_f(x')] = 1/|\mathcal{U}|$, and $\mathrm{E}_{f,g}[\phi_g(y')] = 1/|\mathcal{V}|$. As indicated in (109)–(113), the dependence of γ_1, γ_2, γ_3, δ_1, and δ_2 on x, y, x', y', f, and g is implicit in our notation.

To bound $\mathrm{E}_{f,g}[\beta_3^\rho]$ further, we study some of the terms on the RHS of (108) separately, starting with the second, which involves the sum over y'. For fixed $x \in \mathcal{X}$, $y \in \mathcal{Y}$, and $y' \in \mathcal{Y} \setminus \{y\}$,

$$\begin{aligned}
\mathrm{E}_{f,g}[\delta_1]\mathrm{E}_{f,g}[\delta_1^{\rho-1}] &\leq \mathrm{E}_{f,g}[\delta_1^\rho]^{\frac{1}{\rho}} \mathrm{E}_{f,g}[\delta_1^\rho]^{\frac{\rho-1}{\rho}} && (114)\\
&= \mathrm{E}_{f,g}[\delta_1^\rho] && (115)\\
&\leq 2^{\rho^2} \max\{\mathrm{E}_{f,g}[\delta_1], \mathrm{E}_{f,g}[\delta_1]^\rho\} && (116)\\
&\leq 2^{\rho^2} \left(\mathrm{E}_{f,g}[\delta_1] + \mathrm{E}_{f,g}[\delta_1]^\rho\right), && (117)
\end{aligned}$$

where (114) follows from Jensen's inequality because $z \mapsto z^{\frac{1}{\rho}}$ and $z \mapsto z^{\frac{\rho-1}{\rho}}$ are both concave on $\mathbb{R}_{\geq 0}$ since $\rho > 1$, and (116) follows from Lemma 7 because $\rho > 1$ and because δ_1 is a sum of independent random variables taking values in $\{0, 1\}$. This implies that for fixed $x \in \mathcal{X}$ and $y \in \mathcal{Y}$,

$$\begin{aligned}
\sum_{y' \neq y} \frac{1}{|\mathcal{V}|} \mathrm{E}_{f,g}[\delta_1]\mathrm{E}_{f,g}[\delta_1^{\rho-1}] &\leq 2^{\rho^2} \sum_{y' \neq y} \frac{1}{|\mathcal{V}|} \left(\mathrm{E}_{f,g}[\delta_1] + \mathrm{E}_{f,g}[\delta_1]^\rho\right) && (118)\\
&= 2^{\rho^2} \mathrm{E}_{f,g}[\beta_3] + 2^{\rho^2} \sum_{y' \neq y} \frac{1}{|\mathcal{V}|} \mathrm{E}_{f,g}[\delta_1]^\rho, && (119)
\end{aligned}$$

where (119) follows from the definitions of δ_1 and β_3. Similarly, for the third term on the RHS of (108),

$$\sum_{x' \neq x} \frac{1}{|\mathcal{U}|} \mathrm{E}_{f,g}[\delta_2]\mathrm{E}_{f,g}[\delta_2^{\rho-1}] \leq 2^{\rho^2} \mathrm{E}_{f,g}[\beta_3] + 2^{\rho^2} \sum_{x' \neq x} \frac{1}{|\mathcal{U}|} \mathrm{E}_{f,g}[\delta_2]^\rho. \tag{120}$$

With the help of (119) and (120), we now go back to (108) and argue that it implies that for fixed $x \in \mathcal{X}$ and $y \in \mathcal{Y}$,

$$\mathrm{E}_{f,g}[\beta_3^\rho] \leq 2 \cdot 4^{\rho^2} \left[\mathrm{E}_{f,g}[\beta_3] + \sum_{y' \neq y} \frac{1}{|\mathcal{V}|} \mathrm{E}_{f,g}[\delta_1]^\rho + \sum_{x' \neq x} \frac{1}{|\mathcal{U}|} \mathrm{E}_{f,g}[\delta_2]^\rho + \mathrm{E}_{f,g}[\beta_3]^\rho\right]. \tag{121}$$

To prove this, we consider four cases depending on which term on the RHS of (108) achieves the maximum: If $\mathrm{E}_{f,g}[\beta_3]$ achieves the maximum, then (121) holds because $4^\rho \leq 2 \cdot 4^{\rho^2}$. If the LHS of (118) achieves the maximum, then (121) follows from (119) because $4^\rho \cdot 2^{\rho^2} \leq 2 \cdot 4^{\rho^2}$. If the LHS of (120) achieves the maximum, then (121) follows similarly. Finally, if $\mathrm{E}_{f,g}[\beta_3]\mathrm{E}_{f,g}[\beta_3^{\rho-1}]$ achieves the maximum, then:

$$\begin{aligned}
\mathrm{E}_{f,g}[\beta_3^\rho] &\leq 4^\rho \mathrm{E}_{f,g}[\beta_3]\mathrm{E}_{f,g}[\beta_3^{\rho-1}] && (122)\\
&\leq 4^\rho \mathrm{E}_{f,g}[\beta_3]\mathrm{E}_{f,g}[\beta_3^\rho]^{\frac{\rho-1}{\rho}}, && (123)
\end{aligned}$$

where (123) follows from Jensen's inequality because $z \mapsto z^{\frac{\rho-1}{\rho}}$ is concave on $\mathbb{R}_{\geq 0}$ for $\rho > 1$. Rearranging (123), we obtain:

$$\mathrm{E}_{f,g}[\beta_3^\rho] \leq 4^{\rho^2}\,\mathrm{E}_{f,g}[\beta_3]^\rho, \tag{124}$$

so (121) holds also in this case.

Having established (121), we now take the expectation of its sides to obtain:

$$\textstyle\sum_{x,y} P(x,y)\,\mathrm{E}_{f,g}[\beta_3^\rho] \leq 2\cdot 4^{\rho^2} \sum_{x,y} P(x,y)\left[\mathrm{E}_{f,g}[\beta_3] + \sum_{y'\neq y}\frac{1}{|\mathcal{V}|}\mathrm{E}_{f,g}[\delta_1]^\rho + \sum_{x'\neq x}\frac{1}{|\mathcal{U}|}\mathrm{E}_{f,g}[\delta_2]^\rho + \mathrm{E}_{f,g}[\beta_3]^\rho\right]. \tag{125}$$

We now study the terms on the RHS of (125) separately, starting with the fourth (last). By (85)–(90), which hold also if $\rho > 1$,

$$\sum_{x,y} P(x,y)\,\mathrm{E}_{f,g}[\beta_3]^\rho \leq 2^{-\rho\epsilon}. \tag{126}$$

As for the first term on the RHS of (125),

$$\sum_{x,y} P(x,y)\,\mathrm{E}_{f,g}[\beta_3] \leq 2^{-\epsilon}, \tag{127}$$

which follows from (126) in the same way as (97) followed from (95). As for the second term on the RHS of (125),

$$\begin{aligned}
&\sum_{x,y} P(x,y) \sum_{y'\neq y}\frac{1}{|\mathcal{V}|}\mathrm{E}_{f,g}[\delta_1]^\rho \\
&\quad= \sum_{x,y} P(x,y) \sum_{y'\neq y}\frac{1}{|\mathcal{V}|}\left[\sum_{x'\neq x}\psi(x',y')\frac{1}{|\mathcal{U}|}\right]^\rho && (128)\\
&\quad\leq \sum_{x,y} P(x,y)^{\tilde\rho} \sum_{y'}\frac{1}{|\mathcal{V}|}\left[\sum_{x'} P(x',y')^{\tilde\rho}\frac{1}{|\mathcal{U}|}\right]^\rho && (129)\\
&\quad= \sum_{x,y} P(x,y)^{\tilde\rho} \sum_{y'}\left[\sum_{x'} P(x',y')^{\tilde\rho}\frac{1}{|\mathcal{U}\times\mathcal{V}|^\rho}\right]^{\frac{1}{\rho}} \cdot \left[\sum_{x'} P(x',y')^{\tilde\rho}\frac{1}{|\mathcal{U}|^{\frac{\rho}{1+\rho}}}\right]^{(1+\rho)\cdot\frac{\rho-1}{\rho}} && (130)\\
&\quad\leq \sum_{x,y} P(x,y)^{\tilde\rho}\left\{\sum_{y'}\sum_{x'} P(x',y')^{\tilde\rho}\frac{1}{|\mathcal{U}\times\mathcal{V}|^\rho}\right\}^{\frac{1}{\rho}} \cdot \left\{\sum_{y'}\left[\sum_{x'} P(x',y')^{\tilde\rho}\frac{1}{|\mathcal{U}|^{\frac{\rho}{1+\rho}}}\right]^{1+\rho}\right\}^{\frac{\rho-1}{\rho}} && (131)\\
&\quad= \left\{\frac{1}{|\mathcal{U}\times\mathcal{V}|^\rho}\left[\sum_{x,y} P(x,y)^{\tilde\rho}\right]^{1+\rho}\right\}^{\frac{1}{\rho}} \cdot \left\{\frac{1}{|\mathcal{U}|^\rho}\sum_{y'}\left[\sum_{x'} P(x',y')^{\tilde\rho}\right]^{1+\rho}\right\}^{\frac{\rho-1}{\rho}} && (132)\\
&\quad\leq (2^{-\rho\epsilon})^{\frac{1}{\rho}} \cdot (2^{-\rho\epsilon})^{\frac{\rho-1}{\rho}} && (133)\\
&\quad= 2^{-\rho\epsilon}, && (134)
\end{aligned}$$

where in (129), we extended the inner summations and used that $\psi(x',y') \leq [P(x',y')/P(x,y)]^{\tilde\rho}$; (131) follows from Hölder's inequality; and (133) follows from (89)–(90) and (81)–(82). In the same way, we obtain for the third term on the RHS of (125):

$$\sum_{x,y} P(x,y) \sum_{x'\neq x}\frac{1}{|\mathcal{U}|}\mathrm{E}_{f,g}[\delta_2]^\rho \leq 2^{-\rho\epsilon}. \tag{135}$$

From (125), (127), (134), (135), and (126), we deduce:

$$\sum_{x,y} P(x,y)\,\mathrm{E}_{f,g}[\beta_3^\rho] \le 2\cdot 4^{\rho^2}\left(2^{-\epsilon}+2^{-\rho\epsilon}+2^{-\rho\epsilon}+2^{-\rho\epsilon}\right) \tag{136}$$

$$\le 8\cdot 4^{\rho^2}\cdot 2^{-\epsilon}, \tag{137}$$

where (137) holds because $2^{-\rho\epsilon} \le 2^{-\epsilon}$ since $\rho > 1$ and $\epsilon > 0$. Finally, (68), (99), (100), and (137) imply:

$$\mathrm{E}_{f,g}\left[\mathrm{E}[G(X,Y|f(X),g(Y))^\rho]\right] \le 1 + 4^\rho\left(2\cdot 2^{\rho^2+1}\cdot 2^{-\epsilon} + 8\cdot 4^{\rho^2}\cdot 2^{-\epsilon}\right) \tag{138}$$

$$\le 1 + 4^{(\rho+1)^2}\cdot 2^{-\epsilon} \tag{139}$$

and thus prove the existence of $f\colon \mathcal{X} \to \mathcal{U}$ and $g\colon \mathcal{Y} \to \mathcal{V}$ satisfying (64). □

Corollary 2. *For any $\rho > 0$, rate pairs in the interior of $\mathcal{R}(\rho)$ are achievable.*

Proof. Let $(\mathsf{R_X}, \mathsf{R_Y})$ be in the interior of $\mathcal{R}(\rho)$. Then, (6)–(8) hold with strict inequalities, and there exists a $\delta > 0$ such that for all sufficiently large n,

$$\log\lfloor 2^{n\mathsf{R_X}}\rfloor \ge H_{\tilde\rho}(X^n|Y^n) + n\delta, \tag{140}$$

$$\log\lfloor 2^{n\mathsf{R_Y}}\rfloor \ge H_{\tilde\rho}(Y^n|X^n) + n\delta, \tag{141}$$

$$\log\lfloor 2^{n\mathsf{R_X}}\rfloor + \log\lfloor 2^{n\mathsf{R_Y}}\rfloor \ge H_{\tilde\rho}(X^n,Y^n) + n\delta. \tag{142}$$

Using Theorem 3 with $\mathcal{X}' \triangleq \mathcal{X}^n$, $\mathcal{Y}' \triangleq \mathcal{Y}^n$, $\mathcal{U} \triangleq \{1,\dots,\lfloor 2^{n\mathsf{R_X}}\rfloor\}$, $\mathcal{V} \triangleq \{1,\dots,\lfloor 2^{n\mathsf{R_Y}}\rfloor\}$, $P_{X'Y'} \triangleq P_{X^nY^n}$, and $\epsilon_n \triangleq n\delta$ shows that, for all sufficiently large n, there exist encoders $f_n\colon \mathcal{X}^n \to \mathcal{U}$ and $g_n\colon \mathcal{Y}^n \to \mathcal{V}$ and a guessing function G_n satisfying:

$$\mathrm{E}\left[G_n\left(X^n,Y^n|f_n(X^n),g_n(Y^n)\right)^\rho\right] \le \begin{cases} 1+4^{\rho+1}\cdot 2^{-\rho\epsilon_n} & \text{if } \rho\in(0,1],\\ 1+4^{(\rho+1)^2}\cdot 2^{-\epsilon_n} & \text{if } \rho>1. \end{cases} \tag{143}$$

Because ϵ_n tends to infinity as n tends to infinity, the RHS of (143) tends to one as n tends to infinity, which implies that the rate pair $(\mathsf{R_X}, \mathsf{R_Y})$ is achievable. □

Author Contributions: Writing—original draft preparation, A.B., A.L. and C.P.; writing—review and editing, A.B., A.L. and C.P.

Funding: This research received no external funding.

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

References

1. Massey, J.L. Guessing and entropy. In Proceedings of the 1994 IEEE International Symposium on Information Theory (ISIT), Trondheim, Norway, 27 June–1 July 1994; p. 204. [CrossRef]
2. Arıkan, E. An inequality on guessing and its application to sequential decoding. *IEEE Trans. Inf. Theory* **1996**, *42*, 99–105. [CrossRef]
3. 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]
4. Bracher, A.; Hof, E.; Lapidoth, A. Guessing attacks on distributed-storage systems. *arXiv* **2017**, arXiv:1701.01981v1.
5. Graczyk, R.; Lapidoth, A. Variations on the guessing problem. In Proceedings of the 2018 IEEE International Symposium on Information Theory (ISIT), Vail, CO, USA, 17–22 June 2018; pp. 231–235. [CrossRef]
6. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2006; ISBN 978-0-471-24195-9.
7. Fehr, S.; Berens, S. On the conditional Rényi entropy. *IEEE Trans. Inf. Theory* **2014**, *60*, 6801–6810. [CrossRef]
8. Csiszár, I. Generalized cutoff rates and Rényi's information measures. *IEEE Trans. Inf. Theory* **1995**, *41*, 26–34. [CrossRef]

9. Bracher, A.; Lapidoth, A.; Pfister, C. Distributed task encoding. In Proceedings of the 2017 IEEE International Symposium on Information Theory (ISIT), Aachen, Germany, 25–30 June 2017; pp. 1993–1997. [CrossRef]
10. Lapidoth, A.; Pfister, C. Two measures of dependence. In Proceedings of the 2016 IEEE International Conference on the Science of Electrical Engineering (ICSEE), Eilat, Israel, 16–18 November 2016; pp. 1–5. [CrossRef]
11. Rosenthal, H.P. On the subspaces of $L^p(p > 2)$ spanned by sequences of independent random variables. *Isr. J. Math.* **1970**, *8*, 273–303. [CrossRef]
12. Boztaş, S. Comments on "An inequality on guessing and its application to sequential decoding". *IEEE Trans. Inf. Theory* **1997**, *43*, 2062–2063. [CrossRef]
13. Hanawal, M.K.; Sundaresan, R. Guessing revisited: A large deviations approach. *IEEE Trans. Inf. Theory* **2011**, *57*, 70–78. [CrossRef]
14. Christiansen, M.M.; Duffy, K.R. Guesswork, large deviations, and Shannon entropy. *IEEE Trans. Inf. Theory* **2013**, *59*, 796–802. [CrossRef]
15. Sundaresan, R. Guessing based on length functions. In Proceedings of the 2007 IEEE International Symposium on Information Theory (ISIT), Nice, France, 24–29 June 2007; pp. 716–719. [CrossRef]
16. Sason, I. Tight bounds on the Rényi entropy via majorization with applications to guessing and compression. *Entropy* **2018**, *20*, 896. [CrossRef]
17. Sundaresan, R. Guessing under source uncertainty. *IEEE Trans. Inf. Theory* **2007**, *53*, 269–287. [CrossRef]
18. Arıkan, E.; Merhav, N. Guessing subject to distortion. *IEEE Trans. Inf. Theory* **1998**, *44*, 1041–1056. [CrossRef]
19. Bunte, C.; Lapidoth, A. On the listsize capacity with feedback. *IEEE Trans. Inf. Theory* **2014**, *60*, 6733–6748. [CrossRef]
20. Gallager, R.G. *Information Theory and Reliable Communication*; John Wiley & Sons: Hoboken, NJ, USA, 1968; ISBN 0-471-29048-3.
21. Arıkan, E.; Merhav, N. Joint source-channel coding and guessing with application to sequential decoding. *IEEE Trans. Inf. Theory* **1998**, *44*, 1756–1769. [CrossRef]
22. Bunte, C.; Lapidoth, A. Encoding tasks and Rényi entropy. *IEEE Trans. Inf. Theory* **2014**, *60*, 5065–5076. [CrossRef]
23. Rényi, A. On measures of entropy and information. In Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA, USA, 20 June–30 July 1960; Volume 1, pp. 547–561.
24. Arimoto, S. Information measures and capacity of order α for discrete memoryless channels. In *Topics in Information Theory*; Csiszár, I., Elias, P., Eds.; North-Holland Publishing Company: Amsterdam, The Netherlands, 1977; pp. 41–52, ISBN 0-7204-0699-4.
25. Sason, I.; Verdú, S. Arimoto–Rényi conditional entropy and Bayesian *M*-Ary hypothesis testing. *IEEE Trans. Inf. Theory* **2018**, *64*, 4–25. [CrossRef]