Guessing with a bit of help

entropy

MDPI

Article

Guessing with a Bit of Help †

Nir Weinberger [1,*] and Ofer Shayevitz [2]

[1] Institute for Data, Systems, and Society and Laboratory for Information & Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[2] Department of Electrical Engineering-Systems, Tel Aviv University, Tel Aviv 69978, Israel; ofersha@eng.tau.ac.il
* Correspondence: nir.wein@gmail.com
† This paper is an extended version of our paper published in proceedings of IEEE International Symposium on Information Theory (ISIT), Vail, CO, USA, 17–22 June 2018.

Received: 29 August 2019; Accepted: 23 December 2019; Published: 26 December 2019

Abstract: What is the value of just a few bits to a guesser? We study this problem in a setup where Alice wishes to guess an independent and identically distributed (i.i.d.) random vector and can procure a fixed number of k information bits from Bob, who has observed this vector through a memoryless channel. We are interested in the *guessing ratio*, which we define as the ratio of Alice's guessing-moments with and without observing Bob's bits. For the case of a uniform binary vector observed through a binary symmetric channel, we provide two upper bounds on the guessing ratio by analyzing the performance of the dictator (for general $k \geq 1$) and majority functions (for $k = 1$). We further provide a lower bound via maximum entropy (for general $k \geq 1$) and a lower bound based on Fourier-analytic/hypercontractivity arguments (for $k = 1$). We then extend our maximum entropy argument to give a lower bound on the guessing ratio for a general channel with a binary uniform input that is expressed using the strong data-processing inequality constant of the reverse channel. We compute this bound for the binary erasure channel and conjecture that greedy dictator functions achieve the optimal guessing ratio.

Keywords: boolean functions; fourier analysis; guessing moments; guessing with a helper; hypercontractivity; maximum entropy; strong data-processing inequalities

1. Introduction

In the classical guessing problem, Alice wishes to learn the value of a discrete random variable (r.v.) X as quickly as possible by sequentially asking yes/no questions of the form "Is $X = x$?", until she makes a correct guess. A guessing strategy corresponds to an ordering of the alphabet of X according to which the guesses are made and induces a random guessing time. It is well known and simple to verify that the guessing strategy which simultaneously minimizes all the positive moments of the guessing time is to order the alphabet according to a decreasing order of probability. Formally, for any $s > 0$, the *minimal sth-order guessing-time moment* of X is

$$G_s(X) := \mathbb{E}\left(\mathrm{ORD}_X^s(X)\right), \tag{1}$$

where $\mathrm{ORD}_X(x)$ returns the index of the symbol x relative to the order induced by sorting the probabilities in a descending order, with ties broken arbitrarily. For brevity, we refer to $G_s(X)$ as the *guessing-moment* of X.

Several motivating problems for studying guesswork are fairness in betting games, computational complexity of sequential decoding [1], computational complexity of lossy source coding and database search algorithms (see the introduction of Reference [2] for a discussion), secrecy systems [3–5], and

crypt-analysis (password cracking) [6,7]. The guessing problem was first introduced and studied in an information-theoretic framework by Massey [8], who drew a relation between the average guessing time of an r.v. to its entropy. It was later explored more systematically by Arikan [1], who also introduced the problem of guessing with side information. In this problem, Alice is in possession of another r.v. Y that is jointly distributed with X, and then, the optimal conditional guessing strategy is to guess by decreasing order of conditional probabilities. Hence, the associated *minimal conditional sth-order guessing-time moment of X given Y* is

$$G_s(X|Y) := \mathbb{E}\left(\mathrm{ORD}^s_{X|Y}(X \mid Y)\right), \tag{2}$$

where $\mathrm{ORD}_{X|Y}(x \mid y)$ returns the index of x relative to the order induced by sorting the conditional probabilities of X given that $Y = y$ in a descending order. Arikan showed that, as intuition suggests, side information reduces the guessing-moments ([1], Corollary 1)

$$G_s(X|Y) \leq G_s(X). \tag{3}$$

Furthermore, he showed that, if $\{(X_i, Y_i)\}_{i=1}^n$ is an i.i.d. sequence, then ([1], Proposition 5)

$$\lim_{n\to\infty} \frac{1}{n} \log G_s^{1/s}(X^n|Y^n) = H_{\frac{1}{1+s}}(X_1 \mid Y_1), \tag{4}$$

where $H_\alpha(X \mid Y)$ is the Arimoto-Rényi conditional entropy of order α. As was noted by Arikan a few years later [9], the guessing moments are related to the large deviations behavior of the random variable $\frac{1}{n} \log \mathrm{ORD}_{X^n|Y^n}(X^n \mid Y^n)$. However, in Reference [9], he was only able to obtain right-tail large deviation bounds since asymptotically tight bounds on $G_s(X^n \mid Y^n)$ were only known for positive moments ($s > 0$). Large deviation principle for the normalized logarithm of the guessing time was later established in Reference [10] using substantial results from References [11,12]. Throughout the years, information-theoretic analysis of the guessing problem was extended in multiple directions, such as guessing until the distortion between the guess and the true value is below a certain threshold [2], guessing under source uncertainty [13], and improved bounds at finite blocklength [14–16], to name a few.

In the conditional setting described above, one may think of Y^n as side information observed by a "helper", say Bob, who sends his observations to Alice. Nonetheless, as other problems employing a helper (e.g., source coding [17,18]), it is more realistic to impose communication constraints and to assume that Bob can only send a compressed description of Y^n to Alice. This setting was recently addressed by Graczyk and Lapidoth [19,20], who considered the case where Bob encodes Y^n at a positive rate using nR bits before sending this description to Alice. They then characterized the best possible guessing-moments attained by Alice for general distributions as a function of the rate R. In this paper, we take this setting to its extreme and attempt to quantify the value of *k bits* in terms of reducing the guessing-moments by allowing Bob to use only a k-bit description of Y^n. The major difference from previous work is that, here, k is finite and does not increase with n, and for some of our results, we further concentrate on the extreme case of $k = 1$—a single bit of help. To that end, we define (Section 2) the *guessing ratio*, which is the (asymptotically) best possible ratio of the guessing-moments of X^n obtained with and without observing a function $f(Y^n) \in \{0,1\}^k$, i.e., the minimal possible ratio $G_s(X^n \mid f(Y^n))/G_s(X^n)$ as a function of $s > 0$, in the limit of large n.

Sharply characterizing the guessing ratio appears to be a difficult problem in general. Here, we mostly focus on the special case where X^n is uniformly distributed over the Boolean cube $\{0,1\}^n$ and Y^n is obtained by passing X^n through a memoryless binary symmetric channel (BSC) with crossover probability δ (Section 3). We derive two upper bounds and two lower bounds on the guessing ratio in this case. The upper bounds are derived by analyzing the ratio attained by two specific functions, *k-Dictator*, to wit $f(Y^n) = Y^k$, and *Majority*, to wit $f(Y^n) = \mathbb{1}(\sum_{i=1}^n Y_i > \frac{n}{2})$, where $\mathbb{1}(\cdot)$ is the indicator

function, and for simplicity, we henceforth assume that n is odd when discussing majority functions. For $k = 1$, we demonstrate that neither of these functions is better than the other for all values of the moment order s. The first lower bound is based on relating the guessing-moment to entropy using maximum-entropy arguments (generalizing a result of Reference [8]), and the second one on Fourier-analytic techniques combined with a hypercontractivity argument [21]. Furthermore, for the restricted class of functions for which the constituent k-bit functions operate on disjoint sets of bits, a general method is proposed for transforming a lower bound valid for $k = 1$ to a lower bound valid for any $k \geq 1$. Nonetheless, we remark that our bounds are valid for $s > 0$ and obtaining similar bounds for $s < 0$ in order to obtain large deviation principle for the normalized logarithm of the guessing time remains an open problem. In Section 4, we briefly discuss the more general case where X^n is still uniform over the Boolean cube, but Y^n is obtained from X^n via a general binary-input, arbitrary-output channel. We generalize our entropy lower bound to this case using the strong data-processing inequality (SDPI) applied to the reverse channel (from Y to X). We then discuss the case of the binary erasure channel (BEC), for which we also provide an upper bound by analyzing the *greedy dictator* function, namely where Bob sends the first bit that has not been erased. We conjecture that this function minimizes the guessing-moments simultaneously at all erasure parameters and all moments s.

Related Work. As mentioned above, Graczyk and Lapidoth [19,20] considered the same guessing question if Bob can communicate with Alice at some positive rate R, i.e., can use $k = nR$ bits to describe Y^n. This setup facilitates the use of large-deviation-based information-theoretic techniques, which allowed the authors to characterize the optimal reduction in the guessing-moments as a function of R to the first order in the exponent. This type of argument cannot be applied in our setup of finite number of bits. Furthermore, as we shall see, in our setup, the exponential order of the guessing moment with help is equal to the one without it and the performance is therefore more finely characterized by bounding the ratio of the guessing-moments. For a single bit of help $k = 1$, characterizing the guessing ratio in the case of the BSC with a uniform input can also be thought of as a guessing variant of the *most informative Boolean function problem* introduced by Kumar and Courtade [22]. There, the maximal reduction in the entropy of X^n obtainable by observing a Boolean function $f(Y^n)$ is sought after. It was conjectured in Reference [22] that a *dictator function*, e.g., $f(y^n) = y_1$, is optimal simultaneously at all noise levels; see References [23–26] for some recent progress. As in the guessing case, allowing Bob to describe Y^n using nR bits renders the problem amenable to an exact information-theoretic characterization [27]. In another related work [28], we have asked about the Boolean function Y^n that maximizes the reduction in the sequential mean-squared prediction error of X^n and showed that the majority function is optimal in the noiseless case. There is, however, no single function that is simultaneously optimal at all noise levels. Finally, in a recent line of works [29,30], the average guessing time using the help of a noisy version of $f(X^n)$ has been considered. The model in this paper is different since the noise is applied to the inputs of the function rather than to its output.

2. Problem Statement

Let X^n be an i.i.d. vector from a distribution P_X, which is transmitted over a memoryless channel of conditional distribution $P_{Y|X}$. A helper observes $Y^n \in \mathcal{Y}^n$ at the output of the channel and can send k bits $f(Y^n)$, $f : \mathcal{Y}^n \to \{0,1\}^k$ to a guesser of X^n. Our goal is to characterize the best possible multiplicative reduction in guessing-moments offered by a function f, in the limit of large n. Precisely, we wish to characterize the *guessing ratio*, defined as

$$\gamma_{s,k}(P_X, P_{Y|X}) := \limsup_{n\to\infty} \min_{f:\mathcal{Y}^n\to\{0,1\}^k} \frac{G_s(X^n \mid f(Y^n))}{G_s(X^n)} \tag{5}$$

for an arbitrary $s > 0$. In this paper, we are mostly interested in the case where $P_X = (1/2, 1/2)$, i.e., X^n is uniformly distributed over $\{0,1\}^n$, and where the channel is a BSC with crossover probability

$\delta \in [0, 1/2]$. With a slight abuse of notation, we denote the guessing ratio in this case by $\gamma_{s,k}(\delta)$. Furthermore, some of the results will be restricted to the case of a single bit of help ($k = 1$), and in this case, we will further abbreviate the notation from $\gamma_{s,1}(\delta)$ to $\gamma_s(\delta)$. We note the following basic facts.

Proposition 1. *The following properties hold:*

1. *The minimum in Equation (5) is achieved by a sequence of deterministic functions.*
2. *$\gamma_{s,k}(\delta)$ is a non-decreasing function of $\delta \in [0, 1/2]$ which satisfies $\gamma_{s,k}(0) = 2^{-sk}$ and $\gamma_{s,k}(1/2) = 1$. In addition, $\gamma_{s,k}(0)$ is attained by any sequence of functions f_n such that $f_n(Y^n)$ is a uniform Bernoulli vector, i.e., $\Pr(f_n(Y^n) = b^k) = 2^{-k}$ for all $b^k \in \{0, 1\}^k$.*
3. *For a BSC $P_{Y|X}$, the limit-supremum in Equation (5) defining $\gamma_{s,k}(\delta)$ is a regular limit.*
4. *If $k = 1$ and X^n is a uniformly distributed vector, then the optimal guessing order given that $f(Y^n) = 0$ is reversed to the optimal guessing order when $f(Y^n) = 1$.*

Proof. See Appendix A. □

3. Guessing Ratio for a Binary Symmetric Channel

3.1. Main Results

We begin by presenting the bound on the guessing ratio $\gamma_{s,k}(\delta)$ obtained by *k-dictator* functions and then proceed to the bound obtained by *majority* functions for a single bit of help, $k = 1$. The proofs are given in the next two subsections.

Theorem 1. *Let $L_{k,w} := \sum_{v=0}^{w} \binom{k}{v}$ for $w \in \{0, 1, \ldots, k\}$. The guessing ratio is upper bounded as*

$$\gamma_{s,k}(\delta) \leq (1 - 2\delta) \cdot 2^{-sk} \cdot \sum_{w=0}^{k-1} (1 - \delta)^{k-1-w} \delta^w \cdot L_{k,w}^{s+1} + (2\delta)^k, \tag{6}$$

and this upper bound is achieved by k-dictator functions, $f(y^n) = y^k$.

Specifically, for $k = 1$, Theorem 1 implies

$$\gamma_s(\delta) \leq (1 - 2\delta) \cdot 2^{-s} + 2\delta. \tag{7}$$

Theorem 2. *Let $\beta := \frac{1-2\delta}{\sqrt{4\delta(1-\delta)}}$ and $Z \sim \mathcal{N}(0, 1)$, and denote by $Q(\cdot)$ the tail distribution function of the standard normal distribution. Then, the guessing ratio is upper bounded as*

$$\gamma_s(\delta) \leq 2 \cdot (s + 1) \cdot \mathbb{E}\left[Q(\beta Z) \cdot (1 - Q(Z))^s\right], \tag{8}$$

and this upper bound is achieved by majority functions, $f(y^n) = \mathbb{1}(\sum_{i=1}^{n} y_i > \frac{n}{2})$.

We remark that, if $k = 1$, the guessing ratio of functions similar to the dictator and majority functions, such as single-bit dictator on $j > 1$ inputs ($f(y^n) = 1$ if and only if $y^j = 1^j$) or unbalanced majority ($f(y^n) = \mathbb{1}(\sum_{i=1}^{n} y_i > t)$ for some t), may also be analyzed in a similar way. However, numerical computations indicate that they do not improve the bounds of Theorems 1 and 2, and thus, their analysis is omitted.

We next present two lower bounds on the guessing ratio $\gamma_{s,k}(\delta)$. The first is based on maximum-entropy arguments, and the second is based on Fourier-analytic arguments.

Theorem 3. *The guessing ratio satisfies the following lower bound:*

$$\gamma_{s,k}(\delta) \geq e^{-1} \cdot \frac{s^{s-1} \cdot (s+1)}{\Gamma^s(\frac{1}{s})} \cdot 2^{-sk(1-2\delta)^2} \tag{9}$$

where $\Gamma(z) := \int_0^\infty t^{z-1} e^{-t} \mathrm{d}t$ *is Euler's Gamma function (defined for* $\Re\{z\} > 0$*).*

Remark 1. *When restricted to* $k = 1$*, the proof of Theorem 3 utilizes the bound* $H(X^n|f(Y^n)) \geq n - (1-2\delta)^2$ *(see Equation (63)). For balanced functions, this bound was improved in Reference [23] for* $1/2(1 - 1/\sqrt{3}) \leq \delta \leq 1/2$*. Using this improved bound here leads to an immediate improvement in the bound of Theorem 3. Furthermore, it is known [24] that there exists* δ_0 *such that the most informative Boolean function conjecture holds for all* $\delta_0 \leq \delta \leq 1/2$*. For such crossover probabilities,*

$$H(X^n|f(Y^n)) \geq n - 1 + h(\delta) \tag{10}$$

holds, and then, Theorem 3 may be improved to

$$\gamma_s(\delta) \geq e^{-1} \cdot \frac{s^{s-1} \cdot (s+1)}{\Gamma^s(\frac{1}{s})} \cdot 2^{-s(1-h(\delta))}. \tag{11}$$

Our Fourier-based bound for $k = 1$ is as follows:

Theorem 4. *Let* $\tau := 1 + (1-2\delta)^{2(1-\lambda)}$*. The guessing ratio is lower bounded as*

$$\gamma_s(\delta) \geq \max_{0 \leq \lambda \leq 1} \left[1 - \frac{(s+1) \cdot (1-2\delta)^\lambda}{(\tau s + 1)^{1/\tau}} \right]. \tag{12}$$

This bound can be weakened by the possibly suboptimal choice $\lambda = 1$, which leads to a simpler yet explicit bound:

Corollary 1.

$$\gamma_s(\delta) \geq 1 - \frac{(s+1) \cdot (1-2\delta)}{\sqrt{1+2s}}. \tag{13}$$

The bound in Theorem 4 is only valid for the case $k = 1$. An interesting problem is to find a general way of "transforming" a lower bound which assumes $k = 1$ to a bound useful for $k > 1$. In principle, such a result could stem from the observation that a k bit function provides k different conditional optimal guessing orders for each of its output bits. For a general function, however, distilling a useful bound from this observation seems challenging since the relation between the optimal guessing order induced by each of the bits and the optimal guessing order induced by all k bits might be involved. Nonetheless, such a result is possible to obtain if each of the k single-bit functions operate on a different set of input bits. For this restricted set of functions, there is a simple bound which relates the optimal ordering given each of the bits and all the k bits together. It is reasonable to conjecture that this restricted sub-class is optimal or at least close to optimal, since it seems that more information is transferred to the guesser when the k functions operate on different sets of bits, which make the k functions statistically independent.

Specifically, let us specify a k-bit function $f\colon \mathcal{Y}^n \to \{0,1\}^k$ by its k constituent one-bit functions $f_j\colon \mathcal{Y}^n \to \{0,1\}$, $j \in [k]$. Let $\mathcal{F}_k$ be the set of sequences of functions $\{f^{(n)}\}$, $f^{(n)} : \mathcal{Y}^n \to \{0,1\}^k$, such that each specific sequence of functions $\{f^{(n)}\}$ satisfies the following property: There exists a sequence of partitions $\{\{I_j^{(n)}\}_{j\in[k]}\}_{n=1}^\infty$ of $[n]$, such that, for all $n \geq 1$ and $j \in [k]$, $f_j^{(n)}(Y^n)$ only depends on $\{Y_i\}_{i \in I_j^{(n)}}$ and $\lim_{n\to\infty} |I_j^{(n)}| = \infty$ for all $j \in [k]$. In particular, this implies that $\{f_j^{(n)}(Y^n)\}_{j\in[k]}$ is

mutually independent for all $n \geq 1$. For example, when $k = 2$, $f_1(x^n) = x_1$, and $f_2(x^n) = x_2$, we can choose $I_j^{(n)}$ to be the odd/even indices. For $f_1 = \mathrm{Maj}(y_1^{n/2})$ and $f_2 = \mathrm{Maj}(y_{n/2+1}^n)$, the sets are the first and second halves of $[n]$. As in Equation (5), we may define the guessing ratio of this constrained set of functions as

$$\tilde{\gamma}_{s,k}(\delta) := \min_{\{f^{(n)}\}_{n=1}^{\infty} \in \mathcal{F}_k} \limsup_{n\to\infty} \frac{G_s(X^n \mid f^{(n)}(Y^n))}{G_s(X^n)}, \tag{14}$$

where, in general, $\tilde{\gamma}_{s,k}(\delta) \geq \gamma_{s,k}(\delta)$.

Proposition 2.

$$\tilde{\gamma}_{s,k}(\delta) \geq \frac{\tilde{\gamma}_{s,1}^k(\delta)}{(s+1)^{k-1}}. \tag{15}$$

We demonstrate our results for $k = 1$ in Figure 1 (resp. Figure 2) which display the bounds on $\gamma_s(\delta)$ for fixed values of s (resp. δ). The numerical results show that, for the upper bounds, when $s \lesssim 3.5$, dictator dominates majority (for all values of δ), whereas for $s \gtrsim 4.25$, majority dominates dictator. For $3.5 \lesssim s \lesssim 4.25$, there exists δ_s' such that majority is better for $\delta \in (0, \delta_s')$ and dictator is better for $\delta \in (\delta_s', 1/2)$. Figure 2 demonstrates the switch from dictator to majority as s increases (depending on δ). As for lower bounds, we first remark that the conjectured maximum-entropy bound (Equation (11)) is also plotted (see Remark 1). The numerical results show that the maximum-entropy bound is better for low values of δ whereas the Fourier-analysis bound is better for high values of δ. As a function of s, the maximum-entropy bound (resp. Fourier-analysis bound) is better for high (resp. low) values of s. We also mention that, in these figures, the maximizing parameter in the Fourier-based bound (Theorem 4) is $\lambda = 1$ and the resulting bound is as in Equation (13). However, for values of s as low as 10, the maximizing λ may be far from 1, and in fact, it continuously and monotonically increases from 0 to 1 as δ increases from 0 to $1/2$. Finally, Figure 3 demonstrates the behavior of the k-dictator and maximum-entropy bounds on $\gamma_{s,k}(\delta)$ as a function of k.

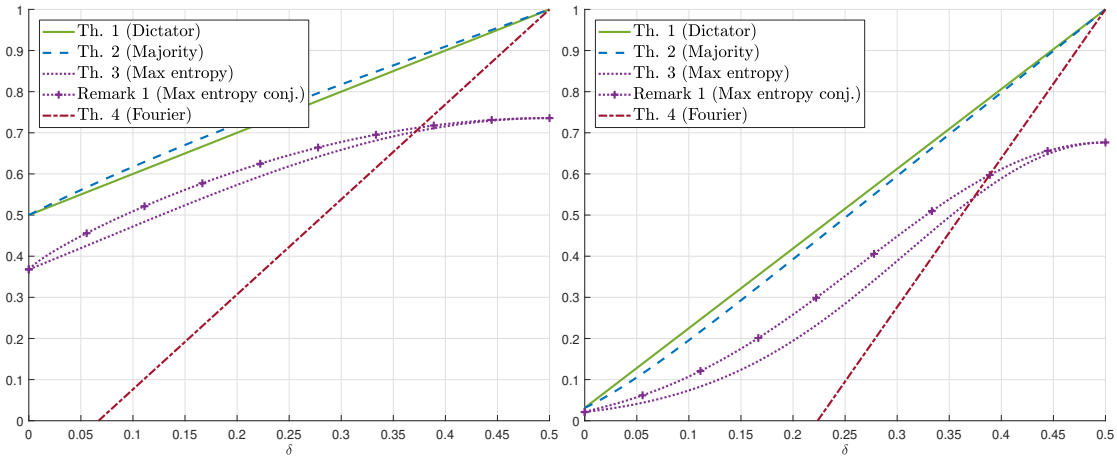

Figure 1. Bounds on $\gamma_s(\delta)$ for $s = 1$ (**left**) and $s = 5$ (**right**) as a function of $\delta \in [0, 1/2]$.

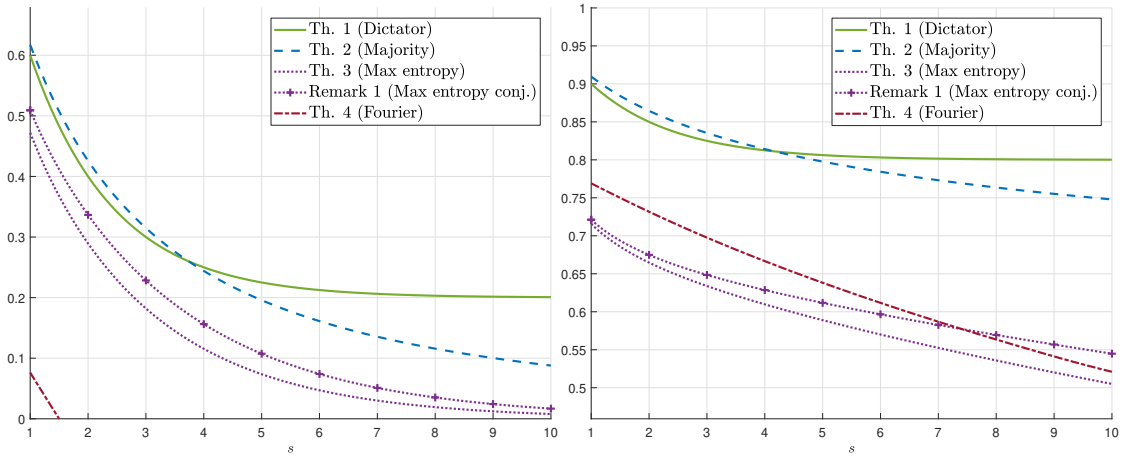

Figure 2. Bounds on $\gamma_s(\delta)$ for $\delta = 0.1$ (**left**) and $\delta = 0.4$ (**right**) as a function of $s \in [1, 10]$.

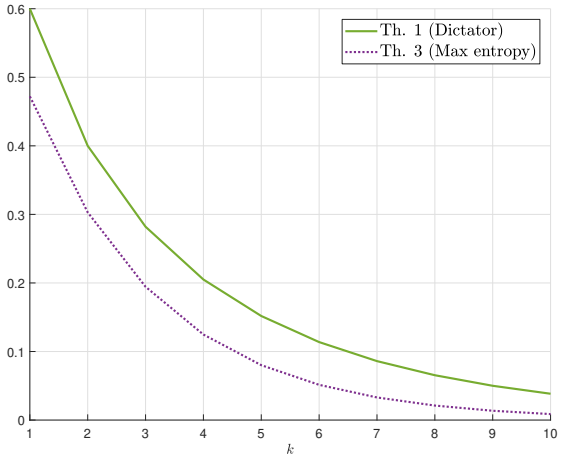

Figure 3. Bounds on $\gamma_{s,k}(\delta)$ for $\delta = 0.1$ and $s = 1$ as a function of k.

3.2. Proofs of the Upper Bounds on $\gamma_{s,k}(\delta)$

Let $a, b \in \mathbb{N}$, $a \leq b$ be given. The following sum will be useful for the proofs in the rest of the paper:

$$K_s(a,b) := \frac{1}{b-a} \sum_{i=a+1}^{b} i^s, \tag{16}$$

where we will abbreviate $K_s(b) := K_s(0,b)$. For a pair of sequences $\{a_n\}_{n=1}^{\infty}$,$\{b_n\}_{n=1}^{\infty}$, we will let $a_n \doteq b_n$ mean that $\lim_{n\to\infty} \frac{a_n}{b_n} = 1$.

Lemma 1. *Let $\{a_n\}_{n=1}^{\infty}$ and $\{b_n\}_{n=1}^{\infty}$ be non-decreasing integer sequences such that $a_n < b_n$ for all n and $\lim_{n\to\infty}(a_n+1)/b_n = 0$. Then,*

$$K_s(a_n, b_n) \doteq \frac{1}{s+1} \cdot \frac{b_n^{s+1} - a_n^{s+1}}{b_n - a_n}. \tag{17}$$

Specifically, $G_s(X^n) = K_s(2^n) \doteq \frac{2^{sn}}{s+1}$.

Proof. See Appendix A. □

We next prove Theorem 1.

Proof of Theorem 1. Consider a k-dictator function which directly outputs k of the bits of y^n, say, without loss of generality (w.l.o.g.) $f(y^n) = y^k$. Let $d_H(x^n, y^n)$ be the Hamming distance of x^n and y^n,

and recall the assumption $0 < \delta < 1/2$. It is easily verified that the optimal guessing order of X^n given y^k has $k+1$ parts, such that the wth part, $w \in \{0, 1, \ldots, k\}$, is comprised of an arbitrary ordering of the $\binom{k}{w} \cdot 2^{n-k}$ vectors for which $d_H(x^k, y^k) = w$. From symmetry, $G_s(X^n \mid f(Y^n)) = G_s(X^n \mid f(Y^n) = b^k)$ for any $b^k \in \{0,1\}^k$. Then, from Lemma 1

$$G_s(X^n \mid f(Y^n) = b^k) = \sum_{w=0}^{k} \binom{k}{w} (1-\delta)^{k-w} \delta^w \cdot K_s(2^{n-k} \cdot L_{k,w-1}, 2^{n-k} \cdot L_{k,w}) \tag{18}$$

$$= \sum_{w=0}^{k} \binom{k}{w} (1-\delta)^{k-w} \delta^w \cdot K_s(2^{n-k} \cdot L_{k,w-1}, 2^{n-k} \cdot L_{k,w}) \tag{19}$$

$$\doteq \sum_{w=0}^{k} \binom{k}{w} (1-\delta)^{k-w} \delta^w \cdot \frac{2^{s(n-k)}}{s+1} \cdot \frac{L_{k,w}^{s+1} - L_{k,w-1}^{s+1}}{\binom{k}{w}} \tag{20}$$

$$= \frac{2^{s(n-k)}}{s+1} \sum_{w=0}^{k} (1-\delta)^{k-w} \delta^w \cdot \left[L_{k,w}^{s+1} - L_{k,w-1}^{s+1} \right] \tag{21}$$

$$= \frac{2^{s(n-k)}}{s+1} \left((1-2\delta) \sum_{w=0}^{k-1} (1-\delta)^{k-1-w} \delta^w \cdot L_{k,w}^{s+1} + \delta^k 2^{k(s+1)} \right) \tag{22}$$

where in the first equality, $L_{k,-1} := 0$, and the last equality is obtained by telescoping the sum. The result then follows from Equation (5) and Lemma 1. □

We next prove Theorem 2.

Proof of Theorem 2. Recall that we assume for simplicity that n is odd. The analysis for an even n is not fundamentally different. To evaluate the guessing-moment, we first need to find the optimal guessing strategy. To this end, we let $W_H(x^n)$ be the Hamming weight of x^n and note that the posterior probability is given by

$$\Pr(X^n = x^n \mid \mathrm{Maj}(Y^n) = 1) = \frac{\Pr(\mathrm{Maj}(Y^n) = 1 \mid X^n = x^n) \cdot \Pr(X^n = x^n)}{\Pr(\mathrm{Maj}(Y^n) = 1)} \tag{23}$$

$$= 2^{1-n} \cdot \Pr\left(\sum_{i=1}^{n} Y_i > n/2 \mid X^n = x^n \right) \tag{24}$$

$$= 2^{1-n} \cdot \Pr\left(\sum_{i=1}^{n} Y_i > n/2 \mid W_H(X^n) = W_H(x^n) \right) \tag{25}$$

$$=: 2^{1-n} \cdot r_n(W_H(x^n)), \tag{26}$$

where Equation (25) follows from symmetry. Evidently, $r_n(w)$ is an increasing function of $w \in \{0, 1, \ldots, n\}$. Indeed, let $\mathrm{Bin}(n, \delta)$ be a binomial r.v. of n trials and success probability δ. Then, for any $w \leq n-1$, as $\delta \leq 1/2$,

$$r_n(w+1)$$

$$= \Pr\left(\mathrm{Bin}(w+1, 1-\delta) + \mathrm{Bin}(n-w-1, \delta) > n/2\right) \tag{27}$$

$$= \Pr\left(\mathrm{Bin}(w, 1-\delta) + \mathrm{Bin}(1, 1-\delta) + \mathrm{Bin}(n-w-1, \delta) > n/2\right) \tag{28}$$

$$\geq \Pr\left(\mathrm{Bin}(w, 1-\delta) + \mathrm{Bin}(1, \delta) + \mathrm{Bin}(n-w-1, \delta) > n/2\right) \tag{29}$$

$$= \Pr\left(\mathrm{Bin}(w, 1-\delta) + \mathrm{Bin}(n-w, \delta) > n/2\right) \tag{30}$$

$$= r_n(w), \tag{31}$$

where, in each of the above probabilities, the summation is over an independent binomial r.v. Hence, we deduce that, whenever $\mathrm{Maj}(Y^n) = 1$ (resp. $\mathrm{Maj}(Y^n) = 0$), the optimal guessing strategy is by

decreasing (resp. increasing) Hamming weight (with arbitrary order for inputs of equal Hamming weight).

We can now turn to evaluate the guessing-moment for the optimal strategy given the majority of Y^n. Let $M_{n,w} := \sum_{v=0}^{w} \binom{n}{v}$ for $w \in \{0, 1, \ldots, n\}$. From symmetry,

$$G_s(X^n \mid \mathrm{Maj}(Y^n)) = G_s(X^n \mid \mathrm{Maj}(Y^n) = 1) \tag{32}$$

$$= \sum_{w=0}^{n} \binom{n}{w} 2^{1-n} r_n(w) \sum_{i=M_{n,w-1}+1}^{M_{n,w}} i^s \tag{33}$$

where $M_{n,-1} := 0$. Thus,

$$G_s(X^n \mid \mathrm{Maj}(Y^n)) \geq \sum_{w=0}^{n} \binom{n}{w} 2^{1-n} r_n(w) M_{n,w-1}^s \tag{34}$$

$$= 2^{sn+1} \cdot \mathbb{E}\left[r_n(W) \left(\frac{M_{n,W-1}}{2^n} \right)^s \right] \tag{35}$$

$$= 2^{sn+1} \cdot \mathbb{E}\left[r_n(W) \Pr\left(W' \leq W - 1\right)^s \right], \tag{36}$$

where $W, W' \sim \mathrm{Bin}(n, 1/2)$ and is independent. For evaluating the asymptotic behavior (for large n) of this expression, we note that the Berry–Esseen central-limit theorem ([31], Chapter XVI.5, Theorem 2) leads to (see, e.g., Reference [28], proof of Lemma 15)

$$r_n(w) = Q\left(\beta \cdot \frac{2}{\sqrt{n}} \left[\frac{n}{2} - w \right] \right) + \frac{a_\delta}{\sqrt{n}}, \tag{37}$$

for some universal constant a_δ. Using the Berry–Esseen central-limit theorem again, we have that $\frac{2}{\sqrt{n}}(\frac{n}{2} - W') \xrightarrow{d} Z$, where $Z \sim \mathcal{N}(0,1)$ and $\xrightarrow{d}$ denote convergence in distribution. Thus for a given w,

$$\Pr\left(W' \leq w - 1\right) = 1 - \Pr\left(\frac{2}{\sqrt{n}} \left(\frac{n}{2} - W' \right) \geq \frac{2}{\sqrt{n}} \left(\frac{n}{2} - w - 1 \right) \right) \tag{38}$$

$$= 1 - Q\left(\frac{2}{\sqrt{n}} \left(\frac{n}{2} - w - 1 \right) \right) - \frac{a_{1/2}}{\sqrt{n}} \tag{39}$$

$$= 1 - Q\left(\frac{2}{\sqrt{n}} \left(\frac{n}{2} - w \right) \right) - O\left(\frac{1}{\sqrt{n}} \right), \tag{40}$$

where the last equality follows from the fact that $|Q'(t)| \leq \frac{1}{\sqrt{2\pi}}$ for all $t \in \mathbb{R}$. Using the Berry–Esseen theorem once again, we have that $\frac{2}{\sqrt{n}}(\frac{n}{2} - w) \xrightarrow{d} Z$. Hence, Portmanteau's lemma (e.g., Reference [31], Chapter VIII.1, Theorem 1) and the fact the $Q(t)$ is continuous and bounded result in the following:

$$G_s(X^n \mid \mathrm{Maj}(Y^n)) \geq 2^{sn+1} \cdot \mathbb{E}\left[Q\left(\beta N\right) \cdot \left(1 - Q(N)\right)^s \right] + O\left(\frac{1}{n^{s/2}} \right). \tag{41}$$

Similarly to Equation (34), the upper bound

$$G_s(X^n \mid \mathrm{Maj}(Y^n)) \leq \sum_{w=0}^{n} \binom{n}{w} 2^{1-n} r_n(w) M_w^s, \tag{42}$$

holds, and a similar analysis leads to an expression which asymptotically coincides with the right-hand side (r.h.s.) of Equation (41). The result then follows from Equation (5) and Lemma 1. □

3.3. Proofs of the Lower Bounds on $\gamma_{s,k}(\delta)$

To prove Theorem 3, we first prove the following maximum entropy result. With a standard abuse of notation, we will write the guessing-moment and the entropy of a random variable as functions of its distribution.

Lemma 2. *The maximal entropy under guessing-moment constraint satisfies*

$$\max_{P:G_s(P)=g} H(P) = \log\left(e^{1/s}s^{(1-s)/s}\cdot G_s^{1/s}(P)\cdot\Gamma\left(\frac{1}{s}\right)\right)+o(1), \tag{43}$$

where $o(1)$ vanishes as $g\to\infty$.

Proof. To solve the maximum entropy problem ([32], Chapter 12) in Equation (43) (note that the support of P is only restricted to be countable), we first relax the constraint $G_s(P)=g$ to

$$\sum_{i=1}^{\infty} P(i)\cdot i^s = g, \tag{44}$$

i.e., we omit the requirement that $\{P(i)\}$ is a decreasing sequence. Assuming momentarily that the entropy is measured in nats, it is easily verified (e.g., using the theory of exponential families ([33], Chapter 3) or by Lagrange duality ([34], Chapter 5)) that the entropy maximizing distribution is

$$P_\lambda(i) := \frac{\exp(-\lambda i^s)}{Z(\lambda)} \tag{45}$$

for $i\in\mathbb{N}_+$, where $Z(\lambda):=\sum_{i=1}^{\infty}\exp(-\lambda i^s)$ is the *partition function* and $\lambda>0$ is chosen such that $\sum_{i=1}^{\infty}P_\lambda(i)\cdot i^s=g$. Evidently, $P_\lambda(i)$ is in decreasing order (and so is $G_s(P_\lambda)=g$) and is therefore the solution to Equation (43). The resulting maximum entropy is then given in a parametric form as

$$H(P_\lambda)=\lambda G_s(P_\lambda)+\ln Z(\lambda). \tag{46}$$

Evidently, if $g=G_s(P_\lambda)\to\infty$, then $\lambda\to 0$. In this case, we may approximate the limit of the partition function as $\lambda\to 0$ by a Riemann integral. Specifically, by the monotonicity of $e^{-\lambda i^s}$ in $i\in\mathbb{N}$,

$$\begin{aligned} Z(\lambda) &= \sum_{i=1}^{\infty} e^{-\lambda i^s} && (47)\\ &= \frac{1}{2}\left(\sum_{i=-\infty}^{\infty}\exp\left(-\left(\frac{|i|}{\lambda^{-1/s}}\right)^s\right)-1\right) && (48)\\ &\geq \frac{1}{2}\left(\int_{-\infty}^{\infty}\exp\left(-\left(\frac{|t|}{\lambda^{-1/s}}\right)^s\right)\mathrm{d}t-1\right) && (49)\\ &= \frac{1}{s}\lambda^{-1/s}\cdot\Gamma\left(\frac{1}{s}\right)-\frac{1}{2}, && (50) \end{aligned}$$

where the last equality follows from the definition of the Gamma function (see Theorem 3) or from the identification of the integral as an unnormalized generalized Gaussian distribution of zero mean, scale parameter $\lambda^{-1/s}$, and shape parameter s [35]. Further, by the convexity of $e^{-\lambda t^s}$ in $t\in\mathbb{R}_+$, Jensen's inequality implies that

$$e^{-\lambda i^s}\leq\int_{-i-1/2}^{i+1/2}\exp\left(-\lambda|t|^s\right)\mathrm{d}t \tag{51}$$

for every $i \geq 1$ (the r.h.s. can be considered as averaging over a uniform random variable $[i-1/2, i+1/2]$) and so, similarly to Equation (50),

$$Z(\lambda) \leq \frac{1}{2}\left(\int_{-\infty}^{\infty} \exp\left(-\left(\frac{|t|}{\lambda^{-1/s}}\right)^s\right) \mathrm{d}t\right). \tag{52}$$

Therefore,

$$Z(\lambda) = (1+a_\lambda) \cdot \frac{1}{s}\lambda^{-1/s} \cdot \Gamma\left(\frac{1}{s}\right) \tag{53}$$

where $a_\lambda \to 0$ as $\lambda \to 0$. In the same spirit,

$$G_s(P_\lambda) = \sum_{i=1}^{\infty} i^s \cdot \frac{\exp(-\lambda i^s)}{Z(\lambda)} \tag{54}$$

$$= \frac{\int_0^\infty t^s \exp\left(-\left(\frac{|t|}{\lambda^{-1/s}}\right)^s\right) \mathrm{d}t + b_\lambda}{(1+a_\lambda)\frac{1}{s}\lambda^{-1/s} \cdot \Gamma\left(\frac{1}{s}\right)} \tag{55}$$

$$= \frac{\frac{1}{s}\lambda^{-\frac{s+1}{s}} \cdot \Gamma\left(\frac{s+1}{s}\right) + b_\lambda}{(1+a_\lambda)\frac{1}{s}\lambda^{-1/s} \cdot \Gamma\left(\frac{1}{s}\right)} \tag{56}$$

$$= \frac{\frac{1}{s^2}\lambda^{-\frac{s+1}{s}} \cdot \Gamma\left(\frac{1}{s}\right) + b_\lambda}{(1+a_\lambda)\frac{1}{s}\lambda^{-1/s} \cdot \Gamma\left(\frac{1}{s}\right)} \tag{57}$$

$$= \frac{1}{s\lambda}(1+c_\lambda), \tag{58}$$

where in Equation (56), $b_\lambda \to 0$ as $\lambda \to 0$; in Equation (57), the identity $\Gamma(t+1) = t\Gamma(t)$ for $t \in \mathbb{R}_+$ was used; and in Equation (58), $c_\lambda \to 0$ as $\lambda \to 0$.

Returning to measure entropy in bits, we thus obtain that, for any distribution P,

$$H(P) \leq \log\left(\frac{e^{1/s}}{s} s^{1/s} \cdot G_s^{1/s}(P) \cdot \Gamma\left(\frac{1}{s}\right)\right) + o(1), \tag{59}$$

or, equivalently,

$$G_s(P) \geq \Psi_s \cdot 2^{sH(P)} \cdot (1+o(1)), \tag{60}$$

where $\Psi_s := e^{-1} \cdot \frac{s^{s-1}}{\Gamma^s(\frac{1}{s})}$ and $o(1)$ is a vanishing term as $G_s(P) \to \infty$. In the same spirit, Equation (60) holds whenever $H(P) \to \infty$. □

Remark 2. *In Reference [8], the maximum-entropy problem was studied for $s = 1$. In this case, the maximum-entropy distribution is readily identified as the geometric distribution. The proof above generalizes that result to any $s > 0$.*

Proof of Theorem 3. Assume that f is taken from a sequence of functions which achieves the minimum in Equation (5). Using Lemma 2 when conditioning on $f(Y^n) = b^k$ for each of possible b^k, we get (see a rigorous justification to Equation (61) in Appendix A)

$$G_s(X^n \mid f(Y^n)) \geq \ell_n \cdot \Psi_s \cdot \sum_{b^k \in \{0,1\}^k} \Pr(f(Y^n) = b^k) \cdot 2^{sH(X^n|f(Y^n)=b^k)} \tag{61}$$

$$\geq \ell_n \cdot \Psi_s \cdot 2^{sH(X^n|f(Y^n))} \tag{62}$$

$$\geq \ell_n \cdot \Psi_s \cdot 2^{s[n-k(1-2\delta)^2]} \tag{63}$$

where in Equation (61), $\ell_n \doteq 1$ and Equation (62) follows from Jensen's inequality. For $k = 1$, the bound in Equation (63) is directly related to the Boolean function conjecture [22] and may be proved in several ways, e.g., using Mrs. Gerber's Lemma ([36], Theorem 1); see ([23], Section IV), References [27,37]. For general $k \geq 1$, the bound $H(X^n|f(Y^n)) \geq n - k(1-2\delta)^2$ was established in Reference ([27], Corollary 1). □

Before presenting the proof of the Fourier-based bound, we briefly remind the reader of the basic definitions and results of Fourier analysis of Boolean functions [21], and to that end, it is convenient to replace the binary alphabet $\{0,1\}$ by $\{-1,1\}$. An inner product between two real-valued functions on the Boolean cube $f, g : \{-1,1\}^n \to \mathbb{R}$ is defined as

$$\langle f, g \rangle := \mathbb{E}\left(f(X^n)g(X^n)\right), \tag{64}$$

where $X^n \in \{-1,1\}^n$ is a uniform Bernoulli vector. A *character* associated with a set of coordinates $S \subseteq [n] := \{1,2,\ldots,n\}$ is the Boolean function $x^S := \prod_{i\in S} x_i$, where by convention, $x^{\emptyset} := 1$. It can be shown ([21], Chapter 1) that the set of all characters forms an orthonormal basis with respect to the inner product (Equation (64)). Furthermore,

$$f(x^n) = \sum_{S\subseteq[n]} \hat{f}_S \cdot x^S, \tag{65}$$

where $\{\hat{f}_S\}_{S\subseteq[n]}$ are the *Fourier coefficients* of f, given by $\hat{f}_S = \langle x^S, f\rangle = \mathbb{E}(X^S \cdot f(X^n))$. *Plancherel's identity* then states that $\langle f, g\rangle = \mathbb{E}(f(X^n)g(X^n)) = \sum_{S\in[n]} \hat{f}_S \hat{g}_S$. The p norm of a function f is defined as $\|f\|_p := [\mathbb{E}|f(X^n)|^p]^{1/p}$.

The *noise operator* operating on a Boolean function f is defined as

$$T_\rho f(x^n) = \mathbb{E}(f(Y^n) \mid X^n = x^n) \tag{66}$$

where $\rho := 1 - 2\delta$ is the *correlation parameter*. The noise operator has a smoothing effect on the function which is captured by the so-called hypercontractivity theorems. Specifically, we shall use the following version.

Theorem 5 ([21], p. 248)**.** *Let $f : \{-1,1\}^n \to \mathbb{R}$ and $0 \leq \rho \leq 1$. Then, $\|T_\rho f\|_2 \leq \|f\|_{\rho^2+1}$.*

With the above, we can prove Theorem 4.

Proof of Theorem 4. From Bayes law (recall that $f(x^n) \in \{-1,1\}$)

$$\Pr(X^n = x^n \mid f(Y^n) = b) = 2^{-(n+1)} \cdot \frac{1 + bT_\rho f(x^n)}{\Pr(f(Y^n) = b)}, \tag{67}$$

and from the law of total expectation

$$G_s(X^n \mid f(Y^n)) = \Pr(f(Y^n) = 1) \cdot G_s(X^n \mid f(Y^n) = 1) + \Pr(f(Y^n) = -1) \cdot G_s(X^n \mid f(Y^n) = -1). \tag{68}$$

Let us denote $\hat{f}_\phi = \mathbb{E}f(X^n)$ and $g := f - \hat{f}_\phi$ and abbreviate $\mathrm{ORD}_f(x^n) := \mathrm{ORD}_{X^n|f(Y^n)}(x^n \mid 1)$. Then, the first addend on the r.h.s. of Equation (68) is given by

$$\Pr(f(Y^n) = 1) \cdot G_s(X^n \mid f(Y^n) = 1) = 2^{-(n+1)} \sum_{x^n} \left(1 + \hat{f}_\phi + T_\rho g(x^n)\right) \cdot \mathrm{ORD}^s_{T_\rho g}(x^n) \tag{69}$$

$$= \frac{(1+\hat{f}_\phi)}{2} \cdot \mathbb{E}\left(\mathrm{ORD}^s_{T_\rho g}(X^n)\right) + \frac{1}{2}\langle T_\rho g, \mathrm{ORD}^s_{T_\rho g}\rangle \tag{70}$$

$$= \frac{(1+\hat{f}_\phi)}{2} \cdot K_s(2^n) + \frac{1}{2}\langle T_\rho g, \mathrm{ORD}^s_{T_\rho g}\rangle \tag{71}$$

$$= \frac{(1+\hat{f}_\phi)}{2} \cdot \ell_n \cdot \frac{2^{sn}}{s+1} + \frac{1}{2}\langle T_\rho g, \mathrm{ORD}^s_{T_\rho g}\rangle, \tag{72}$$

where, in the last equality, $\ell_n \doteq 1$ (Lemma 1). Let $\lambda \in [0,1]$, and denote $\rho_1 := \rho^\lambda$ and $\rho_2 = \rho^{1-\lambda}$. Then, the inner-product term in Equation (72) is upper bounded as

$$\begin{aligned}
\left|\langle T_\rho g, \mathrm{ORD}^s_{T_\rho g}\rangle\right| &= \left|\langle T_{\rho_1} g, T_{\rho_2}\, \mathrm{ORD}^s_{T_\rho g}\rangle\right| && (73)\\
&\le \|T_{\rho_1} g\|_2 \cdot \|T_{\rho_2}\, \mathrm{ORD}^s_{T_\rho g}\|_2 && (74)\\
&\le \rho_1 \cdot \sqrt{1-\hat{f}^2_\phi} \cdot \|T_{\rho_2}\, \mathrm{ORD}^s_{T_\rho g}\|_2 && (75)\\
&\le \rho_1 \cdot \sqrt{1-\hat{f}^2_\phi} \cdot \|\,\mathrm{ORD}^s_{T_\rho g}\|_{1+\rho_2^2} && (76)\\
&= \rho_1 \cdot \sqrt{1-\hat{f}^2_\phi} \cdot \left(K_{(1+\rho_2^2)s}(2^n)\right)^{1/(1+\rho_2^2)} && (77)\\
&= \rho_1 \cdot \sqrt{1-\hat{f}^2_\phi} \cdot \left[k_n \cdot \frac{1}{(1+\rho_2^2)s+1}\right]^{1/(1+\rho_2^2)} \cdot 2^{sn}, && (78)
\end{aligned}$$

where Equation (73) holds since T_ρ is a self-adjoint operator and Equation (74) follows from the Cauchy–Schwarz inequality. To justify Equation (75), we note that

$$\begin{aligned}
\|T_\rho g\|_2^2 &= \langle T_\rho g, T_\rho g\rangle && (79)\\
&= \sum_{S\in[n]} \rho^{2|S|}\hat{g}_S^2 && (80)\\
&= \sum_{S\in[n]\setminus\phi} \rho^{2|S|}\hat{f}_S^2 && (81)\\
&\le \rho^2 \cdot (1-\hat{f}^2_\phi), && (82)
\end{aligned}$$

where Equation (80) follows from Plancherel's identity, Equation (81) is since $\hat{g}_S = \hat{f}_S$ for all $S \neq \phi$ and $\hat{g}_\phi = 0$, and Equation (82) follows from $\sum_{S\in[n]} \hat{f}_S^2 = \|f\|_2^2 = \mathbb{E}f^2 = 1$. Equation (76) follows from Theorem 5, and in Equation (78), $k_n \doteq 1$. The second addend on the r.h.s. of Equation (68) can be bounded in the same manner. Hence,

$$\begin{aligned}
G_s(X^n \mid f(Y^n)) &\ge \max_{0\le\lambda\le 1} 2^{sn} \cdot \left[\ell_n \cdot \frac{1}{s+1} - \rho^\lambda \cdot \sqrt{1-\hat{f}^2_\phi} \cdot \left[k_n \cdot \frac{1}{(1+\rho^{2(1-\lambda)})s+1}\right]^{1/(1+\rho^{2(1-\lambda)})}\right] && (83)\\
&\ge \max_{0\le\lambda\le 1} 2^{sn} \cdot \left[\ell_n \cdot \frac{1}{s+1} - \rho^\lambda \left[k_n \cdot \frac{1}{(1+\rho^{2(1-\lambda)})s+1}\right]^{1/(1+\rho^{2(1-\lambda)})}\right] && (84)\\
&\to 2^{sn} \cdot \max_{0\le\lambda\le 1} \left[\frac{1}{s+1} - \frac{\rho^\lambda}{\left[(1+\rho^{2(1-\lambda)})s+1\right]^{1/(1+\rho^{2(1-\lambda)})}}\right] && (85)
\end{aligned}$$

as $n \to \infty$. □

We close this section with the following proof of Proposition 2:

Proof of Proposition 2. Let $I = (i_1, \ldots, i_L)$ be a vector of indices in $[n]$ such that $1 \le i_1 < i_2 < \cdots < i_L \le n$, and let $x^n(I) = (x_{i_1}, \ldots, x_{i_L})$ be the components of x^n in those indices. Further, let $\{f^{(n)}\}_{n=1}^\infty \in \mathcal{F}_k$. Then, it holds that

$$\Pr\left[X^n = x^n, f^{(n)}(Y^n)\right] = \prod_{j=1}^k \Pr\left[X^n(I_j) = x^n(I_j), f_j^{(n)}(y^n)\right], \tag{86}$$

as well as

$$\mathrm{ORD}_{X^n|f^{(n)}(Y^n)}(x^n \mid b^k) \geq \prod_{j=1}^{k} \left[\mathrm{ORD}_{X^n(I_j)|f_j^{(n)}(Y^n)}(x^n(I_j) \mid b_j) - 1 \right]. \tag{87}$$

Hence,

$$G_s(X^n \mid f^{(n)}(Y^n)) \geq \prod_{j=1}^{k} [G_s^{(n)}(X^n(I_j) \mid f_j^{(n)}(Y^n)) - 1] \tag{88}$$

and the stated bound is deduced after taking limits and normalizing by $G_s(X^n) \doteq \frac{2^{sn}}{s+1}$. □

4. Guessing Ratio for a General Binary Input Channel

In this section, we consider the guessing ratio for general channels with a uniform binary input. The lower bound of Theorem 3 can be easily generalized to this case. To that end, consider the SDPI constant [38,39] of the reverse channel $(P_Y, P_{X|Y})$, given by

$$\eta(P_Y, P_{Y|X}) := \sup_{Q_Y:\, Q_Y \neq P_Y} \frac{D(Q_X||P_X)}{D(Q_Y||P_Y)}, \tag{89}$$

where Q_X is the X-marginal of $Q_Y \circ P_{X|Y}$. As was shown in Reference ([40], Theorem 2), the SDPI constant of $(P_Y, P_{X|Y})$ is also given by

$$\eta(P_Y, P_{Y|X}) = \sup_{P_{W|Y}:\, W-Y-X,\, I(W;Y)>0} \frac{I(W;X)}{I(W;Y)}. \tag{90}$$

Theorem 6. *We have*

$$\gamma_{s,k}(P_X, P_{Y|X}) \geq e^{-1} \cdot \frac{s^{s-1} \cdot (s+1)}{\Gamma^s(\frac{1}{s})} \cdot 2^{-s \cdot k \cdot \eta(P_Y, P_{X|Y})}. \tag{91}$$

Proof. See Appendix A. □

Remark 3. *The bound for the BSC case (Theorem 3) is indeed a special case of Theorem 6 as the reverse BSC channel is also a BSC with uniform input and the same crossover probability. For BSCs, it is well known that the SDPI constant is* $(1-2\delta)^2$ *([38], Theorem 9).*

Next, we consider in more detail the case where the observation channel is a BEC. We restrict the discussion to the case of a single bit of help, $k = 1$.

4.1. Binary Erasure Channel

Suppose that $Y^n \in \{0, 1, \mathrm{e}\}^n$ is obtained from X^n by erasing each bit independently with probability $\epsilon \in [0, 1]$. As before, Bob observes the channel output Y^n and can send one bit $f : \{0, 1, \mathrm{e}\}^n \to \{0, 1\}$ to Alice, who wishes to guess X^n. With a slight abuse of notation, the guessing ratio in Equation (5) will be denoted by $\gamma_s(\epsilon)$.

To compute the lower bound of Theorem 6, we need to find the SDPI constant associated with the reverse channel, which is easily verified to be

$$P_{X|Y=y}(x) = \begin{cases} \mathbb{1}(x = y), & y = 0 \text{ or } y = 1 \\ \mathrm{Ber}(1/2), & y = \mathrm{e}, \end{cases} \tag{92}$$

with an input distribution $P_Y = (\frac{1-\epsilon}{2}, \epsilon, \frac{1-\epsilon}{2})$. Letting $Q_Y(y) = q_y$ for $y \in \{0, 1, \mathrm{e}\}$ yields $Q_X(x) = q_x + \frac{q_\mathrm{e}}{2}$ for $x \in \{0, 1\}$. The computation of $\eta(P_Y, P_{X|Y})$ is now a simple three-dimensional constrained optimization problem. We plotted the resulting lower bound for $s = 1$ in Figure 4.

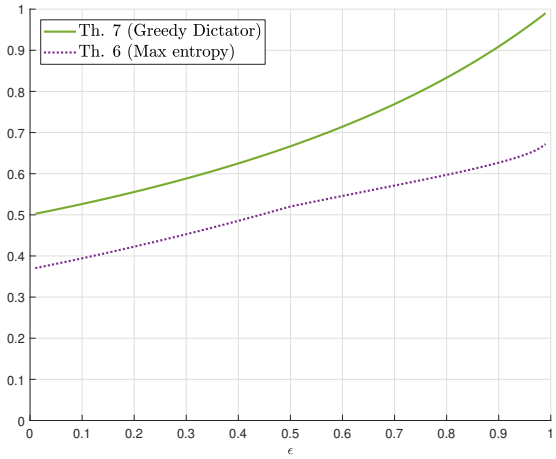

Figure 4. Bounds on $\gamma_s(\delta)$ for $s = 1$ as a function of $\epsilon \in [0, 1]$.

Let us now turn to upper bounds and focus for simplicity on the average guessing time, i.e., the guessing-moment for $s = 1$. To begin, let S represent the set of indices of the symbols that were not erased, i.e., $i \in S$ if and only if $Y_i \neq \mathrm{e}$. Any function $f : \{0, 1, \mathrm{e}\}^n \to \{0, 1\}$ is then uniquely associated with a set of Boolean functions $\{f_S\}_{S \in [n]}$, where $f_S : \{0, 1\}^{|S|} \to \{0, 1\}$ designates the operation of the function when S is the set of non-erased symbols. We also let $\Pr(S) = (1-\epsilon)^{|S|} \cdot \epsilon^{|S^c|}$ be the probability that the non-erased symbols have index set S. Then, the joint probability distribution is given by

$$\begin{aligned} \Pr(X^n = x^n, f(Y^n) = 1) &= \Pr(X^n = x^n) \cdot \Pr(f(Y^n) = 1 \mid X^n = x^n) && (93)\\ &= 2^{-n} \cdot \sum_{S \subseteq [n]} \Pr(S) \cdot \Pr(f(Y^n) = 1 \mid X^n = x^n, S) && (94)\\ &= 2^{-n} \cdot \sum_{S \subseteq [n]} \Pr(S) \cdot f_S(x^n), && (95) \end{aligned}$$

and, similarly,

$$\begin{aligned} \Pr(X^n = x^n, f(Y^n) = 0) &= 2^{-n} \cdot \sum_{S \subseteq [n]} \Pr(S) \cdot (1 - f_S(x^n)) && (96)\\ &= 2^{-n} - 2^{-n} \cdot \sum_{S \subseteq [n]} \Pr(S) \cdot f_S(x^n). && (97) \end{aligned}$$

In accordance with Proposition 1, the optimal guessing order given that $f(Y^n) = 0$ is reversed to the optimal guessing order when $f(Y^n) = 1$. It is also apparent that the posterior probability is determined by a mixture of 2^n different Boolean functions $\{f_S\}_{S \in [n]}$. This may be contrasted with the BSC case, in which the posterior is determined by a *single* Boolean function (though with noisy input).

A seemingly natural choice is a *greedy dictator* function, for which $f(Y^n)$ sends the first non-erased bit. Concretely, letting

$$k(y^n) := \begin{cases} n+1, & y^n = \mathrm{e}^n \\ \min\{i \colon y_i \neq \mathrm{e}\}, & \text{otherwise} \end{cases}, \tag{98}$$

the *greedy dictator* function is defined by

$$\text{G-Dict}(y^n) := \begin{cases} \mathrm{Ber}(1/2), & y^n = \mathrm{e}^n \\ y_{k(y^n)}, & \text{otherwise} \end{cases}, \tag{99}$$

where $\mathrm{Ber}(\alpha)$ is a Bernoulli r.v. of success probability α. From an analysis of the posterior probability, it is evident that, conditioned on $f(Y^n) = 0$, an optimal guessing order must satisfy that x^n is guessed before z^n whenever

$$\sum_{i=1}^{n} \epsilon^{i-1} \cdot x_i \leq \sum_{i=1}^{n} \epsilon^{i-1} \cdot z_i, \tag{100}$$

(see Appendix A for a proof of Equation (100)). This rule can be loosely thought of as comparing the "base $1/\epsilon$ expansion" of x^n and z^n. Furthermore, when ϵ is close to 1, then the optimal guessing order tends toward a *minimum Hamming weight* rule (or maximum Hamming weight in case $f = 1$).

The greedy dictator function is "locally optimal" when $\epsilon \in [0, 1/2]$, in the following sense:

Proposition 3. *If $\epsilon \in [0, 1/2]$, then an optimal guessing order conditioning on* G-Dict$(Y^n) = 0$ *(resp.* G-Dict$(Y^n) = 1$*) is lexicographic (reverse lexicographic). Also, given lexicographic (resp. reverse lexicographic) order when the received bit is* 0 *(resp.* 1*), the optimal function f is a greedy dictator.*

Proof. See Appendix A. □

The guessing ratio of the greedy dictator function can be evaluated for $s = 1$, and the analysis leads to the following upper bound:

Theorem 7. *For $s = 1$, the guessing ratio is upper bounded as*

$$\gamma_1(\epsilon) \leq \frac{1}{2 - \epsilon}, \tag{101}$$

and the r.h.s. is achieved with equality by the greedy dictator function in Equation (99) for $\epsilon \in [0, 1/2]$.

Proof. See Appendix A. □

The upper bound of Theorem 7 is plotted in Figure 4. Based on Proposition 3 and numerical computations for moderate values of n, we conjecture:

Conjecture 1. *Greedy dictator functions attain $\gamma_s(\epsilon)$ for the BEC.*

Supporting evidence for this conjecture include the local optimality property stated in Proposition 3 (although there are other locally optimal choices) as well as the following heuristic argument: Intuitively, Bob should reveal as much as possible regarding the bits he has seen and as little as possible regarding the erasure pattern. So, it seems reasonable to find a smallest possible set of balanced functions from which to choose all the functions f_S, so that they coincide as much as possible. Greedy dictator is a greedy solution to this problem: it uses the function x_1 for half of the erasure patterns, which is the maximum possible. Then, it uses the function x_2 for half of the remaining patterns, and so on. Indeed, we were not able to find a better function than G-Dict for small values of n.

However, applying standard techniques in attempt to prove Conjecture 1 has not been fruitful. One possible technique is induction. For example, assume that the optimal functions for dimension $n-1$ are $f_S^{(n-1)}$. Then, it might be perceived that there exists a bit, say x_1, such that the optimal functions for dimension n satisfy $f_S^{(n)} = f_S^{(n-1)}$ if x_1 is erased; in that case, it remains only to determine $f_S^{(n)}$ when x_1 is not erased. However, observing Equation (95), it is apparent that the optimal choice of $f_S^{(n)}$ should satisfy two contradicting goals—on the one hand, to match the order induced by

$$\sum_{S \subseteq [n]:\, 1 \notin S} \Pr(S) \cdot f_S(x^n) \tag{102}$$

and, on the other hand, to minimize the average guessing time of

$$\sum_{S\subseteq[n]:\, 1\in S} \Pr(S)\cdot f_S(x^n). \tag{103}$$

It is easy to see that taking a greedy approach toward satisfying the second goal would result in $f_S^{(n)}(x^n) = x_1$ if $1 \in S$ and performing the recursion steps would indeed lead to a greedy dictator function. Interestingly, taking a greedy approach toward satisfying the first goal would also lead to a greedy dictator function, but one which operates on a cyclic permutation of the inputs (specifically, Equation (99) applied to (y_2^n, y_1)). Nonetheless, it is not clear that choosing $\{f_S^{(n)}\}_{S:\, 1\in S}$ with some loss in the average guessing time induced by Equation (103) could not lead to a gain in the second goal (matching the order of Equation (102)), which outweighs that loss.

Another possible technique is majorization. It is known that, if one probability distribution majorizes another, then all the nonnegative guessing-moments of the first are no greater than the corresponding moments of the second ([29], Proposition 1). (The proof in Reference [29] is only for $s = 1$, but it is easily extended to the general $s > 0$ case.) Hence, one approach toward identifying the optimal function could be to try and find a function in which induced posterior distributions majorize the corresponding posteriors that induces by any other functions with the same bias (it is of course not clear that such a function even exists). This approach unfortunately fails for the greedy dictator. For example, the posterior distributions induced by setting f_S to be majority functions are not always majorized by those induces by the greedy dictator (although they seem to be "almost" majorized) even though the average guessing time of greedy dictator is lower (this happens, e.g., for $n = 5$ and $\epsilon = 0.4$). In fact, the guessing moments for greedy dictator seem to be better than these of majority irrespective of the value of s.

Author Contributions: Conceptualization, N.W. and O.S.; Investigation, N.W. and O.S.; Methodology, N.W. and O.S.; Writing—original draft, N.W. and O.S.; Writing—review and editing, N.W. and O.S. Both authors contributed equally to the research work and to the writing process of the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by an ERC grant no. 639573. The research of N. Weinberger was partially supported by the MIT-Technion fellowship and the Viterbi scholarship, Technion.

Acknowledgments: We are very grateful to Amos Lapidoth and Robert Graczyk for discussing their recent work on guessing with a helper [19,20] during the second author's visit to ETH, which provided the impetus for this work. We also thank the anonymous reviewer for helping us clarify the connection between the guessing moments and large deviation principle of the normalized logarithm of the guessing time.

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

Abbreviations

The following abbreviations are used in this manuscript:

BEC	binary erasure channel
BSC	binary symmetric channel
i.i.d.	independent and identically distributed
r.h.s.	right-hand side
r.v.	random variable
SDPI	strong data-processing inequality
w.l.o.g.	without loss of generality

Appendix A. Miscellaneous Proofs

Proof of Proposition 1 . The claim that random functions do not improve beyond deterministic ones follows directly from that property that conditioning reduces guessing-moment ([1], Corollary 1). Monotonicity follows from the fact that Bob can always simulate a noisier channel. Now, if $\delta = 1/2$,

then X^n and Y^n are independent and $G_s(X^n \mid f(Y^n)) = G_s(X^n) \doteq \frac{2^{sn}}{s+1}$ for any f (Lemma 1). For $\delta = 0$, let

$$\gamma_{s,k}^{(n)}(\delta) := \min_{f:\{0,1\}^n \to \{0,1\}^k} \frac{G_s(X^n \mid f(Y^n))}{G_s(X^n)}, \tag{A1}$$

and let $\{f_{n,k}^*\}_{n=1}^{\infty}$ be a sequence of functions such that $f_{n,k}^*$ achieves $\gamma_{s,k}^{(n)}(\delta)$. We show that $f_{n,k}^*$ must satisfy $\Pr[f(Y^n) = b^k] = 2^{-k}$ for all $b^k \in \{0,1\}^k$. If we denote $B^k = f_{n,k}^*(Y^n)$, then this is equivalent to showing that $\Pr[B_l = 1 \mid B_1 = b_1, \ldots, B_{l-1} = b_{l-1}, B_l = b_l, \ldots B_k = b_k] = 1/2$ for all $l \in [k]$ and $(b_1, \ldots, b_{l-1}, b_l, \ldots, b_k) \in \{0,1\}^{k-1}$. Assume towards contradiction that the optimal function does not satisfy this property for, say, $l = k$. Let us denote $\Pr(f_{n,k}^*(Y^n) = b^k) := q(b^k)$ and assume w.l.o.g. that $q(b^{k-1}, 0) > q(b^{k-1}, 1)$ for all $b^{k-1} \in \{0,1\}^{k-1}$ (for notational simplicity). Further, let $\overline{q}(b^{k-1}) := \frac{1}{2}[q(b^{k-1}, 0) + q(b^{k-1}, 1)]$. Then,

$$\begin{aligned}
&G_s(X^n \mid f_{n,k}^*(Y^n)) \\
&= \sum_{b^{k-1} \in \{0,1\}^{k-1}} q(b^{k-1}, 0) \cdot G_s(X^n \mid f_{n,k}^*(Y^n) = (b^{k-1}, 0)) + q(b^{k-1}, 1) \cdot G_s(X^n \mid f_{n,k}^*(Y^n) = (b^{k-1}, 1)) \\
&= \sum_{b^{k-1} \in \{0,1\}^{k-1}} q(b^{k-1}, 0) \cdot K_s\left(q(b^{k-1}, 0) \cdot 2^n\right) + q(b^{k-1}, 1) \cdot K_s\left(q(b^{k-1}, 1) \cdot 2^n\right) \\
&= 2^{-n} \sum_{b^{k-1} \in \{0,1\}^{k-1}} \left(\sum_{i=1}^{q(b^{k-1},0) \cdot 2^n} i^s + \sum_{i=1}^{q(b^{k-1},1) \cdot 2^n} i^s \right) \\
&= 2^{-n} \sum_{b^{k-1} \in \{0,1\}^{k-1}} \left(\sum_{i=1}^{\overline{q}(b^{k-1}) \cdot 2^n} i^s + \sum_{i=\overline{q}(b^{k-1})+1}^{q(b^{k-1},0) \cdot 2^n} i^s + \sum_{i=1}^{\overline{q}(b^{k-1}) \cdot 2^n} i^s - \sum_{i=q(b^{k-1},1)}^{\overline{q}(b^{k-1}) \cdot 2^n} i^s \right) \\
&> 2^{-(n-1)} \sum_{b^{k-1} \in \{0,1\}^{k-1}} \sum_{i=1}^{\overline{q}(b^{k-1}) \cdot 2^n} i^s.
\end{aligned}$$

As equality can be achieved if we modify $f_{n,k}^*$ to satisfy $q(b^{k-1}, 0) = q(b^{k-1}, 1)$ for all $b^{k-1} \in \{0,1\}^{k-1}$, this contradicts the assumed optimality of $f_{n,k}^*$. The minimal $G_s(X^n \mid f(Y^n))$ is thus obtained by any function for which $f(Y^n) \in \{0,1\}^k$ is a uniform Bernoulli vector and equals to $K_s(2^{n-k}) \doteq \frac{2^{s(n-k)}}{s+1}$ (Lemma 1).

To prove that the limit in Equation (5) exists, we note that

$$\begin{aligned}
G_s(X^{n+1}) &= 2^{-(n+1)} \sum_{i=1}^{2^{n+1}} i^s && \text{(A2)} \\
&= 2^{-(n+1)} \sum_{i=1}^{2^n} (2i-1)^s + (2i)^s && \text{(A3)} \\
&\geq 2^s \cdot 2^{-n} \sum_{i=1}^{2^n} (i-1)^s && \text{(A4)} \\
&= \ell_n \cdot 2^s \cdot 2^{-n} \sum_{i=1}^{2^n} i^s && \text{(A5)} \\
&= \ell_n \cdot 2^s \cdot G_s(X^n), && \text{(A6)}
\end{aligned}$$

where

$$\ell_n := \frac{\sum_{i=1}^{2^n} (i-1)^s}{\sum_{i=1}^{2^n} i^s}. \tag{A7}$$

As before, let $\{f^*_{n,k}\}_{n=1}^{\infty}$ be a sequence of functions such that $f^*_{n,k}$ achieves $\gamma^{(n)}_{s,k}(\delta)$. Denote the order induced by the posterior $\Pr(X^n = x^n \mid f^*_{n,k}(Y^n) = b^k)$ as $\mathrm{ORD}_{b^k,n,n}$, $b^k \in \{0,1\}^k$ and the order induced by $\Pr(X^{n+1} = x^{n+1} \mid f^*_n(Y^n) = b^k)$ as $\mathrm{ORD}_{b^k,n,n+1}$. As before (when breaking ties arbitrarily)

$$\mathrm{ORD}_{b^k,n,n+1}(x^n, 0) = 2\,\mathrm{ORD}_{b^k,n,n}(x^n) \tag{A8}$$

and

$$\mathrm{ORD}_{b^k,n,n+1}(x^n, 1) = 2\,\mathrm{ORD}_{b^k,n,n}(x^n) - 1 \leq 2\,\mathrm{ORD}_{b^k,n,n}(x^n). \tag{A9}$$

Thus,

$$\begin{aligned}
&G_s(X^{n+1} \mid f^*_{n+1,k}(Y^{n+1})) \\
&\leq G_s(X^{n+1} \mid f^*_{n,k}(Y^n)) && \text{(A10)}\\
&= \sum_{b^k \in \{0,1\}^k} \Pr(f^*_{n,k}(Y^{n+1}) = b^k) \cdot G_s(X^{n+1} \mid f^*_{n,k}(Y^n) = b^k) && \text{(A11)}\\
&\leq \sum_{b^k \in \{0,1\}^k} \sum_{x^{n+1}} \Pr(X^{n+1} = x^{n+1}, f^*_{n,k}(Y^n) = b^k) \cdot \mathrm{ORD}^s_{b^k,n,n+1}(x^{n+1}) \\
&\leq 2^s \cdot \sum_{b^k \in \{0,1\}^k} \sum_{x^n} \Pr(X^n = x^n, f^*_{n,k}(Y^n) = b^k) \cdot \mathrm{ORD}^s_{b^k,n,n}(x^n) && \text{(A12)}\\
&= 2^s \cdot G_s(X^n \mid f^*_{n,k}(Y^n)). && \text{(A13)}
\end{aligned}$$

Hence,

$$\gamma^{(n+1)}_{s,k}(\delta) \leq \ell_n^{-1} \cdot \gamma^{(n)}_{s,k}(\delta). \tag{A14}$$

To continue, we further analyze ℓ_n. The summation in the numerator of Equation (A7) may be started from from $i = 2$, and so Equations (A31) and (A33) (proof of Lemma 1 below) imply that

$$\begin{aligned}
1 &\geq \ell_n && \text{(A15)}\\
&\geq \frac{\frac{1}{s+1} \cdot \frac{2^{n(s+1)}-1}{2^n-1}}{\frac{1}{s+1} \cdot \frac{(2^n+1)^{s+1}-1}{2^n}} && \text{(A16)}\\
&\geq \frac{2^{n(s+1)}-1}{(2^n+1)^{s+1}} && \text{(A17)}\\
&= \left(\frac{2^n}{2^n+1}\right)^{s+1} - \frac{1}{2^{n(s+1)}} && \text{(A18)}\\
&= \left(1 + \frac{1}{2^n}\right)^{-(s+1)} - \frac{1}{2^{n(s+1)}} && \text{(A19)}\\
&= 1 - \frac{(s+1)}{2^n} + O\left(\frac{1}{2^{2n}}\right) - \frac{1}{2^{n(s+1)}} && \text{(A20)}\\
&= 1 - \frac{(s+1)}{2^n} + O\left(\frac{1}{2^{n \cdot \min\{1+s,2\}}}\right). && \text{(A21)}
\end{aligned}$$

Thus, there exists $c, C > 0$ such that

$$\begin{aligned}
\log \prod_{n=1}^{\infty} \ell_n^{-1} &= \sum_{n=1}^{\infty} \log \ell_n^{-1} && \text{(A22)}\\
&\leq -\sum_{n=1}^{\infty} \log\left[1 - \frac{c}{2^n}\right] && \text{(A23)}\\
&\leq C + \sum_{n=1}^{\infty} \frac{c}{2^n} + O\left(\frac{1}{2^{2n}}\right) && \text{(A24)}
\end{aligned}$$

$$< \infty, \tag{A25}$$

and consequently,

$$d_n := \prod_{j=n}^{\infty} \ell_j^{-1} \to 1 \tag{A26}$$

as $n \to \infty$. Hence, Equation (A14) implies that

$$e_n := d_n \cdot \gamma_s^{(n)}(\delta) \tag{A27}$$

is a non-increasing sequence which is bounded below by 0 and, thus, has a limit. Since $d_n \to 1$ as $n \to \infty$, $\gamma_s^{(n)}(\delta)$ also has a limit.

We finally show the reverse ordering property for $k = 1$. The guessing order given that $f(Y^n) = 1$ is determined by ordering

$$\Pr(X^n = x^n \mid f(Y^n) = 1) = \frac{\Pr(X^n = x^n) \cdot \Pr(f(Y^n) = 1 \mid X^n = x^n)}{\Pr(f(Y^n) = 1)}, \tag{A28}$$

or equivalently, by ordering $\Pr(f(Y^n) = 1 \mid X^n = x^n)$. It then follows that the order, given that $f(Y^n) = 0$, is reversed compared to the order given that $f(Y^n) = 1$ since

$$\Pr(f(Y^n) = 0 \mid X^n = x^n) + \Pr(f(Y^n) = 1 \mid X^n = x^n) = 1. \tag{A29}$$

□

Proof of Lemma 1. The monotonicity of i^s and standard bounds on sums using integrals lead to the bounds

$$K_s(a, b) \leq \int_{a+1}^{b+1} \frac{t^s}{b-a} \cdot \mathrm{d}t \tag{A30}$$

$$= \frac{1}{s+1} \cdot \frac{(b+1)^{s+1} - (a+1)^{s+1}}{b-a} \tag{A31}$$

and

$$K_s(a, b) \geq \int_{a}^{b} \frac{t^s}{b-a} \cdot \mathrm{d}t \tag{A32}$$

$$= \frac{1}{s+1} \cdot \frac{b^{s+1} - a^{s+1}}{b-a}. \tag{A33}$$

The ratio between the upper and lower bound is

$$\kappa_s(a, b) := \frac{(b+1)^{s+1} - (a+1)^{s+1}}{b^{s+1} - a^{s+1}} \tag{A34}$$

which satisfies $\kappa_s(a_n, b_n) \to 1$ given the premise of the lemma. □

Proof of Equation (61). Denote by f_n^* a function which achieves the minimal guessing ratio in Equation (5). Then, it holds that $G_s(X^n \mid f^*(Y^n) = b^k)$ is a monotonic non-increasing function of n. To see this, suppose that f_{n+1}^* is an optimal function for $n+1$. This function f_{n+1}^* can be used for guessing X^n on the basis of k bit of help computed from Y^n as follows: Given Y^n, the helper randomly generates $Y_{n+1} \sim P_{Y|X}(\cdot|0)$, computes $b^k = f_{n+1}^*(Y^{n+1})$, and send these bits to the guesser. The guesser of X^n then uses the bits b^k to guess X^n, and the resulting conditional guessing moment is $G_s(X^{n+1} \mid f_{n+1}^*(Y^{n+1}) = b^k, X_{n+1} = 0)$, which is less than $G_s(X^{n+1} \mid f_{n+1}^*(Y^{n+1}) = b^k)$ since conditioning reduces guessing moments. Thus, the optimal function f_n^* can only achieve lower

guessing moments, which implies the desired monotonicity property. For brevity, we henceforth simply write the optimal function as f (with dimension and optimality being implicit).

Define the set

$$\mathcal{B}_k := \{b^k \in \{0,1\}^k \colon \sup_n G_s(X^n \mid f(Y^n) = b^k) = \infty\}, \tag{A35}$$

to wit, the set of k-tuples such that the conditional guessing moment grows without bound when conditioned on that k-tuple. By the law of total expectation

$$\begin{aligned} G_s(X^n \mid f(Y^n)) &= \sum_{b^k \in \mathcal{B}_k} \Pr(f(Y^n) = b^k) \cdot G_s(X^n \mid f(Y^n) = b^k) \\ &\quad + \sum_{b^k \in \{0,1\}^k \setminus \mathcal{B}_k} \Pr(f(Y^n) = b^k) \cdot G_s(X^n \mid f(Y^n) = b^k) \qquad \text{(A36)} \\ &=: G_n^{(1)} + G_n^{(2)}. \qquad \text{(A37)} \end{aligned}$$

So, since $G_s(X^n \mid f(Y^n))$ grows without bound as a function of n, it must hold that $\mathcal{B}_k$ is not empty and that there exists ℓ_n such that $G_s(X^n \mid f(Y^n)) = \ell_n G_n^{(1)}$, where $\ell_n \to 1$ as $n \to \infty$. Let $\eta > 0$ be given. The monotonicity property previously established and Equation (60) imply that there exists $n_0(\eta)$ such that for all $n \geq n_0(\eta)$ both

$$G_s(X^n \mid f(Y^n) = b^k) \geq (1-\eta) \cdot \Psi_s \cdot 2^{sH(X^n \mid f(Y^n) = b^k)} \tag{A38}$$

and

$$\Psi_s \cdot 2^{sH(X^n \mid f(Y^n) = b^k)} \geq (1-\eta) \cdot G_s(X^n \mid f(Y^n) = b^k) \tag{A39}$$

hold for any $b^k \in \mathcal{B}_k$. Thus, also

$$G_s(X^n \mid f(Y^n)) \geq \ell_n (1-\eta) \sum_{b^k \in \mathcal{B}_k} \Pr(f(Y^n) = b^k) \cdot \Psi_s \cdot 2^{sH(X^n \mid f(Y^n) = b^k)} \tag{A40}$$

and

$$\sum_{b^k \in \mathcal{B}_k} \Pr(f(Y^n) = b^k) \cdot \Psi_s \cdot 2^{sH(X^n \mid f(Y^n) = b^k)} \geq \sum_{b^k \in \mathcal{B}_k} \Pr(f(Y^n) = b^k)(1-\eta) \cdot G_s(X^n \mid f(Y^n) = b^k) \tag{A41}$$

hold, and the last equation implies that the term on its left-hand side is unbounded. Moreover, Equation (60) and the sentence that follows it both imply that, if $G_s(X^n \mid f(Y^n) = b^k)$ is bounded, then $H(X^n \mid f(Y^n) = b^k)$ is bounded too. Thus, there exists k_n which satisfies $k_n \to 1$ as $n \to \infty$ such that

$$\sum_{b^k \in \mathcal{B}_k} \Pr(f(Y^n) = b^k) \cdot \Psi_s \cdot 2^{sH(X^n \mid f(Y^n) = b^k)} = k_n \cdot \sum_{b^k \in \{0,1\}^n} \Pr(f(Y^n) = b^k) \cdot \Psi_s \cdot 2^{sH(X^n \mid f(Y^n) = b^k)}. \tag{A42}$$

Combining Equation (A40) with the last equation and noting that $\eta > 0$ is arbitrary completes the proof. □

Proof of Theorem 6. The proof follows the same lines as the proof of Theorem 3 up to Equation (62), yielding

$$G_s(X^n \mid f(Y^n)) \geq k_n \cdot \Psi_s \cdot 2^{s[n - I(X^n; f(Y^n))]}. \tag{A43}$$

Now, let $W^{(n)}$ be such that $X^n - Y^n - W^{(n)}$ forms a Markov chain. Then,

$$\begin{aligned} \sup_{f \colon \mathcal{Y}^n \to \{0,1\}} \frac{I(X^n; f(Y^n))}{I(Y^n; f(Y^n))} &\leq \sup_{P_{W^{(n)} \mid Y^n}} \frac{I(X^n; W^{(n)})}{I(Y^n; W^{(n)})} \qquad \text{(A44)} \\ &= \eta(P_{Y^n}, P_{X^n \mid Y^n}) \qquad \text{(A45)} \end{aligned}$$

$$= \eta(P_Y, P_{X|Y}), \tag{A46}$$

where Equation (A46) follows since the SDPI constant tensorizes (see Reference [40] for an argument obtained by relating the SDPI constant to the hypercontractivity parameter or its extended version, Reference ([40], p. 5), for a direct proof). Thus, for all f,

$$I(X^n; f(Y^n)) \leq \eta(P_Y, P_{X|Y}) \cdot I(Y^n; f(Y^n)) \tag{A47}$$

$$\leq \eta(P_Y, P_{X|Y}) \cdot H(f(Y^n)) \tag{A48}$$

$$\leq \eta(P_Y, P_{X|Y}) \cdot k. \tag{A49}$$

Inserting Equation (A49) into Equation (A43) yields

$$G_s(X^n \mid f(Y^n)) \geq k_n \cdot \Psi_s \cdot 2^{s\left[n - k \cdot \eta(P_Y, P_{X|Y})\right]}, \tag{A50}$$

and substituting this in the definition of the guessing ratio of Equation (5) completes the proof. □

Proof of Equation (100). Let us evaluate the posterior probability conditioned on G-Dict$(Y^n) = 0$. Since G-Dict is balanced, Bayes law implies that

$$\Pr(X^n = x^n \mid \text{G-Dict}(Y^n) = 0)$$

$$= 2^{-(n-1)} \cdot \Pr(\text{G-Dict}(Y^n) = 0 \mid X^n = x^n) \tag{A51}$$

$$= 2^{-(n-1)} \cdot \sum_{i=1}^{n+1} \Pr\left(k(y^n) = i \mid X^n = x^n\right) \cdot \Pr(\text{G-Dict}(Y^n) = 0 \mid X^n = x^n, k(y^n) = i) \tag{A52}$$

$$= 2^{-(n-1)} \cdot \left\{ \sum_{i=1}^{n} (1-\epsilon)\epsilon^{i-1} \cdot \mathbb{1}\left\{x_i = 0\right\} + \tfrac{1}{2}\epsilon^n \right\}. \tag{A53}$$

This immediately leads to the guessing rule in Equation (100). From Proposition 1, the guessing rule for G-Dict$(Y^n) = 1$ is on reverse order. □

Proof of Proposition 3. We denote the lexicographic order by ORD_{lex}. Assume that G-Dict$(Y^n) = 0$ and that $\text{ORD}_{\text{lex}}(x^n) \leq \text{ORD}_{\text{lex}}(z^n)$. Then, there exists $j \in [n]$ such that $x^{j-1} = z^{j-1}$ (where x^0 is the empty string) and $x_j = 0 < z_j = 1$. Then,

$$\Pr(X^n = x^n \mid \text{G-Dict}(Y^n) = 0) - \Pr(X^n = z^n \mid \text{G-Dict}(Y^n) = 0)$$

$$= \epsilon^{j-1} + \sum_{i=j+1}^{n} \epsilon^{i-1} \cdot (z_i - x_i) \tag{A54}$$

$$\geq \epsilon^{j-1} - \sum_{i=j+1}^{n} \epsilon^{i-1} \tag{A55}$$

$$= \frac{\epsilon^{j-1}}{1-\epsilon}\left(1 - 2\epsilon + \epsilon^{n-j+1}\right) \tag{A56}$$

$$\geq 0. \tag{A57}$$

This proves the first statement of the proposition. Now, let ORD_0 (ORD_1) be the guessing order given that the received bit is 0 (resp. 1), and let $\{f_S\}$ be the Boolean functions (which are not necessarily optimal). Then, from Equations (97) and (95)

$$G_1(X^n \mid f(Y^n))$$

$$= \sum_{x^n} \Pr(X^n = x^n, f(Y^n) = 0) \cdot \text{ORD}_0(x^n) + \Pr(X^n = x^n, f(Y^n) = 1) \cdot \text{ORD}_1(x^n) \tag{A58}$$

$$= 2^{-n} \cdot \sum_{S \subseteq [n]} \Pr(S) \sum_{x^n} [(1 - f_S(x^n)) \cdot \mathrm{ORD}_0(x^n) + f_S(x^n) \cdot \mathrm{ORD}_1(x^n)] \tag{A59}$$

$$= 2^{-n} \cdot \sum_{S \subseteq [n]} \Pr(S) \sum_{x^S} \left[(1 - f_S(x^n)) \cdot \mathrm{PORD}_0(x^S || S) + f_S(x^n) \cdot \mathrm{PORD}_1(x^S || S) \right] \tag{A60}$$

$$\geq 2^{-n} \cdot \sum_{S \subseteq [n]} \Pr(S) \sum_{x^n} \min \left\{ \mathrm{PORD}_0(x^S || S), \mathrm{PORD}_1(x^S || S) \right\}, \tag{A61}$$

where for $b \in \{0, 1\}$, the *projected orders* are defined as

$$\mathrm{PORD}_b(x^S || S) := \sum_{x^{(S^c)}} \mathrm{ORD}_b(x^n). \tag{A62}$$

It is easy to verify that, if ORD_0 (ORD_1) is the lexicographic (resp. revered lexicographic) order, then the greedy dictator achieves Equation (A61) with equality due to the following simple property: If $\mathrm{ORD}_{\mathrm{lex}}(x^n) < \mathrm{ORD}_{\mathrm{lex}}(z^n)$, then

$$\sum_{x^{(S^c)}} \mathrm{ORD}_{\mathrm{lex}}(x^n) \leq \sum_{x^{(S^c)}} \mathrm{ORD}_{\mathrm{lex}}(z^n) \tag{A63}$$

for all $S \in [n]$. This can be proved by induction over n. For $n = 1$, the claim is easily asserted. Suppose it holds for $n - 1$, let us verify it for n. If $1 \in S$, then whenever $\mathrm{ORD}_{\mathrm{lex}}(x^n) < \mathrm{ORD}_{\mathrm{lex}}(z^n)$

$$\sum_{x^{(S^c)}} \mathrm{ORD}_{\mathrm{lex}}(x^n) = \sum_{x^{(S^c)}} \mathrm{ORD}_{\mathrm{lex}}(x_1, x_2^n) \tag{A64}$$

$$= x_1 \cdot 2^{n-1} + \sum_{x^{(S^c)}} \mathrm{ORD}_{\mathrm{lex}}(x_2^n) \tag{A65}$$

$$\leq z_1 \cdot 2^{n-1} + \sum_{z^{(S^c)}} \mathrm{ORD}_{\mathrm{lex}}(z_2^n) \tag{A66}$$

$$= \sum_{z^{(S^c)}} \mathrm{ORD}_{\mathrm{lex}}(z^n) \tag{A67}$$

where the inequality follows from the induction assumption and since $x_1 \leq z_1$. If $1 \notin S$ then, similarly,

$$\sum_{x^{(S^c)}} \mathrm{ORD}_{\mathrm{lex}}(x^n) = \sum_{x^{(S^c \setminus \{1\})}} \left[2^{n-1} + 2 \cdot \mathrm{ORD}_{\mathrm{lex}}(x_2^n) \right] \tag{A68}$$

$$\leq \sum_{z^{(S^c \setminus \{1\})}} \left[2^{n-1} + 2 \cdot \mathrm{ORD}_{\mathrm{lex}}(z_2^n) \right] \tag{A69}$$

$$= \sum_{z^{(S^c)}} \mathrm{ORD}_{\mathrm{lex}}(z^n). \tag{A70}$$

□

Proof of Theorem 7. We denote the lexicographic order by $\mathrm{ORD}_{\mathrm{lex}}$. Then,

$$G_1(X^n \mid \text{G-Dict}(Y^n)) = G_1(X^n \mid \text{G-Dict}(Y^n) = 0) \tag{A71}$$

$$\leq \sum_{x^n} \Pr(X^n = x^n \mid \text{G-Dict}(Y^n) = 0) \cdot \mathrm{ORD}_{\mathrm{lex}}(x^n) \tag{A72}$$

$$= 2^{-(n-1)} \cdot \sum_{x^n} \sum_{i=1}^{n} (1 - \epsilon) \epsilon^{i-1} \cdot \mathbb{1}\{x_i = 0\} \cdot \mathrm{ORD}_{\mathrm{lex}}(x^n) + \epsilon^n K_1(2^n) \tag{A73}$$

$$= 2^{-(n-1)} \cdot (1 - \epsilon) \sum_{i=1}^{n} \epsilon^{i-1} \cdot \sum_{x^n} \mathbb{1}\{x_i = 0\} \cdot \mathrm{ORD}_{\mathrm{lex}}(x^n) + \epsilon^n K_1(2^n) \tag{A74}$$

$$= (1 - \epsilon) J_n + \epsilon^n K_1(2^n), \tag{A75}$$

where $J_1 := \frac{1}{2}$ and for $n \geq 2$

$$J_n := 2^{-(n-1)} \cdot \sum_{i=1}^{n} \epsilon^{i-1} \cdot \sum_{x^n} \mathbb{1}\{x_i = 0\} \cdot \mathrm{ORD}_{\mathrm{lex}}(x^n) \tag{A76}$$

$$= 2^{-(n-1)} \sum_{x^n} \mathbb{1}\{x_i = 0\} \cdot \mathrm{ORD}_{\mathrm{lex}}(x^n) + 2^{-(n-1)} \sum_{i=2}^{n} \epsilon^{i-1} \cdot \sum_{x^n} \mathbb{1}\{x_i = 0\} \cdot \mathrm{ORD}_{\mathrm{lex}}(x^n) \tag{A77}$$

$$= K_1(2^{n-1}) + 2^{-(n-1)} \sum_{i=2}^{n} \epsilon^{i-1} \cdot \sum_{x_2^n} [\mathbb{1}\{x_1 = 0, x_i = 0\} \cdot \mathrm{ORD}_{\mathrm{lex}}(0, x_2^n) + \mathbb{1}\{x_1 = 1, x_i = 0\} \cdot \mathrm{ORD}_{\mathrm{lex}}(1, x_2^n)] \tag{A78}$$

$$= K_1(2^{n-1}) + 2^{-(n-1)} \epsilon \sum_{i=1}^{n-1} \epsilon^{i-1} \cdot \sum_{x^{n-1}} \mathbb{1}\{x_i = 0\} \, \mathrm{ORD}_{\mathrm{lex}}(x^{n-1}) + 2^{-(n-1)} \epsilon \sum_{i=1}^{n-1} \epsilon^{i-1} \cdot \sum_{x^{n-1}} \mathbb{1}\{x_i = 0\} \left[2^{n-1} + \mathrm{ORD}_{\mathrm{lex}}(x^{n-1})\right] \tag{A79}$$

$$= K_1(2^{n-1}) + \epsilon J_{n-1} + \sum_{i=1}^{n-1} \epsilon^i \cdot \sum_{x^{n-1}} \mathbb{1}\{x_i = 0\} \tag{A80}$$

$$= K_1(2^{n-1}) + \epsilon J_{n-1} + 2^{n-2} \cdot \frac{\epsilon - \epsilon^n}{1 - \epsilon}. \tag{A81}$$

So,

$$J_n = K_1(2^{n-1}) + \epsilon \left[K_1(2^{n-2}) + \epsilon J_{n-2} + 2^{n-3} \cdot \frac{\epsilon - \epsilon^{n-1}}{1 - \epsilon}\right] + 2^{n-2} \cdot \frac{\epsilon - \epsilon^n}{1 - \epsilon} \tag{A82}$$

$$= K_1(2^{n-1}) + \epsilon K_1(2^{n-2}) + \epsilon^2 J_{n-2} + 2^{n-3} \cdot \frac{\epsilon^2 - \epsilon^n}{1 - \epsilon} + 2^{n-2} \cdot \frac{\epsilon - \epsilon^n}{1 - \epsilon} \tag{A83}$$

$$= \sum_{i=1}^{n} \epsilon^{i-1} K_1(2^{n-i}) + \frac{1}{1 - \epsilon} \sum_{i=1}^{n} 2^{i-2} \cdot (\epsilon^{n-i+1} - \epsilon^n). \tag{A84}$$

Hence,

$$G_1(X^n \mid \text{G-Dict}(Y^n)) \leq (1 - \epsilon) \sum_{i=1}^{n} \epsilon^{i-1} K_1(2^{n-i}) + \sum_{i=1}^{n} 2^{i-2} \cdot \left(\epsilon^{n-i+1} - \epsilon^n\right) + \epsilon^n K_1(2^n). \tag{A85}$$

Noting that $K_1(M) = \frac{M+1}{2}$, we get

$$G_1(X^n \mid \text{G-Dict}(Y^n)) \leq 2^{n-1} \frac{(1 - \epsilon)}{\epsilon} \sum_{i=1}^{n} \left(\frac{\epsilon}{2}\right)^i + \frac{(1 - \epsilon)(1 - \epsilon^n)}{2\epsilon} + \frac{1}{4} \sum_{i=1}^{n} \left(\frac{2}{\epsilon}\right)^i \cdot \epsilon^{n+1} - \frac{1}{2}\left(2^n - 1\right)\epsilon^n + 2^{n-1}\epsilon^n + \frac{\epsilon^n}{2} \tag{A86}$$

$$= \frac{1}{2 - \epsilon}\left(2^{n-1} + \frac{\epsilon^n}{2}(1 - \epsilon)\right) + \frac{(1 - \epsilon)(1 - \epsilon^n)}{2\epsilon} \tag{A87}$$

$$\doteq \frac{2^{n-1}}{2 - \epsilon}. \tag{A88}$$

□

References

1. Arikan, E. An inequality on guessing and its application to sequential decoding. *IEEE Trans. Inf. Theory* **1996**, *42*, 99–105. [CrossRef]
2. Arikan, E.; Merhav, N. Guessing subject to distortion. *IEEE Trans. Inf. Theory* **1998**, *44*, 1041–1056. [CrossRef]
3. Merhav, N.; Arikan, E. The Shannon cipher system with a guessing wiretapper. *IEEE Trans. Inf. Theory* **1999**, *45*, 1860–1866. [CrossRef]
4. Hayashi, Y.; Yamamoto, H. Coding theorems for the Shannon cipher system with a guessing wiretapper and correlated source outputs. *IEEE Trans. Inf. Theory* **2008**, *54*, 2808–2817. [CrossRef]
5. Hanawal, M.K.; Sundaresan, R. The Shannon cipher system with a guessing wiretapper: General sources. *IEEE Trans. Inf. Theory* **2011**, *57*, 2503–2516. [CrossRef]
6. Christiansen, M.M.; Duffy, K.R.; du Pin Calmon, F.; Médard, M. Multi-user guesswork and brute force security. *IEEE Trans. Inf. Theory* **2015**, *61*, 6876–6886. [CrossRef]
7. Yona, Y.; Diggavi, S. The effect of bias on the guesswork of hash functions. In Proceedings of the 2017 IEEE International Symposium on Information Theory (ISIT), Aachen, Germany, 25–30 June 2017; pp. 2248–2252.
8. 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.
9. Arikan, E. Large deviations of probability rank. In Proceedings of the 2000 IEEE International Symposium on Information Theory, Washington, DC, USA, 25–30 June 2000; p. 27.
10. Christiansen, M.M.; Duffy, K.R. Guesswork, large deviations, and Shannon entropy. *IEEE Trans. Inf. Theory* **2012**, *59*, 796–802. [CrossRef]
11. Pfister, C.E.; Sullivan, W.G. Rényi entropy, guesswork moments, and large deviations. *IEEE Trans. Inf. Theory* **2004**, *50*, 2794–2800. [CrossRef]
12. Hanawal, M.K.; Sundaresan, R. Guessing revisited: A large deviations approach. *IEEE Trans. Inf. Theory* **2011**, *57*, 70–78. [CrossRef]
13. Sundaresan, R. Guessing under source uncertainty. *IEEE Trans. Inf. Theory* **2007**, *53*, 269–287. [CrossRef]
14. Serdar, B. Comments on "An inequality on guessing and its application to sequential decoding". *IEEE Trans. Inf. Theory* **1997**, *43*, 2062–2063.
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]
16. Sason, I. Tight bounds on the Rényi entropy via majorization with applications to guessing and compression. *Entropy* **2018**, *20*, 896. [CrossRef]
17. Wyner, A. A theorem on the entropy of certain binary sequences and applications—II. *IEEE Trans. Inf. Theory* **1973**, *19*, 772–777. [CrossRef]
18. Ahlswede, R.; Körner, J. Source coding with side information and a converse for degraded broadcast channels. *IEEE Trans. Inf. Theory* **1975**, *21*, 629–637. [CrossRef]
19. Graczyk, R.; Lapidoth, A. Variations on the guessing problem. In Proceedings of the 2018 IEEE International Symposium on Information Theory, Vail, CO, USA, 17–22 June 2018; pp. 231–235.
20. Graczyk, R. Guessing with a Helper. Master's Thesis, ETH Zurich, Zürich, Switzerland, 2017.
21. O'Donnell, R. *Analysis of Boolean Functions*; Cambridge University Press: Cambridge, UK, 2014.
22. Courtade, T.A.; Kumar, G.R. Which Boolean functions maximize mutual information on noisy inputs? *IEEE Trans. Inf. Theory* **2014**, *60*, 4515–4525. [CrossRef]
23. Ordentlich, O.; Shayevitz, O.; Weinstein, O. An improved upper bound for the most informative Boolean function conjecture. In Proceedings of the 2016 IEEE International Symposium on Information Theory, Barcelona, Spain, 10–15 July 2016; pp. 500–504.
24. Samorodnitsky, A. On the entropy of a noisy function. *IEEE Trans. Inf. Theory* **2016**, *62*, 5446–5464. [CrossRef]
25. Kindler, G.; O'Donnell, R.; Witmer, D. Continuous Analogues of the most Informative Function Problem. Available online: http://arxiv.org/pdf/1506.03167.pdf (accessed on 26 December 2015).
26. Li, J.; Médard, M. Boolean functions: Noise stability, non-interactive correlation, and mutual information. In Proceedings of the 2018 IEEE International Symposium on Information Theory, Vail, CO, USA, 17–22 June 2018; pp. 266–270.
27. Chandar, V.; Tchamkerten, A. Most informative quantization functions. Presented at the 2014 Information Theory and Applications Workshop, San Diego, CA, USA, 9–14 February 2014.

28. Weinberger, N.; Shayevitz, O. On the optimal Boolean function for prediction under quadratic Loss. *IEEE Trans. Inf. Theory* **2017**, *63*, 4202–4217. [CrossRef]
29. Burin, A.; Shayevitz, O. Reducing guesswork via an unreliable oracle. *IEEE Trans. Inf. Theory* **2018**, *64*, 6941–6953. [CrossRef]
30. Ardimanov, N.; Shayevitz, O.; Tamo, I. Minimum Guesswork with an Unreliable Oracle. In Proceedings of the 2018 IEEE International Symposium Information Theory, Vail, CO, USA, 17–22 June 2018; pp. 986–990. Extended Version. Available online: http://arxiv.org/pdf/1811.08528.pdf (accessed on 26 December 2018).
31. Feller, W. *An Introduction to Probability Theory and Its Applications*; John Wiley & Sons: New York, NY, USA, 1971; Volume 2.
32. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; Wiley-Interscience: Hoboken, NJ, USA, 2006.
33. Wainwright, M.J.; Jordan, M.I. Graphical models, exponential families, and variational inference. *Found. Trends® Mach. Learn.* **2008**, *1*, 1–305. [CrossRef]
34. Boyd, S.P.; Vandenberghe, L. *Convex Optimization*; Cambridge University Press: Cambridge, UK, 2004.
35. Nadarajah, S. A generalized normal distribution. *J. Appl. Stat.* **2005**, *32*, 685–694. [CrossRef]
36. Wyner, A.; Ziv, J. A theorem on the entropy of certain binary sequences and applications—I. *IEEE Trans. Inf. Theory* **1973**, *19*, 769–772. doi:10.1109/TIT.1973.1055107. [CrossRef]
37. Erkip, E.; Cover, T.M. The efficiency of investment information. *IEEE Trans. Inf. Theory* **1998**, *44*, 1026–1040. [CrossRef]
38. Ahlswede, R.; Gács, P. Spreading of sets in product spaces and hypercontraction of the Markov operator. *Ann. Probab.* **1976**, 925–939. [CrossRef]
39. Raginsky, M. Strong data processing inequalities and Φ–Sobolev inequalities for discrete channels. *IEEE Trans. Inf. Theory* **2016**, *62*, 3355–3389. [CrossRef]
40. Anantharam, V.; Gohari, A.; Kamath, S.; Nair, C. On hypercontractivity and a data processing inequality. In Proceedings of the 2014 IEEE International Symposium on Information Theory, Honolulu, HI, USA, 29 June–4 July 2014; pp. 3022–3026.