AICR · SOLUTION

LHL for Random Functions

Resolves. The public-seed conjecture of *Randomness Extraction from Unpredictable Random-Oracle Sources: A Leftover-Hash-Lemma Conjecture, Public Seed* (see `main.tex` in this folder) — proven true. Its secret-seed companion, which no longer has a statement note of its own and is recorded below, is not addressed and remains open; a natural but false transposition of it to the public-seed game is also disproven here (Proposition 1). Both statements are reproduced verbatim below.

AI: Claude Opus 5 (Anthropic)
Prompts: Pooya Farshim
Reviewers: Nobody
Date: 14 August 2026

The conjecture, as posed

The two conjectures below share a setting and differ only in whether the distinguisher is given the seed. Conjecture 2 is the whole of `main.tex` in this folder, reproduced verbatim; it is the statement this note resolves. Conjecture 1 is its secret-seed companion, which had a statement note and a page of its own until both were removed from the site on 28 August 2026, and which is reproduced here from that note, verbatim, as the only surviving record of it. It is printed rather than merely cited because Proposition 1 is a statement about an expression derived from it, and Remark 1 compares against it. They are numbered here secret seed first, because the argument below reaches the public-seed statement by way of the secret-seed one; the numbering is this document's own. That shared setting is summarised first and then given in full, in the form used throughout the proof, in Section 2.

Setting. Fix nonempty finite sets $\mathcal{K}$ (seeds), $\mathcal{D}$ (inputs), $\mathcal{R}$ (outputs), and write $K := |\mathcal{K}|$, $D := |\mathcal{D}|$, $R := |\mathcal{R}|$. Let $\mathsf{Fun}(\mathcal{K}\times\mathcal{D},\mathcal{R})$ be the set of all functions $\mathcal{K}\times\mathcal{D} \to \mathcal{R}$, let $\mathrm{SD}(\cdot,\cdot)$ denote statistical distance and $U_{\mathcal{R}}$ the uniform distribution on $\mathcal{R}$. The source S, the predictor P and the distinguisher D are computationally unbounded and receive the *entire* function table of H as an explicit input. In the prediction game Game Pred_P^S the predictor is to guess the source's output x from H and the source's auxiliary output z; in the two extraction games a seed $sd \xleftarrow{\$} \mathcal{K}$ is drawn after S has terminated and D is to tell $H(sd,x)$ from uniform, given H and z. The two extraction games differ only in whether the seed sd is kept secret from D (the game Game Ext) or given to it (the game Game Ext-pub); Game Pred and Game Ext-pub are written out in Figure 1 and Definition 1 below.

Advantages. Set $\mathbf{Adv}^{\mathrm{pred}}_{\mathcal{K},\mathcal{D},\mathcal{R},S}(P) := \Pr[\mathrm{Pred}_P^S \Rightarrow 1]$ and

$$\mathbf{Adv}^{\mathrm{ext}}_{\mathcal{K},\mathcal{D},\mathcal{R}}(S,\mathsf{D}) := 2\Pr[\mathrm{Ext}_{\mathsf{D}}^S \Rightarrow 1]-1, \qquad \mathbf{Adv}^{\mathrm{ext\text{-}pub}}_{\mathcal{K},\mathcal{D},\mathcal{R}}(S,\mathsf{D}) := 2\Pr[\mathrm{Ext\text{-}pub}_{\mathsf{D}}^S \Rightarrow 1]-1.$$

A source S is *ϵ-unpredictable* if $\mathbf{Adv}^{\mathrm{pred}}_{\mathcal{K},\mathcal{D},\mathcal{R},S}(P) \le \epsilon$ for every unbounded predictor P.

Conjecture 1 (Secret seed). *There is a universal constant $c > 0$, independent of $\mathcal{K}$, $\mathcal{D}$, $\mathcal{R}$ and ϵ, such that for all nonempty finite $\mathcal{K}, \mathcal{D}, \mathcal{R}$, all $\epsilon \in (0,1]$, every ϵ-unpredictable source S and every unbounded distinguisher* D,

$$\mathbf{Adv}^{\mathrm{ext}}_{\mathcal{K},\mathcal{D},\mathcal{R}}(S,\mathsf{D}) \;\le\; \delta(\epsilon,K,D,R) \;:=\; c\cdot\sqrt{\frac{\epsilon R + \log_2 D}{K}}\;.$$

Conjecture 2 (Public seed). *There is a universal constant $c > 0$, independent of $\mathcal{K}$, $\mathcal{D}$, $\mathcal{R}$ and ϵ, such that for all nonempty finite $\mathcal{K}, \mathcal{D}, \mathcal{R}$, all $\epsilon \in (0,1]$, every ϵ-unpredictable source S and every unbounded distinguisher* D,

$$\mathbf{Adv}^{\text{ext-pub}}_{\mathcal{K},\mathcal{D},\mathcal{R}}(S,\mathsf{D}) \;\le\; \delta_{\mathrm{pub}}(\epsilon,K,D,R) \;:=\; c\cdot\left(\sqrt{\epsilon R} + \sqrt{\frac{\log_2 D}{K}}\right).$$

Conjecture 2 is what this note settles: it holds with $c = 2$ (Corollary 1), as a consequence of the sharper Theorem 1. Conjecture 1 is a statement about the hidden-seed game Game Ext, is not addressed here, and remains open.

1 Introduction

This note revisits the secret-seed conjecture of Section under the one change that the seed sd is given to the distinguisher. That game is written Game Ext-pub and its advantage $\mathbf{Adv}^{\text{ext-pub}}$, with the suffix carried everywhere, so that no statement here can be mistaken for a statement about the hidden-seed game Game Ext of the companion notes; both are set out in Figure 1 and Definition 1. The conjecture below is Conjecture 1, the secret-seed conjecture, stated there for $\mathbf{Adv}^{\mathrm{ext}}$ and transposed verbatim to $\mathbf{Adv}^{\text{ext-pub}}$. It is not Conjecture 2, the separate and weaker conjecture posed for the public-seed game; that one is treated in Corollary 1 and is not refuted.

Conjecture 3 (Conjecture 1 transposed to the public-seed game). *There is a universal constant $c > 0$, independent of $\mathcal{K}$, $\mathcal{D}$, $\mathcal{R}$ and ϵ, such that for all nonempty finite $\mathcal{K}, \mathcal{D}, \mathcal{R}$, all $\epsilon \in (0,1]$, every ϵ-unpredictable source S and every unbounded distinguisher* D,

$$\mathbf{Adv}^{\text{ext-pub}}_{\mathcal{K},\mathcal{D},\mathcal{R}}(S,\mathsf{D}) \;\le\; \delta(\epsilon,K,D,R) \;:=\; c\cdot\sqrt{\frac{\epsilon R + \log_2 D}{K}}\;.$$

Three things happen. Conjecture 3 becomes *false*, and demonstrably so: a maximum-entropy source with no auxiliary output already refutes it (Proposition 1). The proof technique of the companion proof note, on the other hand, survives essentially unchanged, and delivers

$$\mathbf{Adv}^{\text{ext-pub}}_{\mathcal{K},\mathcal{D},\mathcal{R}}(S,\mathsf{D}) \;\le\; \frac{1}{\sqrt{2}}\sqrt{\epsilon R} \;+\; \min\left\{\frac{6}{5}\sqrt{\frac{1+\ln D}{K}}\;,\;\sqrt{\frac{\ln(2eD\epsilon)}{2K}} + \frac{9}{10\sqrt{K}}\right\}$$

(Theorem 1): the seed's factor $1/K$ is lost from the deficiency term and retained in full by the selection term. The first form inside the brace is the direct transcription of the hidden-seed argument; the second is sensitive to the source's entropy and is the better of the two whenever D is large or ϵ is near its minimum $1/D$, where it has no $\log D$ in it at all. In the closed form of Conjecture 3 the correct statement is $\mathbf{Adv}^{\text{ext-pub}} \le c\sqrt{\epsilon R + \log_2 D/K}$, with the $1/K$ attached to the selection term alone, and $c = 2$ still suffices. Third, Conjecture 2, the public-seed conjecture this note resolves, which already attaches the $1/K$ to the selection term, follows from Theorem 1 with $c = 2$ (Corollary 1); so what Proposition 1 refutes is the transposition, not that conjecture. Sections and lemmas are kept in one-to-one correspondence with the hidden-seed note so that the two can be read against each other.

2 Setting and the two games

Fix nonempty finite sets $\mathcal{K}$ (seeds), $\mathcal{D}$ (inputs) and $\mathcal{R}$ (outputs), and write $K := |\mathcal{K}|$, $D := |\mathcal{D}|$, $R := |\mathcal{R}|$. Let $\mathsf{Fun}(\mathcal{K} \times \mathcal{D}, \mathcal{R})$ be the set of all functions $\mathcal{K} \times \mathcal{D} \to \mathcal{R}$. For distributions P, Q on a countable set Ω, $\mathrm{SD}(P,Q) := \max_{A \subseteq \Omega}\big(P(A) - Q(A)\big) = \frac{1}{2}\sum_{\omega\in\Omega} |P(\omega) - Q(\omega)|$ denotes statistical distance, the maximum being attained at $A^* = \{\omega : P(\omega) > Q(\omega)\}$ whether Ω is finite or not, and $U_{\mathcal{R}}$ is the uniform distribution on $\mathcal{R}$. A *source* S receives a function table H and outputs an input $x \in \mathcal{D}$ together with an auxiliary string $z \in \{0,1\}^*$. The source S, the predictor P and the distinguisher D of Figure 1 are all computationally unbounded, and all three receive the *entire* table of H, all $K \cdot D$ entries, as an explicit input; there is no query budget. The distinguisher now also receives the seed sd, so the only object hidden from D is the coins of S.

2.1 Standard facts

The seven facts below are the whole of what the note borrows from outside itself, apart from the one result of the companion note quoted as Fact 8. They are collected here so that the arguments of Sections 3 and 5 rest on nothing that is not written down.

Fact 1 (Optimal distinguishing advantage). Let P_0 and P_1 be distributions on a countable set Ω. Sample $b \xleftarrow{\$} \{0,1\}$ and then $\omega \xleftarrow{\$} P_b$. For every function $\mathsf{A}\colon \Omega \to \{0,1\}$, possibly randomised and subject to no complexity bound,

$$2\Pr[\mathsf{A}(\omega) = b] - 1 \;\le\; \mathrm{SD}(P_0, P_1),$$

with equality for the maximum-a-posteriori test, which returns 1 exactly when $P_1(\omega) > P_0(\omega)$. In particular the maximum over A is attained.

Fact 2 (Bounded differences; McDiarmid [1]). Let $V_1, \ldots, V_N$ be independent random variables, V_i taking values in a set $\mathcal{V}_i$, and let the measurable function $g\colon \mathcal{V}_1 \times \cdots \times \mathcal{V}_N \to \mathbb{R}$ satisfy

$$|g(v) - g(v')| \;\le\; c_i \qquad \text{whenever } v \text{ and } v' \text{ differ in the } i\text{th coordinate only.}$$

Then for every $\lambda > 0$,

$$\Pr\big[\, g(V_1,\ldots,V_N) \;\ge\; \mathbb{E}\, g(V_1,\ldots,V_N) + \lambda \,\big] \;\le\; \exp\Big(-\frac{2\lambda^2}{\sum_{i=1}^N c_i^2}\Big).$$

Fact 3 (Binomial estimate). For integers $1 \le t \le D$, $\binom{D}{t} \le (eD/t)^t$.

Fact 4 (Mean absolute deviation). For a real random variable X of finite variance, $\mathbb{E}\,|X - \mathbb{E}X| \le \sqrt{\operatorname{Var} X}$.

Fact 5 (Jensen's inequality). Let $I \subseteq \mathbb{R}$ be an interval, $\varphi\colon I \to \mathbb{R}$ concave, and X an integrable random variable taking values in I. Then $\mathbb{E}[\varphi(X)] \le \varphi(\mathbb{E}X)$. If φ is in addition nondecreasing and $\mathbb{E}X \le m$ with $m \in I$, then $\mathbb{E}[\varphi(X)] \le \varphi(m)$.

Fact 6 (Cauchy–Schwarz). For reals $\alpha_1,\alpha_2,\beta_1,\beta_2$, $(\alpha_1\beta_1 + \alpha_2\beta_2)^2 \le (\alpha_1^2 + \alpha_2^2)(\beta_1^2 + \beta_2^2)$.

Fact 7 (Coverage of a random row). The map $R \mapsto (1 - 1/R)^R$ is increasing on the integers $R \ge 2$, with value $\frac{1}{4}$ at $R = 2$ and supremum e^{-1}. In particular $(1 - 1/R)^R \ge \frac{1}{4}$ for every integer $R \ge 2$.

Game Pred_P^S
$H \xleftarrow{\$} \mathsf{Fun}(\mathcal{K} \times \mathcal{D}, \mathcal{R})$
$(x,z) \xleftarrow{\$} S(H)$
$x' \xleftarrow{\$} P(H,z)$
return $(x = x')$

Game $\mathsf{Ext\text{-}pub}_{\mathsf{D}}^S$
$H \xleftarrow{\$} \mathsf{Fun}(\mathcal{K} \times \mathcal{D}, \mathcal{R})$
$(x,z) \xleftarrow{\$} S(H)$; $sd \xleftarrow{\$} \mathcal{K}$
$y_0 \leftarrow H(sd,x)$; $y_1 \xleftarrow{\$} \mathcal{R}$; $b \xleftarrow{\$} \{0,1\}$
$b' \xleftarrow{\$} \mathsf{D}(H, sd, y_b, z)$
return $(b = b')$

Figure 1: Prediction game (left) and extraction game (right). The first argument of S, P and D is the *full function table* of H; the seed sd *is* given to D, but is sampled only after S has terminated.

Definition 1. For the games of Figure 1, set

$$\mathbf{Adv}^{\mathrm{pred}}_{\mathcal{K},\mathcal{D},\mathcal{R},S}(P) := \Pr[\mathsf{Pred}_P^S \Rightarrow 1], \qquad \mathbf{Adv}^{\mathrm{ext\text{-}pub}}_{\mathcal{K},\mathcal{D},\mathcal{R}}(S,\mathsf{D}) := 2\Pr[\mathsf{Ext\text{-}pub}_{\mathsf{D}}^S \Rightarrow 1] - 1.$$

In Game Ext-pub the distinguisher is called as $b' \xleftarrow{\$} \mathsf{D}(H, sd, y_b, z)$: it is given the seed. The suffix pub is carried throughout to distinguish this from the game Game Ext of the companion notes, which is identical except that D is called as $b' \xleftarrow{\$} \mathsf{D}(H, y_b, z)$ and the seed is withheld; the two advantages $\mathbf{Adv}^{\mathrm{ext\text{-}pub}}$ and $\mathbf{Adv}^{\mathrm{ext}}$ are different quantities and no result about one is a result about the other. A source S is *ϵ-unpredictable* if $\mathbf{Adv}^{\mathrm{pred}}_{\mathcal{K},\mathcal{D},\mathcal{R},S}(P) \le \epsilon$ for every unbounded predictor P; this is a property of S alone and is the same notion in both models.

The hidden-seed game just described is the subject of the companion note, whose bound is quoted here for the comparison drawn in Remark 1. Nothing in this note is proved about $\mathbf{Adv}^{\mathrm{ext}}$, and Fact 8 is used nowhere in the proofs.

Fact 8 (Companion note, hidden-seed model). For all nonempty finite $\mathcal{K}, \mathcal{D}, \mathcal{R}$, all $\epsilon \in (0,1]$, every ϵ-unpredictable source S and every unbounded distinguisher D in the game Game Ext of Definition 1,

$$\mathbf{Adv}^{\mathrm{ext}}_{\mathcal{K},\mathcal{D},\mathcal{R}}(S,\mathsf{D}) \;\le\; \frac{1}{\sqrt{2}}\sqrt{\frac{\epsilon R}{K}} \;+\; \frac{6}{5}\sqrt{\frac{1+\ln D}{K}}.$$

The prediction game is untouched by the change, and the choice not to hand sd to P costs nothing: sd is sampled after S terminates and is therefore independent of (X,H,Z), so a predictor given sd has the same optimal success probability. What the change does affect is the sampling order in Game Ext-pub, and that order is now the entire content of the model. The source must commit to x before sd exists; were S allowed to see sd it could choose x with $H(sd,x)$ whatever value it liked and no bound of any shape would hold.

Write (X,Z) for the output distribution of S, and for a view (H,z) of positive probability put

$$p^{H,z}_x \;:=\; \Pr[X=x \mid H,z], \qquad \epsilon_{H,z} \;:=\; \max_{x\in\mathcal{D}} p^{H,z}_x.$$

For $k\in\mathcal{K}$ let $\nu^{H,z}_k$ denote the conditional law of $H(k,X)$ given $(H,Z)=(H,z)$, that is, the pushforward of $p^{H,z}$ along the single row $H(k,\cdot)$, and put

$$\Delta_{H,z} \;:=\; \frac{1}{K}\sum_{k\in\mathcal{K}} \mathrm{SD}\big(\nu^{H,z}_k, U_{\mathcal{R}}\big).$$

Lemma 1 (Analytic forms). *It holds that* $\max_P \mathbf{Adv}^{\mathrm{pred}}_{\mathcal{K},\mathcal{D},\mathcal{R},S}(P) = \mathbb{E}_{(H,Z)}[\epsilon_{H,Z}]$*, so that S is ϵ-unpredictable if and only if* $\mathbb{E}_{(H,Z)}[\epsilon_{H,Z}] \le \epsilon$*; and*

$$\max_{\mathsf{D}} \mathbf{Adv}^{\mathrm{ext\text{-}pub}}_{\mathcal{K},\mathcal{D},\mathcal{R}}(S,\mathsf{D}) \;=\; \mathbb{E}_{(H,Z)}\big[\Delta_{H,Z}\big]. \tag{1}$$

Proof. Each claim identifies an optimal unbounded adversary with a statistical quantity. For the first, the predictor that outputs a mode of $p^{H,z}$ succeeds with probability $\epsilon_{H,z}$ on the view (H,z), and its success is averaged over (H,Z). For the second, the argument conditions on $(H,Z)=(H,z)$, identifies the optimal advantage with the statistical distance between the two view distributions, and expands that distance over the pair (sd,y_b); since sd is uniform under both hypotheses, the sum factors through the K rows and equals $\Delta_{H,z}$. Averaging over (H,Z) gives (1).

The first claim is unchanged from the hidden-seed setting: given its view (H,z) a predictor succeeds with probability at most $\epsilon_{H,z}$, with equality for the (unbounded) predictor that outputs a mode of $p^{H,z}$; averaging over (H,Z) gives the claim. (The view space $\mathsf{Fun}(\mathcal{K}\times\mathcal{D},\mathcal{R})\times\{0,1\}^*$ is countably infinite, but the maxima over P, and below over D, are attained all the same: the mode predictor and the maximum-a-posteriori test of Fact 1 are defined view by view.)

For the second, condition on $(H,Z)=(H,z)$. The view of D is (H,sd,y_b,z), and the two hypotheses differ only in the pair (sd,y_b): under $b=0$ it is distributed as $(sd,H(sd,X))$ with $sd \xleftarrow{\$} \mathcal{K}$ independent of X, and under $b=1$ as $U_{\mathcal{K}}\otimes U_{\mathcal{R}}$. By Fact 1 the optimal advantage $2\Pr[b'=b]-1$ equals the statistical distance between the two view distributions; since the component (H,Z) has the same law under $b=0$ and $b=1$, and the first coordinate sd is uniform under both hypotheses,

$$\begin{aligned}\mathrm{SD}\big((sd,H(sd,X)) \mid (H,z),\ U_{\mathcal{K}}\otimes U_{\mathcal{R}}\big) &= \tfrac{1}{2}\sum_{k,r}\tfrac{1}{K}\big|\nu_k^{H,z}(r)-\tfrac{1}{R}\big| \\ &= \frac{1}{K}\sum_{k}\mathrm{SD}\big(\nu_k^{H,z},U_{\mathcal{R}}\big) \ = \ \Delta_{H,z},\end{aligned}$$

and averaging over (H,Z) gives (1). □

The one structural change is visible in the display above: the average over k now sits *outside* the statistical distance. In the hidden-seed model D sees a mixture of the K rows and the mixing itself smooths the distribution; here D is told which row it is looking at, and the K rows are graded separately. Everything below follows from that one displacement.

3 The conjectured bound is false in this model

Proposition 1. *For every $c>0$ there are nonempty finite $\mathcal{K},\mathcal{D},\mathcal{R}$, a value $\epsilon\in(0,1]$, an ϵ-unpredictable source S and a distinguisher* D *with*

$$\mathbf{Adv}^{\text{ext-pub}}_{\mathcal{K},\mathcal{D},\mathcal{R}}(S,\mathsf{D}) \ \geq\ \tfrac{1}{4}\ >\ c\sqrt{\frac{\epsilon R+\log_2 D}{K}}.$$

The witness has $z=\perp$ and X uniform on all of $\mathcal{D}$, that is, the largest entropy a source over $\mathcal{D}$ can have.

Proof. A witness is exhibited and its two sides are computed. The source outputs a uniform x and $z=\perp$, and $\mathcal{R}$ is taken with $R=D$; by Lemma 1 it is $(1/D)$-unpredictable, so $\epsilon R=1$ and the conjectured bound reads $c\sqrt{(1+\log_2 D)/K}$. The distinguisher returns 0 when its challenge lies in the image of the row $H(sd,\cdot)$, which its view determines. Under $b=0$ this is correct with probability 1; under $b=1$ it is correct with probability $(1-1/R)^D$, computed from the D i.i.d. entries of the row and bounded below by Fact 7. Combining the two gives advantage at least $\frac{1}{4}$ for every K, and K is then taken large enough to put the conjectured bound below $\frac{1}{4}$.

Take $\mathcal{D}$ and $\mathcal{R}$ with $R=D\geq 2$, let $S(H)$ output $x\xleftarrow{\$}\mathcal{D}$ and $z\leftarrow\perp$, and let $\mathcal{K}$ be arbitrary. For each (H,z) the output X is uniform on $\mathcal{D}$, so $\epsilon_{H,z}=1/D$ and S is $(1/D)$-unpredictable by Lemma 1; thus $\epsilon=1/D$ and $\epsilon R=1$. Let $\mathsf{D}(H,sd,y,z)$ return 0 if y lies in the image of the row $H(sd,\cdot)$ and 1 otherwise. The test is computable from the view because D holds both H and sd. If $b=0$ then $y_0=H(sd,X)$ is in that image by construction, so D is correct with probability 1. If $b=1$ then $y_1\xleftarrow{\$}\mathcal{R}$ is independent of H and sd, and D is correct with probability $\mathbb{E}\big[1-|\mathrm{Im}\,H(sd,\cdot)|/R\big]$; the D entries of the row are i.i.d. uniform on $\mathcal{R}$, so a fixed r is missed with probability $(1-1/R)^D$ and this expectation equals $(1-1/R)^D=(1-1/R)^R$, which is at least $\frac{1}{4}$ by Fact 7. Hence $\mathbf{Adv}^{\text{ext-pub}}=2\cdot\frac{1}{2}\big(1+(1-1/R)^R\big)-1\geq\frac{1}{4}$, uniformly in K. Meanwhile

$c\sqrt{(\epsilon R + \log_2 D)/K} = c\sqrt{(1+\log_2 D)/K} < \frac{1}{4}$ as soon as $K > 16c^2(1+\log_2 D)$, and K is unconstrained. □

The constant $\frac{1}{4}$ is attained rather than approached: by Fact 7 the proposition in fact gives $\mathbf{Adv}^{\text{ext-pub}} = (1-1/R)^R \geq \frac{1}{4}$.

The source of the failure is not the auxiliary output and not adversarial selection: it is that once sd is public, the map $x \mapsto H(sd, x)$ is known to D in full, so the only randomness available is that of X itself and the seed cannot contribute entropy. Any correct bound must therefore contain a term in ϵR that does not decrease in K. That is the only defect: with the $1/K$ moved onto the selection term alone, Conjecture 3 holds with $c = 2$.

4 The corrected bound

Theorem 1. *For all nonempty finite $\mathcal{K}, \mathcal{D}, \mathcal{R}$, all $\epsilon \in (0,1]$, every ϵ-unpredictable source S and every unbounded distinguisher* D *in the game of Figure 1,*

$$\mathbf{Adv}^{\text{ext-pub}}_{\mathcal{K},\mathcal{D},\mathcal{R}}(S, \mathsf{D}) \;\leq\; \frac{1}{\sqrt{2}}\sqrt{\epsilon R} \;+\; \frac{6}{5}\sqrt{\frac{1+\ln D}{K}} \;\leq\; 2\sqrt{\epsilon R + \frac{\log_2 D}{K}}\,, \tag{2}$$

$$\mathbf{Adv}^{\text{ext-pub}}_{\mathcal{K},\mathcal{D},\mathcal{R}}(S, \mathsf{D}) \;\leq\; \frac{1}{\sqrt{2}}\sqrt{\epsilon R} \;+\; \sqrt{\frac{\ln(2eD\epsilon)}{2K}} \;+\; \frac{9}{10\sqrt{K}}\,. \tag{3}$$

As in the hidden-seed setting the last inequality of (2) is not an unconditional fact about the parameters: it holds for all $\epsilon \in (0,1]$ whenever $D \geq 2$, and for $D = 1$ it holds at $\epsilon = 1$, the only value at which an ϵ-unpredictable source exists; it can fail for $D = 1$ and ϵ small. Neither of (2) and (3) implies the other; see Remark 4.

Theorem (Informal; see Theorem 1). Publishing the seed costs exactly one factor of K, and costs it in one summand only. The first summand $\frac{1}{\sqrt{2}}\sqrt{\epsilon R}$ does not improve as the seed space grows; the second, which is what the source can extract by choosing its support after seeing H and signalling it through z, keeps its $1/K$ in full. The two forms differ only in that second summand: (2) carries $1+\ln D$ where (3) carries the entropy-sensitive $\ln(2eD\epsilon)$, and (3) pays an additive $\frac{9}{10\sqrt{K}}$ for the exchange.

Section reproduces two conjectures, one for each game, and neither of them is refuted here: what Proposition 1 kills is Conjecture 3, the transposition of the secret-seed Conjecture 1 to the public-seed game. The public-seed conjecture as actually posed is the weaker Conjecture 2,

$$\mathbf{Adv}^{\text{ext-pub}}_{\mathcal{K},\mathcal{D},\mathcal{R}}(S, \mathsf{D}) \;\leq\; c\left(\sqrt{\epsilon R} \;+\; \sqrt{\tfrac{\log_2 D}{K}}\right), \tag{4}$$

which already carries the factor $1/K$ on the selection term alone. That conjecture survives, and Theorem 1 settles it.

Corollary 1 (The public-seed conjecture as posed). *For all nonempty finite $\mathcal{K}, \mathcal{D}, \mathcal{R}$ with $D \geq 2$, all $\epsilon \in (0,1]$, every ϵ-unpredictable source S and every unbounded distinguisher* D, *the bound* ([4](#)) *holds with* $c = \frac{8}{5}$*; the smallest constant obtainable by this route is* $\frac{6}{5}\sqrt{1+\ln 2} = 1.5615\ldots$*, attained at* $D = 2$*. For* $D = 1$*, the only case left,* (4) *holds with* $c = \frac{1}{\sqrt{2}} + \frac{6}{5} < 1.91$*. Hence* $c = 2$ *suffices throughout the range in which an* ϵ*-unpredictable source exists.*

Proof. For $D \geq 2$ the two summands of (2) are compared with the two summands of (4) separately: the first comparison asks for $c \geq 1/\sqrt{2}$, and the second reduces to bounding $(1 + \ln D)/\log_2 D$, which is decreasing in D and so largest at $D = 2$. For $D = 1$ the second summand of (4) vanishes, so the comparison is made against $\sqrt{\epsilon R}$ alone, using that $D = 1$ forces $\epsilon = 1$ and hence $\epsilon R \geq 1$.

By (2) it suffices that $\frac{1}{\sqrt{2}} \leq c$ and that $\frac{6}{5}\sqrt{(1 + \ln D)/K} \leq c\sqrt{\log_2 D/K}$ for every $D \geq 2$, that is, that $c \geq \frac{6}{5}\sqrt{h(D)}$ where $h(D) := (1+\ln D)/\log_2 D$. Writing $\log_2 D = \ln D/\ln 2$ gives $h(D) = (\ln 2)(1 + 1/\ln D)$, which is decreasing on $D \geq 2$; so $h(D) \leq h(2) = (\ln 2)(1+1/\ln 2) = \ln 2+1 = 1.6931\ldots$ and $\frac{6}{5}\sqrt{h(2)} = 1.5615\ldots$. Since $1.5615 \geq 1/\sqrt{2}$, both requirements hold at $c = 1.5615\ldots$, and a fortiori at $c = \frac{8}{5}$.

For $D = 1$ the second summand of (4) vanishes, so what is needed is $\frac{1}{\sqrt{2}}\sqrt{\epsilon R} + \frac{6}{5}K^{-1/2} \leq c\sqrt{\epsilon R}$. Here $\epsilon = 1$, because $\mathbb{E}[\epsilon_{H,Z}] \geq 1/D = 1$ leaves no smaller value admissible, so $\epsilon R = R \geq 1$; together with $K \geq 1$ this gives $\frac{1}{\sqrt{2}}\sqrt{\epsilon R} + \frac{6}{5}K^{-1/2} \leq (\frac{1}{\sqrt{2}} + \frac{6}{5})\sqrt{\epsilon R} < 1.91\sqrt{\epsilon R}$. □

Proposition 1 therefore refutes the verbatim transposition of the secret-seed conjecture, which is strictly stronger than (4), since $\sqrt{(\epsilon R + \log_2 D)/K} \leq \sqrt{\epsilon R} + \sqrt{\log_2 D/K}$. Both are recorded because the transposition is the natural first guess, and locating where it fails is what identifies the term that cannot decay in K.

The rest of the note proves Theorem 1: Section 5.1 fixes the notation, Section 5.2 establishes the three lemmas the proof rests on, and Section 5.3 puts them together. Section 6 discusses what the public seed does and does not cost.

5 Proof of Theorem 1

5.1 Set-up

For a function table H, a seed $k \in \mathcal{K}$ and a nonempty $T \subseteq \mathcal{D}$ let $\mathrm{emp}^H_{k,T}$ be the empirical distribution of the $|T|$ oracle values in row k over T,

$$\mathrm{emp}^H_{k,T}(r) \;:=\; \frac{1}{|T|}\big|\{x \in T : H(k,x) = r\}\big|,$$

and put

$$F_T(H) \;:=\; \frac{1}{K}\sum_{k\in\mathcal{K}} \mathrm{SD}\big(\mathrm{emp}^H_{k,T}, U_{\mathcal{R}}\big), \qquad \Phi_H(t) \;:=\; \max_{T\subseteq\mathcal{D},\,|T|=t} F_T(H).$$

For a uniform H and each fixed k the $|T|$ values $\{H(k,x)\}_{x\in T}$ are $|T|$ i.i.d. uniform draws from $\mathcal{R}$; the K rows are independent of one another, but that is not needed for the mean, only for the deviation. The pair (F_T, Φ_H) is the exact point of departure from the hidden-seed proof, where the corresponding object $Q_{H,T}$ was a single empirical distribution of Kt draws. Here there are K separate empirical distributions of t draws each, and their distances are averaged.

The proof is four steps, in the same order as before: flatten the source (Lemma 2), compute the mean for one fixed support (Lemma 3), control the deviation *uniformly over all supports of all sizes* (Lemma 4), and average (Section 5.3). Lemma 2 and part (i) of Lemma 4 carry over from the hidden-seed note with $F_\cdot(H)$ in place of $\mathrm{SD}(Q_{H,\cdot}, U_{\mathcal{R}})$ and the centering of Lemma 3 in place of the old one (the deviation argument is untouched, but the centering is inherited, so $\frac{1}{2}\sqrt{R/(Kt)}$ becomes $\frac{1}{2}\sqrt{R/t}$ inside W_1), and the convexity step of Lemma 2 now has to pass through the average over k. Of the substance, only Lemma 3 changes, and it changes by exactly one factor of K. Part (ii) of Lemma 4 is new and applies verbatim to the hidden-seed proof as well.

5.2 The three lemmas

Lemma 2 (Flattening). *Fix H and z, and let $t := \lfloor 1/\epsilon_{H,z} \rfloor$. Then $1 \le t \le D$ and $\Delta_{H,z} \le \Phi_H(t)$.*

The effect is that the source enters the rest of the argument only through the single integer t: Lemmas 3 and 4 are statements about flat supports T with $|T| = t$ and about H, and the unpredictability hypothesis reappears only when t is averaged over (H, Z) in Section 5.3.

Proof. The floor gives $1 \le t \le D$ and the pointwise cap $\|p^{H,z}\|_\infty \le 1/t$. The distributions obeying that cap form a hypersimplex, whose vertices are identified as the uniform distributions $\frac{1}{t}\mathbf{1}_T$ on t-element supports, using that t is an integer; so $p^{H,z}$ is a convex combination of them. The map $p \mapsto \Delta_{H,z}$ is then shown to be convex, being an average over k of a convex function of a linear pushforward, and its value at the vertex $\frac{1}{t}\mathbf{1}_T$ is $F_T(H)$. Convexity along the combination bounds $\Delta_{H,z}$ by $\Phi_H(t)$.

Since $\epsilon_{H,z} = \max_x p_x^{H,z} \in [1/D, 1]$ we have $1 \le t \le D$, and $\|p^{H,z}\|_\infty \le 1/t$ by definition of the floor. The set $\{p \in \mathbb{R}^{\mathcal{D}} : p \ge 0,\ \sum_x p_x = 1,\ p_x \le 1/t\}$ is the hypersimplex, whose vertices are precisely the vectors $\frac{1}{t}\mathbf{1}_T$ with $|T| = t$: a vertex has at most one coordinate strictly inside $(0, 1/t)$, and since t is an integer the mass constraint rules that coordinate out. (Integrality of t is essential here; a polytope with a non-integral cap has fractional vertices.) Hence $p^{H,z} = \sum_j \alpha_j \frac{1}{t}\mathbf{1}_{T_j}$ for some convex weights α_j and t-sets T_j. For each fixed k the map $p \mapsto \nu_k$ is linear, being a pushforward, and $\mathrm{SD}(\cdot, U_{\mathcal{R}})$ is convex; a nonnegative average over k of convex functions of linear maps is convex, so $p \mapsto \Delta_{H,z}$ is convex in $p^{H,z}$. Finally the value of that map at the vertex $\frac{1}{t}\mathbf{1}_T$ is $F_T(H)$, since the pushforward of $\frac{1}{t}\mathbf{1}_T$ along $H(k, \cdot)$ is $\mathrm{emp}^H_{k,T}$. Hence $\Delta_{H,z} \le \sum_j \alpha_j F_{T_j}(H) \le \Phi_H(t)$. □

Lemma 3 (Mean at a fixed support). *For every fixed T with $|T| = t$, $\mathbb{E}_H\big[F_T(H)\big] \le \frac{1}{2}\sqrt{R/t}$.*

Proof. The distance is bounded one row at a time. For fixed k, each $\mathrm{emp}^H_{k,T}(r)$ is an average of t i.i.d. Bernoulli$(1/R)$ variables, so its mean absolute deviation from $1/R$ is bounded by its standard deviation; summing over the R values of r and halving gives $\frac{1}{2}\sqrt{R/t}$ for that row, and the average over the K seeds inherits the bound.

Fix k. Each $\mathrm{emp}^H_{k,T}(r)$ is an average of t i.i.d. Bernoulli$(1/R)$ variables, so Fact 4 gives $\mathbb{E}\,|\mathrm{emp}^H_{k,T}(r) - 1/R| \le \sqrt{\mathrm{Var}} = \sqrt{\frac{1}{t}\frac{1}{R}(1-\frac{1}{R})} \le (tR)^{-1/2}$. Summing over the R values of r and halving gives $\mathbb{E}\,\mathrm{SD}(\mathrm{emp}^H_{k,T}, U_{\mathcal{R}}) \le \frac{1}{2}R(tR)^{-1/2} = \frac{1}{2}\sqrt{R/t}$, and averaging over the K seeds preserves the bound. □

Lemma 3 is the whole of the loss. Its hidden-seed version averages Kt draws and yields $\frac{1}{2}\sqrt{R/(Kt)}$; here the count is t, because each row is graded on its own t values and the average over k happens afterwards, outside the distance.

Lemma 4 (Uniform deviation). *Let $a := 1 + \ln D$, let $b_t := \frac{1}{t}\ln\binom{D}{t}$ for $1 \le t \le D$, put $\psi(u) := e^{-u}/(1-e^{-u})$, and set*

$$W_1 := \max_{1\le t\le D}\Big(\Phi_H(t) - \tfrac{1}{2}\sqrt{\tfrac{R}{t}}\Big), \qquad W_2 := \max_{1\le t\le D}\Big(\Phi_H(t) - \tfrac{1}{2}\sqrt{\tfrac{R}{t}} - \sqrt{\tfrac{b_t}{2K}}\Big),$$

with $W_{i,+} := \max\{W_i, 0\}$. Then for every $u > 0$,

$$\text{(i)}\quad \Pr_H\Big[\,W_1 > \sqrt{\tfrac{a+u}{2K}}\,\Big] \le \psi(u), \qquad\qquad \text{(ii)}\quad \Pr_H\Big[\,W_2 > \sqrt{\tfrac{u}{2K}}\,\Big] \le \psi(u),$$

and hence

$$\text{(i)}\quad \mathbb{E}_H[W_{1,+}] \le \sqrt{\frac{a}{2K}} + \frac{\ln 4}{2\sqrt{2Ka}} \le \frac{6}{5}\sqrt{\frac{a}{K}}, \qquad \text{(ii)}\quad \mathbb{E}_H[W_{2,+}] \le \frac{3}{2}\sqrt{\frac{\ln 2}{2K}} \le \frac{9}{10\sqrt{K}}.$$

The two parts bound the same quantity twice and differ only in how the union over supports is paid for. Part (i) subtracts from $\Phi_H(t)$ a single cost $a = 1 + \ln D$ shared by every size t; part (ii) subtracts the size-dependent $b_t = \frac{1}{t}\ln\binom{D}{t}$, which is what $\binom{D}{t}$ costs at that size and which is 0 at $t = D$. Part (i) is what yields (2) and part (ii) is what yields (3).

Proof. The argument runs in two phases. First, for a fixed T of size t, $F_T(H)$ is regarded as a function of the Kt table entries it depends on; a single entry changes it by at most $1/(Kt)$, so Fact 2 bounds $\Pr[F_T(H) \ge \mathbb{E}\,F_T(H) + \lambda]$ by $e^{-2\lambda^2 Kt}$, and Lemma 3 replaces the mean by $\frac{1}{2}\sqrt{R/t}$.

Second, that bound is summed over all T of all sizes under two choices of λ. Part (i) takes $\lambda = \sqrt{(a+u)/(2K)}$, which does not depend on t, and absorbs $\binom{D}{t}$ into e^{ta}; part (ii) takes $\lambda_t = \sqrt{(b_t+u)/(2K)}$, for which that bound becomes $\binom{D}{t}^{-1}e^{-tu}$ and the binomial coefficient cancels against the union bound. Both leave the geometric sum $\sum_{t\ge1} e^{-tu} = \psi(u)$. The two expectations then follow by integrating $\Pr[W_{i,+} > s]$ over s, substituting for u and splitting each integral at $\ln 2$, the point where $\psi(u) = 1$.

Fix T with $|T| = t$ and regard $F_T(H)$ as a function of the Kt independent values $\{H(k,x)\}_{k\in\mathcal{K},x\in T}$. Changing one of them, say $H(k_0,x_0)$, from r_1 to r_2 affects only the summand $k = k_0$: it shifts $\mathrm{emp}^H_{k_0,T}(r_1)$ by $-1/t$ and $\mathrm{emp}^H_{k_0,T}(r_2)$ by $+1/t$ and leaves the other coordinates alone, so it changes $\frac{1}{2}\sum_r |\mathrm{emp}^H_{k_0,T}(r) - 1/R|$ by at most $1/t$, and changes $F_T(H)$, which carries that summand at weight $1/K$, by at most $1/(Kt)$. The bounded-differences constant is therefore $1/(Kt)$, exactly as in the hidden-seed setting, and Fact 2 with $\sum_i c_i^2 = Kt \cdot (Kt)^{-2} = 1/(Kt)$ gives, for every $\lambda > 0$,

$$\Pr\big[F_T(H) \geq \mathbb{E}\, F_T(H) + \lambda\big] \;\leq\; \exp(-2\lambda^2 Kt).$$

The independence hypothesis of Fact 2 is met because H is a uniform *function* table, so its KD entries are i.i.d.; were each row $H(k,\cdot)$ a random permutation or injection, the t values in a row would be sampled without replacement and the inequality would not apply as stated. The same proviso governs the hidden-seed note. Since Lemma 3 gives $\mathbb{E}\, F_T(H) \leq \frac{1}{2}\sqrt{R/t}$, the event $F_T(H) > \frac{1}{2}\sqrt{R/t} + \lambda$ is contained in the event bounded above, for every $\lambda > 0$ we care to choose. The two parts differ only in that choice.

(i). Take $\lambda := \sqrt{(a+u)/(2K)}$, which does *not* depend on t; then the right-hand side is $e^{-t(a+u)}$. Union bounding over all T of all sizes, using Fact 3 in the form $\binom{D}{t} \leq (eD/t)^t \leq e^{t(1+\ln D)} = e^{ta}$,

$$\Pr\Big[W_1 > \lambda\Big] \;\leq\; \sum_{t=1}^{D} \binom{D}{t} e^{-t(a+u)} \;\leq\; \sum_{t\geq 1} e^{ta} e^{-t(a+u)} \;=\; \sum_{t\geq 1} e^{-tu} \;=\; \psi(u).$$

For the expectation, substitute $s = \sqrt{(a+u)/(2K)}$ in $\mathbb{E}[W_{1,+}] = \int_0^\infty \Pr[W_{1,+} > s]\, ds$, the range $s \leq \sqrt{a/(2K)}$ being bounded by $\Pr \leq 1$:

$$\begin{aligned}\mathbb{E}[W_{1,+}] \;&\leq\; \sqrt{\tfrac{a}{2K}} + \int_0^\infty \min\{1, \psi(u)\}\, \frac{du}{2\sqrt{2K(a+u)}} \\ &\leq\; \sqrt{\tfrac{a}{2K}} + \frac{1}{2\sqrt{2Ka}} \int_0^\infty \min\{1, \psi(u)\}\, du.\end{aligned}$$

Since $\psi(u) \geq 1$ exactly for $u \leq \ln 2$, the last integral equals $\ln 2 + \big[\ln(1 - e^{-u})\big]_{\ln 2}^\infty = 2\ln 2 = \ln 4$. Finally $a \geq 1$ gives $\sqrt{a/(2K)}\,(1 + \frac{\ln 4}{2a}) \leq \frac{1.6932}{\sqrt{2}}\sqrt{a/K} \leq \frac{6}{5}\sqrt{a/K}$.

(ii). The bound $\binom{D}{t} \leq e^{ta}$ is lossy for large t, and charging each size its true cost costs nothing. Take instead $\lambda_t := \sqrt{(b_t + u)/(2K)}$, which does depend on t; the right-hand side of the bound of Fact 2 is now $e^{-t(b_t+u)} = \binom{D}{t}^{-1} e^{-tu}$, so the binomial coefficient cancels against the union bound instead of being dominated by it:

$$\Pr\Big[W_2 > \sqrt{\tfrac{u}{2K}}\Big] \;\leq\; \sum_{t=1}^{D} \binom{D}{t}\binom{D}{t}^{-1} e^{-tu} \;\leq\; \sum_{t\geq 1} e^{-tu} \;=\; \psi(u),$$

where the first inequality uses that $W_2 > \sqrt{u/(2K)}$ forces $\Phi_H(t) - \frac{1}{2}\sqrt{R/t} > \sqrt{b_t/(2K)} + \sqrt{u/(2K)} \geq \lambda_t$ for some t, by subadditivity of the square root. For the expectation, substitute $s = \sqrt{u/(2K)}$, so that $ds = du/(2\sqrt{2Ku})$ and no range needs separate treatment:

$$\mathbb{E}[W_{2,+}] \le \int_0^\infty \min\{1,\psi(u)\}\,\frac{du}{2\sqrt{2Ku}} = \frac{1}{2\sqrt{2K}}\int_0^\infty \frac{\min\{1,\psi(u)\}}{\sqrt{u}}\,du.$$

Splitting at $\ln 2$ again, the first piece is $\int_0^{\ln 2} u^{-1/2}du = 2\sqrt{\ln 2}$ and the second is at most $(\ln 2)^{-1/2}\int_{\ln 2}^\infty \psi(u)\,du = (\ln 2)^{-1/2}\ln 2 = \sqrt{\ln 2}$, so the integral is at most $3\sqrt{\ln 2}$ and $\mathbb{E}[W_{2,+}] \le \frac{3}{2}\sqrt{\ln 2/(2K)} \le \frac{9}{10}K^{-1/2}$, the last step because $\frac{3}{2}\sqrt{\ln 2/2} < 0.8831$. □

Note that $b_D = 0$: at $t = D$ part (ii) charges no selection cost whatever. The assembling step below nevertheless reintroduces a constant there, through the two estimates $b_t \le \ln(eD/t)$ and $t \ge 1/(2\epsilon_{H,z})$; carrying b_t itself further would let the selection term vanish at $\epsilon = 1/D$, as Remark 4 argues it should.

The point to record is that the concentration is *identical* to the hidden-seed case even though the mean is larger by $\sqrt{K}$. A single table entry moves $Q_{H,T}$, an average of Kt values, by $1/(Kt)$; it moves F_T, an average of K statistics of t values each, by $\frac{1}{K}\cdot\frac{1}{t}$, which is the same number. Revealing the seed costs the mean and leaves the fluctuation alone.

5.3 Assembling

Lemma 2 and the definitions of W_1 and W_2 bound $\Delta_{H,z}$ by $\frac{1}{2}\sqrt{R/t}$ plus a deviation, in two forms. The estimate $t \ge 1/(2\epsilon_{H,z})$ turns the first summand into $\frac{1}{2}\sqrt{2R\epsilon_{H,z}}$, and the deviations depend on H alone, so they average to the constants of Lemma 4. Taking $\mathbb{E}_{(H,Z)}$ and applying Fact 5 with $\mathbb{E}[\epsilon_{H,Z}] \le \epsilon$ gives (2); for (3), b_t is first bounded by $\ln(2eD\epsilon_{H,z})$ and Fact 5 applied a second time, on the range where $s \mapsto \sqrt{\ln(2eDs)}$ is concave. The closed form of (2) follows by bounding the ratio of the two bounds, by Fact 6 for $D \ge 2$ and directly for $D = 1$.

Write $t := \lfloor 1/\epsilon_{H,z}\rfloor$. By Lemma 2 and the definitions of W_1 and W_2, for every (H,z),

$$\Delta_{H,z} \le \frac{1}{2}\sqrt{\frac{R}{t}} + W_{1,+}, \qquad \text{and} \qquad \Delta_{H,z} \le \frac{1}{2}\sqrt{\frac{R}{t}} + \sqrt{\frac{b_t}{2K}} + W_{2,+}.$$

For $y \ge 1$ one has $\lfloor y\rfloor \ge y/2$, so $t \ge 1/(2\epsilon_{H,z})$ and the first summand of each is at most $\frac{1}{2}\sqrt{2R\,\epsilon_{H,z}}$. Both $W_{1,+}$ and $W_{2,+}$ are functions of H alone, and the H-marginal of the joint (H,Z) is uniform, Z being computed from H and the source's coins, so $\mathbb{E}_{(H,Z)}[W_{i,+}] = \mathbb{E}_H[W_{i,+}]$. Taking $\mathbb{E}_{(H,Z)}$ in the first inequality, using (1) and Fact 5 for the concave nondecreasing $\sqrt{\cdot}$ with $\mathbb{E}[\epsilon_{H,Z}] \le \epsilon$,

$$\mathbf{Adv}^{\text{ext-pub}}_{\mathcal{K},\mathcal{D},\mathcal{R}}(S,\mathsf{D}) \le \frac{1}{\sqrt{2}}\sqrt{\epsilon R} + \mathbb{E}_H[W_{1,+}] \le \frac{1}{\sqrt{2}}\sqrt{\epsilon R} + \frac{6}{5}\sqrt{\frac{1+\ln D}{K}},$$

which is the first inequality of (2). For (3), use the second: $b_t \le \ln(eD/t)$ by Fact 3, and $t \ge 1/(2\epsilon_{H,z})$ gives $eD/t \le 2eD\epsilon_{H,z}$, so the middle summand is at most $\sqrt{\ln(2eD\epsilon_{H,z})/(2K)}$. The function $g(s) := \sqrt{\ln(2eDs)}$ is concave wherever

$\ln(2eDs) > 0$, and $\epsilon_{H,z} \geq 1/D$ gives $2eD\epsilon_{H,z} \geq 2e$, so Fact 5 applies on the range of $\epsilon_{H,Z}$; being also nondecreasing there, g satisfies $\mathbb{E}\, g(\epsilon_{H,Z}) \leq g(\mathbb{E}\,\epsilon_{H,Z}) \leq g(\epsilon)$, whence

$$\mathbf{Adv}^{\text{ext-pub}}_{\mathcal{K},\mathcal{D},\mathcal{R}}(S,\mathrm{D}) \;\leq\; \frac{1}{\sqrt{2}}\sqrt{\epsilon R} \;+\; \sqrt{\frac{\ln(2eD\epsilon)}{2K}} \;+\; \frac{9}{10\sqrt{K}}.$$

For the closed form of (2), note first that an ϵ-unpredictable source exists only for $\epsilon \geq 1/D$, since $\max_x p_x^{H,z} \geq 1/D$ pointwise and hence $\mathbb{E}[\epsilon_{H,Z}] \geq 1/D$; in particular $D = 1$ forces $\epsilon = 1$. Put $P := \epsilon R$, $L := \log_2 D$, and $f(P,L) := \left(\frac{1}{\sqrt{2}}\sqrt{P} + \frac{6}{5}\sqrt{(1+(\ln 2)L)/K}\right)/\sqrt{P + L/K}$, the ratio of the two bounds. If $D = 1$ then $L = 0$ and $P = \epsilon R = R \geq 1$, since $\epsilon = 1$, so $f = \frac{1}{\sqrt{2}} + \frac{6}{5}(KP)^{-1/2} \leq \frac{1}{\sqrt{2}} + \frac{6}{5} < 1.91$. If $D \geq 2$ then $L \geq 1$, and Fact 6 applied to the pairs $\left(\frac{1}{\sqrt{2}}, \frac{6}{5}\sqrt{(1+(\ln 2)L)/L}\right)$ and $(\sqrt{P}, \sqrt{L/K})$ gives

$$f \;\leq\; \sqrt{\tfrac{1}{2} + \tfrac{36}{25}\left(\tfrac{1}{L} + \ln 2\right)} \;\leq\; \sqrt{\tfrac{1}{2} + \tfrac{36}{25}(1 + \ln 2)} \;<\; 1.72.$$

So $f < 1.91$ whenever $D \geq 2$, for every $\epsilon \in (0,1]$, the step of Fact 6 having used only $P \geq 0$; and $f < 1.91$ for $D = 1$, $\epsilon = 1$. In particular the constant $c = 2$ of the corrected conjecture suffices throughout the satisfiable range. (For $D = 1$ the restriction to $\epsilon = 1$ is necessary: for $D = R = 1$ and ϵ small the second inequality fails.) This proves both inequalities of (2) and the inequality (3), which together are Theorem 1. □

6 Discussion

Five things are worth separating: what the public seed costs and what it leaves intact (Remark 1); where the shape of the bound comes from, and which part of the hidden-seed intuition does not survive the change (Remark 2); how far the two summands are matched from below, and which half of that is proved rather than sketched (Remark 3); which of (2) and (3) to prefer at given D and ϵ (Remark 4); and, last, which three features of the model the proof actually consumes.

Remark 1 (What the public seed costs, and what it does not). Comparing with the hidden-seed bound of Fact 8, exactly one factor $1/K$ is lost, and it is lost from the first summand only. The reason is visible in Lemma 1: the average over seeds moves from inside the statistical distance to outside it. Inside, it mixes K rows into one distribution supported on up to K/ϵ values, and the mixture is smooth; outside, each row is judged on its own $1/\epsilon$ values and no amount of averaging over rows repairs a row that is individually far from uniform. Accordingly the available entropy drops from $\log_2 K + \log_2(1/\epsilon)$ to $\log_2(1/\epsilon)$, and $\frac{1}{\sqrt{2}}\sqrt{\epsilon R}$ is the classical leftover-hash shape with an entropy loss of $2\log_2(1/\epsilon)$ bits. The selection term keeps its $1/K$ in full, by Lemma 4. So the correct summary is not that the seed becomes useless: it becomes useless against entropy deficiency, and remains exactly as effective against adversarial selection.

Remark 2 (Where the shape comes from). The step that decides the theorem is again the choice of a deviation λ *independent of* t in Lemma 4. The union-bound cost of ranging over supports of size t is $\ln\binom{D}{t} \approx t\ln D$, while $F_T(H)$ is sub-Gaussian with variance proxy $1/(4Kt)$. In the hidden-seed model that variance proxy came from $Q_{H,T}$ averaging Kt oracle values; here it comes from F_T averaging K row-statistics with weight $1/K$ each, a statistic of t values having bounded-differences constant $1/t$. The arithmetic is the same and the two factors of t cancel, so the required deviation is $\sqrt{t\ln D/(Kt)} = \sqrt{\ln D/K}$ for every t at once, and the selection term enters additively. What does *not* carry over is the accompanying intuition about min-entropy: it is no longer true that a source with more entropy averages away more bias, since each row now sees only t values. The cancellation here is between the union-bound cost and the $1/K$ from averaging over rows, which is a statement about the seed space rather than about the source. Nor does the term enter uniformly in ϵ, as (3) shows: what cancels exactly is t against t, and what survives is the per-size cost $\frac{1}{t}\ln\binom{D}{t}$, which is $\ln D$ only for $t \ll D$ and vanishes at $t = D$. The uniform-in-ϵ reading is an artefact of estimating $\frac{1}{t}\ln\binom{D}{t}$ by $1 + \ln D$.

Remark 3 (Matching lower bounds, sketch). The proof gives $O(\sqrt{\epsilon R}) + O(\sqrt{\log D/K})$. About the first summand two different things can be said, and it is worth separating them. That it must not decay in K at all is *proved*, not sketched: Proposition 1 exhibits, at the single point $\epsilon R = 1$, a source with advantage at least $\frac{1}{4}$ against $\frac{1}{\sqrt{2}}\sqrt{\epsilon R} = 0.708$, uniformly in K. That the shape $\sqrt{\epsilon R}$ and the constant are right throughout $\epsilon R \le 1$ is a sketch, and one point cannot establish it; the witness is instead the fixed-support flat source (T fixed, $|T| = t = 1/\epsilon$, $z = \perp$), for which $\Delta_{H,z} = F_T(H)$ exactly, so that $\mathbb{E}_H[\Delta_{H,z}] = \mathbb{E}\,\mathrm{SD}(\mathrm{emp}^H_{k,T}, U_{\mathcal{R}})$ approaches $\sqrt{2/\pi}\cdot\frac{1}{2}\sqrt{R/t} = 0.3989\sqrt{\epsilon R}$ for $R \le t$, each $|\mathrm{emp}^H_{k,T}(r) - 1/R|$ being asymptotically half-normal. Granting that estimate, the constant of Lemma 3 is tight within $0.5/0.3989 = 1.25$ and the $\frac{1}{\sqrt{2}}$ of (2) within 1.77, a sharper comparison than the factor $0.708/0.25 = 2.83$ that Proposition 1 gives, but a sketch where the proposition is a proof. For the second we only sketch. Fix $A \subseteq \mathcal{R}$ with $|A| = R/2$, let S inspect the whole table, let T consist of the $1/\epsilon$ inputs x maximising $|\{k : H(k,x) \in A\}|$, and let S output $x \xleftarrow{\$} T$ with $z \leftarrow T$. Since $F_T(H) \ge \left|\frac{1}{K}\sum_k(\mathrm{emp}^H_{k,T}(A) - \frac{1}{2})\right|$, extremal-value considerations over the D columns give a bias of order $\sqrt{\log(D\epsilon)/K}$, which is of order $\sqrt{\log D/K}$ only when $1/\epsilon \le D^{1-\Omega(1)}$: selecting the t most biased of D columns buys $\sqrt{\ln(D/t)/K}$, not $\sqrt{\ln D/K}$, and at $t = D$ there is no selection freedom at all. That $\sqrt{\log(D\epsilon)/K}$ is the shape (3) matches, and it is why we state that bound alongside (2); whether the constants can be brought together we do not settle here.

Remark 4 (Which of the two bounds to use). Neither of (2) and (3) implies the other, since the additive $\frac{9}{10}K^{-1/2}$ of the second is the price of dropping $\ln D$ in favour of $\ln(2eD\epsilon)$. Both selection terms are of the form $\gamma/\sqrt{K}$, so the comparison is between the two constants and is independent of K and R. At $\epsilon = 1$ the constants are $\frac{6}{5}\sqrt{1+\ln D}$ and $\sqrt{\ln(2eD)/2} + \frac{9}{10}$, and the second is smaller from $D \ge 26$ onwards (at $D = 25$ they are 2.4648 against 2.4672, at $D = 26$ they are 2.4762 against 2.4734); asymptotically it is smaller by a factor $\frac{6}{5}\sqrt{2} \approx 1.70$. At maximal entropy $\epsilon = 1/D$ the second constant is $\sqrt{\ln(2e)/2} + \frac{9}{10} < 1.83$ *whatever D is*, while the first grows without bound, and the crossover is already at $D = 4$. (Retaining the sharp constant $\frac{3}{2}\sqrt{\ln 2/2} = 0.8831$ of Lemma 4(ii) in place of $\frac{9}{10}$ moves these two thresholds to $D \geq 23$ and 1.81.) The reason is the one recorded in Remark 3: at $\epsilon = 1/D$ the only admissible support is $T = \mathcal{D}$, there is nothing to select, and a selection term growing in D is then an artefact of the estimate $\binom{D}{t} \leq e^{ta}$ rather than a real cost. Everything in this remark and in part (ii) of Lemma 4 applies verbatim to the hidden-seed proof, where it likewise replaces $1 + \ln D$ by $\ln(2eD\epsilon)$.

Remark 5 (What the proof does and does not use). Three features of the model are used, and only these. (a) *sd* is sampled after S terminates and is therefore independent of (X, Z, H): this is what makes the same $p^{H,z}$ appear in every row's pushforward $\nu_k^{H,z}$, and it supplies the $1/K$ in the selection term. Secrecy of *sd* is *not* used anywhere, which is why the argument transfers; independence is used everywhere. (b) The unpredictability hypothesis is used only through $\mathbb{E}_{(H,Z)}[\max_x p_x] \leq \epsilon$, and only via Lemma 2 and one application of Fact 5; no pointwise bound on $\epsilon_{H,z}$ is needed. (c) The source's use of z is never analysed: since Lemma 4 is uniform over *all* subsets of *all* sizes, it covers every measurable rule by which S may pick a support from H and signal it through z. Consequently the theorem holds verbatim for every S whose auxiliary output is an arbitrary function of H and its coins, including $z = H$. Lemma 4 also yields a high-probability refinement, but only for the selection component: with probability $1 - O(e^{-u})$ over H alone, the deviations W_1 and W_2 are small for all supports, hence all sources, simultaneously. No analogous per-H statement holds for the deficiency component: the hypothesis bounds $\epsilon_{H,Z}$ only on average over (H, Z), and an ϵ-unpredictable source may concentrate its predictable mass on an arbitrary fixed H-event of measure ϵ, so the full advantage bound is inherently an average over H.

[1] Colin McDiarmid. On the method of bounded differences. In *Surveys in Combinatorics 1989*, London Mathematical Society Lecture Note Series 141, pages 148–188. Cambridge University Press, 1989. (Fact 2 is the one-sided form of the *independent bounded differences inequality*, Lemma (1.2) there, which is stated two-sidedly with a factor 2; the one-sided bound with the same constant is Theorem (6.7), of which Lemma (1.2) is the independent-coordinate specialisation.)