Evidence Holonomy and Entropy Production: From Universal Coding to Irreversibility

Joshua Winters
Independent Researcher
josh@friendmachine.co

August 25, 2025

Abstract

We formalize an "evidence holonomy" functional on loops of representation transforms applied to sample paths. Using pointwise universality of code lengths for stationary ergodic processes, we prove two complementary reductions: (i) *representation-space holonomy* (measuring evidence in the final representation reached by the loop) converges, up to $o(n)$, to the entropy-rate difference $h(Q) - h(P)$; (ii) *KL holonomy* with *observer transport* (measuring the forward path under a code universal for the loop pushforward law but evaluated in the *original* coordinates) converges to the *relative entropy rate* $\mathsf{d}(P\|Q)$. As corollaries we obtain: gauge invariance for bijective loops; non-negativity for channel-based loops in the KL sense; and, for finite-state Markov chains under a canonical time-reversal loop, equality between holonomy rate and the *entropy production rate*. We also show observer-independence of the holonomy rate across universal coders. A companion battery of numerical tests validates these statements.

1 Setup and Definitions

Alphabet and path space. Fix a finite alphabet $\mathcal{X}$. Let $\mathcal{X}^n$ denote length-n strings and $\mathcal{X}^{\mathbb{N}}$ the one-sided sequence space with its product σ-algebra. Let P be a stationary ergodic probability measure on $(\mathcal{X}^{\mathbb{N}}, \mathcal{F})$. Write $X_{0:n-1}$ for the length-n prefix of a sample from P and P_n for its law on $\mathcal{X}^n$.

Universal codes. A *universal code* on alphabet $\mathcal{A}$ is a map $\mathcal{E}_{\mathcal{A}} : \bigcup_{n\geq 1} \mathcal{A}^n \to \mathbb{R}_+$ assigning a code length in bits to any finite string, such that for every stationary ergodic law Q on $\mathcal{A}^{\mathbb{N}}$,

$$\frac{1}{n}\Big(\mathcal{E}_{\mathcal{A}}(Y_{0:n-1}) + \log_2 Q_n(Y_{0:n-1})\Big) \xrightarrow[n\to\infty]{Q\text{-a.s.}} 0, \tag{1}$$

with convergence also in $L^1(Q)$. Classical examples include LZ78, Krichevsky–Trofimov mixtures for finite-order Markov models, and CTW [1, 2, 3, 6, 7].

Representation transforms and loops. For each n, let $F_{i,n}$ be a measurable map $F_{i,n} : \mathcal{X}_{i-1}^{n_{i-1}(n)} \to \mathcal{X}_i^{n_i(n)}$, where the alphabets $\mathcal{X}_i$ may differ by step, and $n_0(n) = n$. Define successive images

$$x^{(0)} = x \in \mathcal{X}^n, \quad x^{(i)} = F_{i,n} \circ \cdots \circ F_{1,n}(x^{(0)}) \in \mathcal{X}_i^{n_i(n)}.$$

A finite list $\gamma = (F_{1,n}, \ldots, F_{m,n})$ is a *loop at scale* n if $n_m(n) = n$ and $\mathcal{X}_m = \mathcal{X}_0 = \mathcal{X}$. Let $L_n = F_{m,n} \circ \cdots \circ F_{1,n} : \mathcal{X}^n \to \mathcal{X}^n$ be the loop map and let $Q_n := (L_n)\#P_n$ (hence Q_n is a law on $\mathcal{X}^n$).

1.1 Two holonomy functionals

Definition 1.1 (Representation-space holonomy)**.** Given universal codes $\mathcal{E}_{\mathcal{X}_i}$ for intermediate alphabets, define

$$\begin{aligned}\mathrm{Hol}_n^{\mathrm{out},\gamma}(x) &= \sum_{i=1}^{m} \Big(\mathcal{E}_{\mathcal{X}_i}\big(x^{(i)}\big) - \mathcal{E}_{\mathcal{X}_{i-1}}\big(x^{(i-1)}\big) \Big) \\ &= \mathcal{E}_{\mathcal{X}}\big(L_n(x)\big) - \mathcal{E}_{\mathcal{X}}(x).\end{aligned}$$

Definition 1.2 (KL (observer-transported) holonomy)**.** Let $\mathcal{E}_{\mathcal{X}}^{(P)}$ and $\mathcal{E}_{\mathcal{X}}^{(Q)}$ be universal on $\mathcal{X}$ for P and for the stationary pushforward Q with marginals Q_n, respectively. Define the *code-based* estimator

$$\mathrm{Hol}_n^{\mathrm{KL},\gamma}(x) := \mathcal{E}_{\mathcal{X}}^{(Q)}(x) - \mathcal{E}_{\mathcal{X}}^{(P)}(x).$$

Operationally, $\mathcal{E}_{\mathcal{X}}^{(Q)}$ is the universal code trained on samples from Q (e.g., loop-transformed training sequences) but *evaluated on the original sequence* x; hence it estimates $-\log_2 Q_n(x)$, so that $\mathrm{Hol}_n^{\mathrm{KL},\gamma}$ estimates $\log_2 \frac{P_n(x)}{Q_n(x)}$.

2 Reductions via Universality

Lemma 2.1 (Pointwise reductions)**.** *Assume* (1) *for the relevant laws.*

1. *For* $\mathrm{Hol}^{\mathrm{out}}$*: with* $\mathcal{E}_{\mathcal{X}}$ *universal for both* P *and the pushforward process,*

$$\frac{1}{n}\Big(\mathrm{Hol}_n^{\mathrm{out},\gamma}(X_{0:n-1}) - \log_2 \frac{P_n(X_{0:n-1})}{Q_n(L_n(X_{0:n-1}))}\Big) \xrightarrow[n\to\infty]{P\text{-a.s.}} 0. \tag{2}$$

2. *For* $\mathrm{Hol}^{\mathrm{KL}}$*:*

$$\frac{1}{n}\Big(\mathrm{Hol}_n^{\mathrm{KL},\gamma}(X_{0:n-1}) - \log_2 \frac{P_n(X_{0:n-1})}{Q_n(X_{0:n-1})}\Big) \xrightarrow[n\to\infty]{P\text{-a.s.}} 0. \tag{3}$$

Both convergences also hold in $L^1(P)$*.*

Proof. Apply (1) (Barron's strong pointwise coding theorem) to each code/law pair and subtract the limits; see [5, 7, 6]. □

Averaging yields the two central identities.

Theorem 2.2 (Expectation-level reductions)**.** *Under* L^1 *universality,*

$$\frac{1}{n}\,\mathbb{E}_P\big[\mathrm{Hol}_n^{\mathrm{out},\gamma}\big] = h(Q) - h(P) + o(1), \tag{4}$$

$$\frac{1}{n}\,\mathbb{E}_P\big[\mathrm{Hol}_n^{\mathrm{KL},\gamma}\big] = \frac{1}{n}\,\mathrm{D}(P_n\|Q_n) \xrightarrow[n\to\infty]{} \mathsf{d}(P\|Q) \ \geq 0. \tag{5}$$

Proof. Take expectations in (2)–(3) and note $\mathbb{E}_P[-\log_2 P_n(X_{0:n-1})] = H(P_n)$, $\mathbb{E}_P[-\log_2 Q_n(L_n(X_{0:n-1}))] = H(Q_n)$, and $\mathbb{E}_P\big[\log_2 \frac{P_n(X)}{Q_n(X)}\big] = \mathrm{D}(P_n\|Q_n)$. □

Remark 2.3 (Scope)*.* Equation (4) is gauge-invariant and measures net compression or expansion under the loop. Equation (5) is the *irreversibility* functional implemented in our code (KL-rate holonomy): it is observer-transported and non-negative.

3 Canonical Loops and Corollaries

3.1 Gauge invariance for bijective loops

Corollary 3.1 (Gauge invariance)**.** *If each $F_{i,n}$ is a bijection and the loop is the identity on $\mathcal{X}^n$, then*

$$\frac{1}{n}\,\mathrm{Hol}_n^{\mathrm{out},\gamma}(X_{0:n-1}) \to 0 \quad \textit{and} \quad \frac{1}{n}\,\mathrm{Hol}_n^{\mathrm{KL},\gamma}(X_{0:n-1}) \to 0$$

in P-probability and in $L^1(P)$.

Proof. Then $Q_n = P_n$ for all n, so both (4) and (5) vanish. □

3.2 Coarse-graining loops via channels

Let K_n be a (possibly many-to-one) Markov kernel on $\mathcal{X}^n$ and R_n any measurable right-inverse (a "lift") so that $L_n := R_n \circ K_n : \mathcal{X}^n \to \mathcal{X}^n$ is a loop. If $Q_n := L_n \# P_n$ arises from a stationary Q, then (5) gives

$$\frac{1}{n}\,\mathbb{E}_P\big[\mathrm{Hol}_n^{\mathrm{KL},\gamma}\big] \to \mathsf{d}(P\|Q) \geq 0,$$

i.e. KL holonomy is non-negative by construction (data-processing monotonicity of KL under channels [8, Ch. 2]).

3.3 Time reversal and entropy production for Markov chains

Let P be a stationary Markov chain on $\mathcal{X} = \{1, \dots, k\}$ with transition T and stationary π. Its time-reversal P^{rev} has transitions $T^*_{ji} = \frac{\pi_i T_{ij}}{\pi_j}$. For length n, let R_n denote reversal of the string. Consider the canonical loop

$$\textsf{Encode transitions} \;\to\; \textsf{Reverse} \;\to\; \textsf{Decode second},$$

which maps paths back to $\mathcal{X}^n$ (up to a boundary symbol). For Markov P, the pushforward law Q induced by this loop coincides with the path law of P^{rev} on cylinders.

Theorem 3.2 (KL holonomy rate equals entropy production)**.** *For the Markov setting above,*

$$\mathsf{d}(P\|P^{\mathrm{rev}}) \;=\; \sum_{i,j} \pi_i T_{ij}\, \log_2 \frac{\pi_i T_{ij}}{\pi_j T_{ji}} \;=\; \sigma \quad (\textit{bits/step}), \tag{6}$$

and the KL-holonomy satisfies

$$\frac{1}{n}\,\mathbb{E}_P\big[\mathrm{Hol}_n^{\mathrm{KL,time\text{-}rev}}\big] \;\to\; \sigma. \tag{7}$$

Proof. The path log-likelihood ratio between P and the reversed path law under P^{rev} is

$$\log \frac{P_n(X_{0:n-1})}{P_n^{\mathrm{rev}}(R_n(X_{0:n-1}))} = \sum_{t=1}^{n-1} \log \frac{\pi_{X_t}}{\pi_{X_{t-1}}} + \sum_{t=1}^{n-1} \log \frac{T_{X_{t-1}X_t}}{T_{X_t X_{t-1}}}.$$

The stationary term telescopes to $O(1)$; divide by n and take expectations. Identity (6) is standard in stochastic thermodynamics [9, 10]. Equation (7) is (5) with $Q = P^{\mathrm{rev}}$. □

Remark 3.3 (Why not merely $h(Q) - h(P)$?). For stationary Markov chains, $h(P) = h(P^{\mathrm{rev}})$, so representation-space holonomy would vanish. The KL version (observer-transported) returns the irreversible production σ.

3.4 General ergodic reversal

Let P^* be any stationary time-reversed process absolutely continuous w.r.t. P on cylinders, with finite $\mathsf{d}(P\|P^*)$. Then, by the same argument,

$$\frac{1}{n}\,\mathbb{E}_P\big[\mathrm{Hol}_n^{\mathrm{KL,time\text{-}rev}}\big] \;\to\; \mathsf{d}(P\|P^*). \tag{8}$$

4 Observer Independence

Theorem 4.1 (Code-robustness of KL holonomy)**.** *Let $\mathcal{E}^{(1)}$ and $\mathcal{E}^{(2)}$ be universal on $\mathcal{X}$ for the laws appearing in Lemma 2.1. Then, for any fixed loop γ,*

$$\frac{1}{n}\Big|\mathrm{Hol}^{\mathrm{KL},\gamma}_{n,\mathcal{E}^{(1)}}(X_{0:n-1}) - \mathrm{Hol}^{\mathrm{KL},\gamma}_{n,\mathcal{E}^{(2)}}(X_{0:n-1})\Big| \xrightarrow[n\to\infty]{P\text{-}a.s.} 0,$$

and likewise in $L^1(P)$.

Proof. Apply Lemma 2.1 to both codes and subtract. □

5 Numerical validation (UEC battery)

All experiments were run with the companion script `uec_battery.py`. The suite covers: (i) gauge invariance under bijective recoding; (ii) coarse-graining/refinement loops (non-negativity of KL holonomy); (iii) Markov time-reversal, where KL holonomy matches the analytic entropy production (EP) σ in bits/step; (iv) observer-independence trends (KT vs. LZ code lengths per symbol converge); (v) robustness sweeps (random chains, low/high EP regimes, alignment/segment stability), and (vi) bootstrap confidence intervals for windowed estimates.

AoT demos (audio / sensors / finance). For window-level arrow-of-time classification, two choices align AUC with holonomy and our theory: *loop-negatives* (Encode→Reverse→DecodeSecond) instead of literal reversal, and domain preprocessing (`--aot_diff` for audio/sensors, `--aot_logreturn` for finance). These match the time-reversal loop used by the holonomy and avoid negative-KL pathologies. The script logs per-file AUC and bits/step(/s) and writes a scoreboard CSV.

Artifacts and reproducibility. The script writes

- `results/aot_wav.json`, `results/aot_csv.json` (single-file AoT).
- `results/scoreboard.csv`, `results/scoreboard.json` (folder runs).
- `results/summary.json` (aggregated suite summary for the run).

Representative commands and flags for the AoT demos are documented inline in the repository (e.g., `--aot_bins`, `--aot_win`, `--aot_stride`, `--aot_rate`).

Discussion

We distinguished two operational regimes. If one evaluates evidence *in the representation reached by the loop*, holonomy reduces to the *entropy-rate difference* $h(Q) - h(P)$ (Theorem 2.2); this yields gauge invariance and detects net compression/expansion by the loop. If instead one *transports the observer* and evaluates evidence against the loop's pushforward law on the *original coordinates*, holonomy equals the *relative entropy rate* $\mathsf{d}(P\|Q)$, recovering irreversibility and, for Markov time reversal, the entropy production rate.

Technical extensions. The finite-alphabet assumption can be relaxed via quantization and standard approximation. The Markov time-reversal equality extends to hidden Markov models at the level of path measures; holonomy on observed records gives a certified lower bound by data processing (and in the quantum setting by Lindblad/Uhlmann monotonicity [15, 16]). Absolute continuity requirements ensure finite rates (e.g., $\sigma < \infty$ requires $T_{ij} > 0 \Rightarrow T_{ji} > 0$).

Acknowledgements

I used GPT-5 heavily in the creation of this document.

References

[1] J. Ziv and A. Lempel, "Compression of individual sequences via variable-rate coding," *IEEE Trans. Inf. Theory*, 24(5):530–536, 1978.

[2] R. E. Krichevsky and V. K. Trofimov, "The performance of universal encoding," *IEEE Trans. Inf. Theory*, 27(2):199–207, 1981.

[3] F. M. J. Willems, Y. M. Shtarkov, and T. J. Tjalkens, "The Context-Tree Weighting Method: Basic Properties," *IEEE Trans. Inf. Theory*, 41(3):653–664, 1995.

[4] J. Rissanen, "Modeling by shortest data description," *Automatica*, 14(5):465–471, 1978.

[5] A. R. Barron, "The strong ergodic theorem for densities: generalized Shannon–McMillan–Breiman," *Annals of Probability*, 13(4):1292–1303, 1985.

[6] P. C. Shields, *The Ergodic Theory of Discrete Sample Paths*, Graduate Studies in Mathematics, Vol. 13, American Mathematical Society, 1996.

[7] I. Csiszár and P. C. Shields, "Information theory and statistics: A tutorial," *Foundations and Trends in Communications and Information Theory*, 1(4):417–528, 2004.

[8] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed., Wiley, 2006.

[9] J. Schnakenberg, "Network theory of master equation," *Reviews of Modern Physics*, 48(4):571–585, 1976.

[10] U. Seifert, "Stochastic thermodynamics, fluctuation theorems and molecular machines," *Reports on Progress in Physics*, 75:126001, 2012.

[11] R. Kawai, J. M. R. Parrondo, and C. Van den Broeck, "Dissipation: The phase-space perspective," *Phys. Rev. Lett.*, 98:080602, 2007.

[12] G. E. Crooks, "Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences," *Phys. Rev. E*, 60:2721–2726, 1999. (See also *Phys. Rev. E*, 61:2361–2366, 2000.)

[13] T. Hatano and S.-I. Sasa, "Steady-state thermodynamics of Langevin systems," *Phys. Rev. Lett.*, 86:3463–3466, 2001.

[14] É. Roldán and J. M. R. Parrondo, "Estimating Dissipation from Single Stationary Trajectories," *Phys. Rev. Lett.*, 105:150607, 2010.

[15] G. Lindblad, "Completely positive maps and entropy inequalities," *Communications in Mathematical Physics*, 40:147–151, 1975.

[16] A. Uhlmann, "Relative entropy and the Wigner–Yanase–Dyson–Lieb concavity in an interpolation theory," *Communications in Mathematical Physics*, 54:21–32, 1977.