

A hidden arrow of time

in fair-looking coin flips

An entropy threshold, a concentration theorem,
and a certified symmetry-breaking example

Arya Somu

Self-contained research draft. Proofs and exact checks are included.
Correctness and novelty have not received independent expert review.

Abstract

How much time asymmetry can hide in a stationary binary process whose every length- k window is exactly uniform? We impose an entropy-rate lower bound and solve the resulting extremal problem by reducing it to one scalar optimization. Below a sharp entropy-deficit threshold, every optimizer concentrates all its bias in one pair of reversed contexts. At the threshold, the supremum of the time-reversal relative-entropy rate becomes infinite. The reduced problem has a discontinuous symmetry-breaking transition: equal and opposite biases cease to be globally optimal while remaining strictly locally optimal. An explicit rational example certifies this gap. We also prove that arbitrarily large relative-entropy rate can coexist with near-chance time-direction classification on a prescribed finite recording. This distinction is essential: rare events with extreme likelihood ratios can dominate relative entropy. The underlying locally uniform Markov construction is established prior art; the extremal reduction and its consequences are the candidate contributions investigated here.

The conversation starter

Every 20 consecutive flips can be exactly fair, the entropy rate can exceed 0.999998 bits per flip, and the average log-likelihood evidence for the forward direction can grow at any prescribed rate. Yet even the best classifier given 1,000 flips succeeds at most 50.094% of the time.

1 The question, and a small example

Imagine a machine that produces bits indefinitely. At a randomly chosen time, every block of k successive bits has probability 2^{-k} . Thus a single such block has exactly the law of independent fair coin flips. The machine is stationary: this statement does not depend on when one starts observing. It can nevertheless remember the past and behave differently when its output is read backwards.

This is a statement about the distribution of each window. It does not say that many overlapping windows are independent, or that every finite sample has perfectly balanced empirical counts.

Here is an eight-state example, using the previous three bits as the state. Start with a uniformly random three-bit word, then use the following probabilities for the next bit:

Previous bits	000	001	010	011	100	101	110	111
Pr(next = 0)	1/2	9/10	1/10	1/2	1/2	1/10	9/10	1/2

Every triple has probability $1/8$, forever. But

$$\Pr(0010) = \frac{9}{80}, \quad \Pr(0100) = \frac{1}{80}.$$

These two words are reversals of each other. Four bits reveal a time asymmetry that three bits cannot reveal. No transition is forbidden, and the process is irreducible and aperiodic.

The harder question is quantitative. A fair coin has entropy rate $\log 2$ nats per flip. If we allow only a deficit δ from that rate, what is the largest possible time asymmetry, under the exact uniform-window requirement? Does the optimal machine spread its bias across many contexts, or concentrate it? Does a symmetric choice of opposite biases remain optimal?

What was already known. Locally uniform, globally dependent processes are not new. Ryabko and Monarev constructed “two-faced” processes for randomness testing [1]; finite-memory binary chains also have a standard de Bruijn graph representation [2]. Arola-Fernández and Lacasa explain the general obstruction to detecting binary time irreversibility with words shorter than four [3]. We use these as context, not as discoveries of this notebook.

2 Definitions and a complete normal form

All logarithms are natural unless bits are explicitly specified. Let $\mathcal{C}_k$ be the class of stationary binary Markov processes of order at most k , with strictly positive next-bit probabilities and uniform length- k blocks. For a word w , let w^R denote its reversal. Write $P_n(w) = \Pr(X_1 \cdots X_n = w)$ and $P_n^R(w) = P_n(w^R)$. Define

$$h = \lim_{n \rightarrow \infty} \frac{H(P_n)}{n}, \quad \sigma = \lim_{n \rightarrow \infty} \frac{1}{n} D_{\text{KL}}(P_n \| P_n^R).$$

Here σ is an *observable time-reversal relative-entropy rate*. We do not infer a physical heat dissipation rate from it.

Set $N = 2^{k-1}$. The following parameterization accounts for every member of $\mathcal{C}_k$, not merely a chosen family.

Proposition 1 (Normal form). *Choose one number $\theta_u \in (-1, 1)$ for every binary word u of length $k - 1$. Define the next-bit kernel by*

$$K(b \mid au) = \frac{1 + (-1)^{a+b}\theta_u}{2}, \quad a, b \in \{0, 1\}. \quad (1)$$

The uniform distribution on length- k states is stationary. Every process in $\mathcal{C}_k$ has this form. Its observable reverse has parameters $\theta_u^R = \theta_{u^R}$.

Proof. Uniform stationarity at a target state ub requires $K(b \mid 0u) + K(b \mid 1u) = 1$. Together with the two row sums, this says that the 2×2 matrix indexed by a, b is doubly stochastic. Every such positive matrix has the form (1). Conversely, that identity proves stationarity. Positivity gives every edge of the binary de Bruijn graph, so the state chain is irreducible; the self-loop at 0^k makes it aperiodic. For the reversed process,

$$K^R(b \mid au) = \frac{P_{k+1}(bu^R a)}{P_k(u^R a)} = K(a \mid bu^R) = \frac{1 + (-1)^{a+b}\theta_{u^R}}{2}.$$

This also proves that the reverse is an order-at-most- k process. $\square$

Let

$$A(x) = \operatorname{atanh} x = \frac{1}{2} \log \frac{1+x}{1-x}, \quad (2)$$

$$d(x) = \frac{1}{2} [(1+x) \log(1+x) + (1-x) \log(1-x)]. \quad (3)$$

The function d is even, $d(0) = 0$, $d(1) = \log 2$ by continuity, and $d'(x) = A(x)$. It is the entropy deficit of a coin with bias x .

Proposition 2 (Exact finite-word identities). *For every $n \geq k$,*

$$H(P_n) = k \log 2 + (n - k)h, \quad (4)$$

$$D_{\text{KL}}(P_n \| P_n^R) = (n - k)\sigma, \quad (5)$$

$$\log 2 - h = \frac{1}{N} \sum_u d(\theta_u), \quad (6)$$

$$\sigma = \frac{1}{N} \sum_u D\left(\operatorname{Bern}\left(\frac{1+\theta_u}{2}\right) \parallel \operatorname{Bern}\left(\frac{1+\theta_{u^R}}{2}\right)\right). \quad (7)$$

Equivalently, summing once over each unordered nonpalindromic pair,

$$\sigma = \frac{1}{N} \sum_{\{u, u^R\}} (\theta_u - \theta_{u^R})(A(\theta_u) - A(\theta_{u^R})). \quad (8)$$

Thus $\sigma = 0$ exactly when $\theta_u = \theta_{u^R}$ for every u . For $k \leq 2$, all processes in $\mathcal{C}_k$ are reversible.

Proof. Apply the entropy chain rule with a uniform initial k -block. Each subsequent conditional entropy has the same stationary expectation. For relative entropy, divide the forward path product by the reversed kernel path product from Proposition 1. Both initial distributions are uniform, so their ratio is one. Each of the $n - k$ log-kernel ratios has the same expectation. Averaging the two possible oldest bits a gives (6) and (7). The identity

$$D(\operatorname{Bern}((1+x)/2) \parallel \operatorname{Bern}((1+y)/2)) + D(\operatorname{Bern}((1+y)/2) \parallel \operatorname{Bern}((1+x)/2)) = (x - y)(A(x) - A(y))$$

gives (8). Strict monotonicity of A gives the final claims, since every word of length zero or one is a palindrome. $\square$

3 The extremal problem: a sharp threshold and concentration

For $k \geq 3$, define

$$M_k(\delta) = \sup\{\sigma(P) : P \in \mathcal{C}_k, h(P) \geq \log 2 - \delta\}.$$

Let $a : [0, \log 2] \rightarrow [0, 1]$ be the increasing inverse of d restricted to $[0, 1]$. For $0 < r < \log 2$, define the scalar envelope

$$F(r) = \max_{r/2 \leq s \leq r} [a(s) + a(r-s)] [A(a(s)) + A(a(r-s))], \quad F(0) = 0. \quad (9)$$

Theorem 3 (Exact variational solution). *For $k \geq 3$, $N = 2^{k-1}$, and $\delta \geq 0$,*

$$M_k(\delta) = \begin{cases} N^{-1}F(N\delta), & 0 \leq N\delta < \log 2, \\ +\infty, & N\delta \geq \log 2. \end{cases} \quad (10)$$

For $0 < N\delta < \log 2$, the maximum is attained. Every maximizer has zero bias outside one reversed pair of nonpalindromic contexts. Inside that pair its biases have opposite signs, and their total entropy deficit is $N\delta$.

In bits, the threshold is an allowed entropy deficit of exactly $2^{-(k-1)}$ bits per flip.

The theorem replaces an optimization over 2^{k-1} real parameters with one over a single compact interval. It does not give an elementary closed form for the maximizing split s . That remaining scalar problem has a nontrivial change of optimum, treated below.

3.1 A concentration lemma

Write

$$J(x, y) = (x + y)(A(x) + A(y)), \quad C(x, y) = d(x) + d(y) \quad (x, y \geq 0).$$

Thus $J(x, y)$ is the contribution from opposite biases $x, -y$ before division by N .

Lemma 4 (Increasing return along rays). *For $x, y \geq 0$, not both zero, the ratio*

$$\frac{J(tx, ty)}{C(tx, ty)}$$

is strictly increasing in $t > 0$ while $t \max(x, y) < 1$. Consequently $F(r)/r$ is strictly increasing for $0 < r < \log 2$.

Proof. The convergent series are

$$A(z) = \sum_{m \geq 1} \frac{z^{2m-1}}{2m-1}, \quad d(z) = \sum_{m \geq 1} \frac{z^{2m}}{2m(2m-1)}.$$

The ratio of the coefficients of t^{2m} in $J(tx, ty)$ and $C(tx, ty)$ is

$$R_m = 2m \frac{(x+y)(x^{2m-1} + y^{2m-1})}{x^{2m} + y^{2m}}.$$

Assume $x \geq y$ and set $q = y/x \in [0, 1]$. Then

$$\frac{R_m}{2(1+q)} = m \frac{1+q^{2m-1}}{1+q^{2m}} = m + \frac{m(1-q)q^{2m-1}}{1+q^{2m}}.$$

The last fraction lies in $[0, 1/2]$. Indeed, $2m(1-q)q^{2m-1} \leq (1-q) \sum_{j=0}^{2m-1} q^j \leq 1$. It follows that $R_{m+1} - R_m \geq 1+q > 0$.

For power series with positive denominator coefficients, strictly increasing coefficient ratios imply a strictly increasing quotient. To check this directly, differentiate the quotient after putting $z = t^2$: each paired cross term for indices $m > n$ is a positive multiple of $(m-n)(R_m - R_n)z^{m+n-1}$. All series converge absolutely in the stated domain. This proves the first assertion.

For $0 < r_1 < r_2 < \log 2$, choose a maximizer with cost r_1 and scale both coordinates along its ray until its cost is r_2 . Such scaling is possible: the cost when the larger coordinate first reaches one is at least $\log 2$. The first assertion gives $F(r_2)/r_2 > F(r_1)/r_1$. $\square$

3.2 Proof of the extremal theorem

Proof of Theorem 3. Fix $r = N\delta < \log 2$. By (6), each bias has magnitude at most $a(r) < 1$. Thus the feasible parameter set is compact and all its kernels remain strictly positive.

For a fixed pair of bias magnitudes x, y , opposite signs maximize the Jeffreys contribution. Same signs give $(x-y)(A(x) - A(y))$ instead of $J(x, y)$, a difference of $2[xA(y) + yA(x)] \geq 0$. Palindromic contexts contribute no time asymmetry. Let c_i be the cost allocated to the i th nonpalindromic pair. Then, omitting zero-cost terms and using Lemma 4,

$$N\sigma \leq \sum_i F(c_i) \leq \sum_i c_i \frac{F(r)}{r} \leq F(r).$$

One pair with biases $a(s), -a(r-s)$, where s maximizes (9), attains equality. Such a pair exists because $k \geq 3$. If two pairs had positive cost, each would have $c_i < r$ and the middle inequality would be strict. Spending cost on palindromes, or failing to use the full budget, also makes the bound strict. The endpoint split with one magnitude zero is not optimal when $r > 0$: adding a small opposite magnitude increases J at first order, while keeping cost fixed only changes the other magnitude at second order. Hence the two nonzero biases must have opposite signs.

At $r \geq \log 2$, choose one nonpalindromic u and let $\theta_u = t$, with every other bias zero. Then

$$N(\log 2 - h) = d(t) < \log 2, \quad N\sigma = tA(t) \rightarrow \infty \quad \text{as } t \uparrow 1.$$

All members of this sequence have strictly positive transitions. Every individual rate is finite; its supremum is infinite. $\square$

4 Two ends of the envelope

Proposition 5 (Small budgets and threshold asymptotics). *There is a nonempty interval of budgets on which the unique optimal split is symmetric. In particular, for $0 < r \leq d(1/2)$,*

$$F(r) = 4a(r/2)A(a(r/2)). \tag{11}$$

As $r \downarrow 0$,

$$F(r) = 4r + \frac{2}{3}r^2 + \frac{14}{45}r^3 + O(r^4).$$

As $\eta = \log 2 - r \downarrow 0$,

$$F(r) = \frac{1}{2} \log \frac{1}{\eta} + \frac{1}{2} \log \log \frac{1}{\eta} + o(1). \quad (12)$$

Every maximizing pair, ordered as $x \geq y$, satisfies $x \rightarrow 1$ and $y \rightarrow 0$ in this latter limit.

Proof. For a deficit coordinate s , the inverse $a(s)$ is strictly concave, because $a'(s) = 1/A(a(s))$ is decreasing. Also

$$\frac{d}{ds}A(a(s)) = \frac{1}{(1 - a(s)^2)A(a(s))}.$$

The derivative of this expression with respect to $a(s)$ has the sign of $2a(s)A(a(s)) - 1$. It is negative when $a(s) \leq 1/2$. For $r \leq d(1/2)$, both factors in (9) are therefore maximized, strictly, by $s = r/2$, using concavity. Expanding $r = 2d(x) = x^2 + x^4/6 + x^6/15 + \dots$ and $4xA(x) = 4x^2 + 4x^4/3 + 4x^6/5 + \dots$ gives the stated series.

For the other endpoint, the feasible pair $(a(r), 0)$ makes $F(r)$ diverge. Thus any maximizing pair, ordered $x \geq y$, must have $x \rightarrow 1$. Put $p = (1 - x)/2$. The constraint implies

$$d(y) = r - d(x) \leq \log 2 - d(x) = H(\text{Bern}(p)).$$

Since $d(y) \geq y^2/2$, we have $y \leq \sqrt{2H(\text{Bern}(p))}$, so $y \rightarrow 0$ and $yA(x) \rightarrow 0$. Consequently

$$J(x, y) = A(x) + o(1) \leq A(a(r)) + o(1).$$

The one-sided feasible pair gives the matching lower bound $a(r)A(a(r)) = A(a(r)) + o(1)$. Writing $p_r = (1 - a(r))/2$, the equation $\eta = H(\text{Bern}(p_r)) = p_r \log(1/p_r) + p_r + O(p_r^2)$ gives

$$\log(1/p_r) = \log(1/\eta) + \log \log(1/\eta) + o(1).$$

Since $A(a(r)) = \frac{1}{2} \log((1 - p_r)/p_r)$, this proves (12). $\square$

5 A stable symmetric solution that is already wrong

It is tempting to assume that an optimal reversed pair always has equal and opposite biases. Proposition 5 proves that intuition right near perfect randomness and wrong near the threshold. The change happens before the symmetric solution loses local stability.

Proposition 6 (Certified counterexample). *Set*

$$x = \frac{91}{100}, \quad y = \frac{27}{50}, \quad r = d(x) + d(y), \quad z = a(r/2).$$

Then

$$J(x, y) > J(z, z),$$

but (z, z) is a strict local maximum of J on the curve $d(x) + d(y) = r$.

Proof. A constrained perturbation can be written

$$x = z + t + st^2 + O(t^3), \quad y = z - t + st^2 + O(t^3), \quad s = -\frac{1}{2(1 - z^2)A(z)}.$$

Its objective is

$$J(x, y) = 4zA(z) + \frac{2[3z^2 - 1 - z/A(z)]}{(1 - z^2)^2}t^2 + O(t^3). \quad (13)$$

The following inequalities are certified by rational arithmetic in `verify_exact.py`:

$$\begin{aligned} 0.7655109204 < z < 0.7655109205, \\ J(91/100, 27/50) - 4zA(z) > 0.00012999, \\ (3z^2 - 1)A(z) - z < -0.0003697. \end{aligned}$$

The last two lines respectively prove global suboptimality and strict local maximality. These are not floating-point assertions: the verifier uses the first 180 terms of the positive series for A and d , with rational upper bounds on the omitted tails. Explicitly, after L terms the tails are bounded by

$$\frac{z^{2L+1}}{(2L+1)(1-z^2)}, \quad \frac{z^{2L+2}}{(2L+2)(2L+1)(1-z^2)},$$

respectively. Evaluating these expressions at rational endpoints certifies the bracket for z and both strict inequalities. $\square$

For orientation, $r \approx 0.663496882839465$ nats, or $r/\log 2 \approx 0.9572236625$. In the $k = 3$ machine, this example has $\theta_{01} = 0.91$, $\theta_{10} = -0.54$ and the other two parameters zero. The rate difference above is divided by $N = 4$. The effect is small, but its sign is rigorously determined.

Theorem 7 (Discontinuous symmetry breaking). *There is a budget $r_t \in [d(1/2), d(91/100) + d(27/50))$ at which the symmetric split and an unequal split are both global maximizers of F . At this budget the symmetric split is still a strict local maximum. Along a sequence of budgets decreasing to r_t from above, globally maximizing splits remain a positive distance away from equality.*

Proof. Let $S(r) = 4a(r/2)A(a(r/2))$. The envelope F is continuous on $(0, \log 2)$: parameterize its compact interval by $s/r \in [1/2, 1]$ and use uniform continuity of the objective locally in r . We know $F = S$ for $r \leq d(1/2)$ and $F > S$ at the certified budget $r_c = d(91/100) + d(27/50)$. Define

$$r_t = \inf\{r \in (0, r_c] : F(r) > S(r)\}.$$

Continuity implies $F(r_t) = S(r_t)$ and $r_t < r_c$.

The sign of the quadratic coefficient in (13) is the sign of $\phi(z) = (3z^2 - 1)A(z) - z$. For $0 < z \leq 1/\sqrt{3}$, it is negative. For $z \geq 1/\sqrt{3}$,

$$\phi'(z) = 6zA(z) + \frac{3z^2 - 1}{1 - z^2} - 1 > 0,$$

because $A(z) \geq z$. Thus ϕ has exactly one positive zero, and the rational certificate proves that its sign is negative up through r_c . The symmetric point is consequently a strict local maximum throughout the compact budget interval under consideration, with a uniform neighborhood near r_t in which unequal splits have smaller objective.

Choose $r_j \downarrow r_t$ with $F(r_j) > S(r_j)$ and global maximizers at those budgets. They must stay outside that neighborhood. A convergent subsequence of their allocation fractions has an unequal limit, which by continuity attains $F(r_t) = S(r_t)$. This proves coexistence and the claimed nonzero jump. $\square$

The theorem proves existence of a discontinuous transition, not uniqueness of all branches or its exact location. Numerical exploration estimates the first crossing near $r/\log 2 = 0.95707206$, with unequal magnitudes near $(0.90633, 0.54896)$. Local stability is lost later, near $r/\log 2 = 0.95742062$. These decimals are exploratory estimates; the preceding rational counterexample and existence proof do not depend on them.

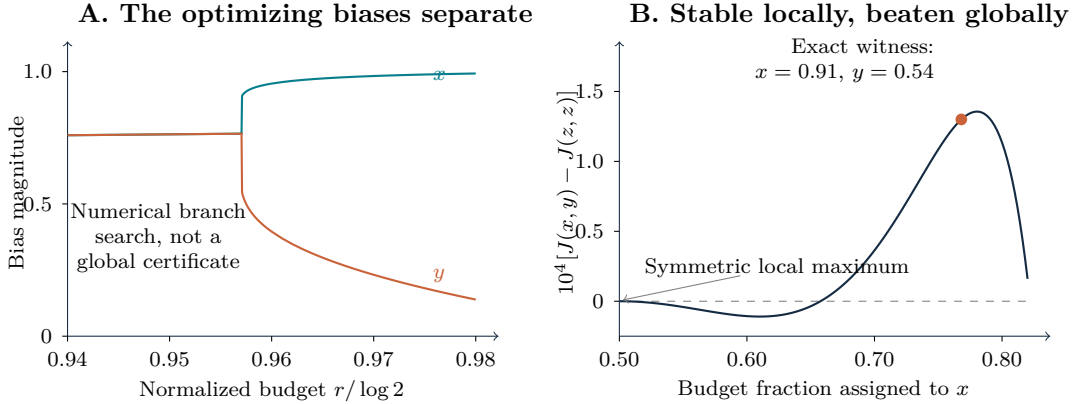


Figure 1: Left: the scalar optimization’s numerical candidates show an abrupt separation of the two magnitudes. Right: at the rational witness budget, small deviations from equal allocation lower the objective, but a larger deviation improves it. The marked witness and local stability are certified with rational bounds in Proposition 6; the plotted curves themselves use floating point.

6 Large relative entropy does not mean an easy guessing game

The term “large arrow of time” can be misleading unless its metric is specified. With equal prior probabilities for a forward or backward recording, the best possible success probability is

$$\text{Acc}_n = \frac{1 + d_{\text{TV}}(P_n, P_n^{\text{R}})}{2}.$$

This depends on total variation, not directly on relative entropy.

Proposition 8 (A finite-recording bound). *For a source whose only nonzero biases are $\theta_u = x$ and $\theta_{u^{\text{R}}} = -y$, with $u \neq u^{\text{R}}$ and $x, y \geq 0$, we have*

$$d_{\text{TV}}(P_{k+1}, P_{k+1}^{\text{R}}) = \frac{x + y}{N}, \quad (14)$$

$$d_{\text{TV}}(P_n, P_n^{\text{R}}) \leq \min \left\{ 1, \frac{(n - k)(x + y)}{N} \right\}, \quad n \geq k. \quad (15)$$

For $n \leq k$, total variation is zero.

Proof. At a context av , the two next-bit distributions have total variation $|\theta_v - \theta_{vR}|/2$. Under the uniform stationary state law, its mean is

$$\frac{1}{2N} \sum_v |\theta_v - \theta_{vR}| = \frac{x + y}{N}.$$

For $n = k + 1$, this is exactly the total variation of the full word laws. For longer words, couple the initial k -blocks identically and maximally couple the two next bits whenever their histories still agree. The probability of first disagreement at a given step is bounded by the unconditional stationary expectation of the displayed conditional total variation. A union bound over $n - k$ steps gives (15). Positivity and stationarity of each marginal are preserved by this coupling. $\square$

Corollary 9 (Nearly random, hard to orient, unbounded divergence). *For every $k \geq 3$ and every prescribed $B > 0$, there is a process in $\mathcal{C}_k$ such that*

$$h_{\text{bits}} > 1 - \frac{1}{2^{k-1}}, \quad \sigma > B,$$

while, for every $n \geq k$,

$$\text{Acc}_n \leq \min \left\{ 1, \frac{1}{2} + \frac{n - k}{2^k} \right\}.$$

Proof. Use the one-sided construction $\theta_u = t$, all other biases zero, and take $t < 1$ sufficiently close to one. Its entropy cost is $d(t)/N < \log 2/N$, its rate is $tA(t)/N \rightarrow \infty$, and Proposition 8 gives the bound, uniformly in t . For an explicit parameter, choose $L = 4N(B + 1)$ and $t = \tanh(L/2)$. Then $t > 1/2$ and $\sigma = tL/(2N) > B$. The smallest next-bit probability is $1/(1 + e^L) > 0$. $\square$

At $k = 20$ and $n = 1000$, the entropy lower bound is $1 - 2^{-19} = 0.999998092651 \dots$ bits per flip, and

$$\text{Acc}_{1000} \leq \frac{1}{2} + \frac{980}{1048576} = 0.500934600830 \dots$$

Meanwhile $D_{\text{KL}}(P_{1000} \| P_{1000}^R) = 980\sigma$ can be arbitrarily large. There is no contradiction: forward and backward laws differ on rare contexts, and a tiny reverse probability makes the log-likelihood ratio on those events enormous. Achieving extreme rates requires extreme transition probabilities, not an ordinary laboratory coin with moderate noise.

7 What a noise floor restores

Proposition 10 (Sharp bound with positive noise). *If all next-bit probabilities lie in $[\varepsilon, 1 - \varepsilon]$, where $0 < \varepsilon < 1/2$, then over $\mathcal{C}_k$*

$$\max \sigma = \left(1 - 2^{-\lfloor (k-1)/2 \rfloor} \right) (1 - 2\varepsilon) \log \frac{1 - \varepsilon}{\varepsilon}. \quad (16)$$

This maximum does not simultaneously impose an additional entropy budget.

Proof. The noise bound is equivalent to $|\theta_u| \leq t = 1 - 2\varepsilon$. For each nonpalindromic pair, the largest Jeffreys contribution is $4tA(t)$, attained by $t, -t$; this follows either from monotonicity of $J(x, y)$ in both nonnegative coordinates, or directly from (8). There are $(N - 2^{\lfloor (k-1)/2 \rfloor})/2$ such pairs. Assign opposite extreme signs to each pair and set palindromic biases to zero. Counting the contributions proves (16), including attainability. $\square$

8 Verification, provenance, and remaining questions

Exact checks. The companion verifier performed 8,850 exact checks across 21 rational sources, plus the certified asymmetric example. It checked transition normalization, uniform stationarity, short-window marginals, observable reversal, and finite-word entropy and divergence identities. Logarithmic identities were represented as rational coefficient vectors of $\log p$ for primes p , avoiding floating-point comparisons. Further exact total-variation checks cover one-pair sources with $3 \leq k \leq 6$. These finite checks test the implementation and examples; the general theorems rest on the proofs above. This is not a formal proof-assistant verification and not an independent referee review.

Numerical exploration. The script `explore.py` records 336 sampled budgets and locally refines candidate maxima of the scalar problem. It also checks 5,000 random multi-pair configurations against a feasible concentrated construction. These floating-point results are exploratory evidence, not certified global optima. In particular, the estimated transition location is not needed for any theorem in this note.

Novelty boundary. The focused literature search covered two-faced processes, binary de Bruijn processes, symbolic time irreversibility, symmetric-divergence bounds, and Markov entropy optimization. It found close foundations [1, 2, 3], general divergence extremal results [4], and perturbative Markov optimization [5]. None of the inspected material supplied the exact concentration theorem, entropy threshold, or certified stable-but-suboptimal symmetric example in the formulation studied here. The search was limited; priority remains unconfirmed. These are candidate results with proofs, pending independent review and broader novelty checks.

What remains unresolved. The exact global transition budget and uniqueness of the unequal maximizing branch are open within this notebook. Theorem 3 is an exact variational solution, not a closed-form solution of that last scalar maximization. Other worthwhile questions are the joint entropy-and-noise constrained frontier, stability under approximate uniformity of short blocks, and the analogue for larger alphabets, where 2×2 doubly stochastic matrices are replaced by higher dimensional ones. No claim about those extensions is made here.

References

- [1] B. Ya. Ryabko and V. A. Monarev. Using information theory approach to randomness testing. *Journal of Statistical Planning and Inference* 133 (2005), 95–110. doi:10.1016/j.jspi.2004.02.010.
- [2] L. Kimpton, P. Challenor, and H. Wynn. Binary De Bruijn Processes. [arXiv:2211.16921](#) (2022).
- [3] L. Arola-Fernández and L. Lacasa. Irreversibility of symbolic time series: a cautionary tale. *Physical Review E* 108 (2023), 014201. doi:10.1103/PhysRevE.108.014201. [arXiv:2303.11868](#).
- [4] I. Sason. Tight Bounds for Symmetric Divergence Measures and a Refined Bound for Lossless Source Coding. [arXiv:1403.7164](#).
- [5] M. Santos Gutiérrez, N. Zagli, and G. Carigi. Markov matrix perturbations to optimize dynamical and entropy functionals. [arXiv:2507.14040](#) (2025).