To Cast a Stone with Six Birds: A Closure-Deficit Account of Randomness under Packaging and Budget

Ioannis Tsiokos

March 9, 2026

Abstract

Randomness is often treated as either ontic indeterminacy or epistemic ignorance. We develop a third, operational account within Six Birds: layer-relative randomness is the predictive residue of non-closure under a chosen packaging and budget. For a micro process X_t and a packaging $\Pi : X \to Y$, we define the micro-closure deficit at scale τ by $\mathsf{CD}_\tau(\Pi) = I(X_t; Y_{t+\tau} \mid Y_t)$, the information about the next packaged state that remains hidden inside the current macro-object. This yields the exact decomposition $H(Y_{t+\tau} \mid Y_t) = H(Y_{t+\tau} \mid X_t) + \mathsf{CD}_\tau(\Pi)$, separating intrinsic substrate uncertainty from randomness introduced by discarded distinctions. We connect this exact quantity to computable diagnostics native to earlier Six Birds work, including route mismatch for coarse-grained Markov dynamics and predictive log-loss gaps under limited-memory models. In controlled Markov benchmarks, closure deficit vanishes on exactly lumpable partitions and rises with within-fiber heterogeneity, while route mismatch tracks it closely. In budgeted prediction, increasing memory order buys back hidden distinctions and lowers held-out log loss. In toy hashing, one-wayness and random-lookingness emerge as the same packaging-and-budget phenomenon: high-entropy inputs obey $q/2^n$ inversion scaling, while low-entropy inputs collapse the effect. The result is a measurable account of randomness that unifies coarse-graining, limited prediction, and feasible one-wayness without treating randomness as primitive.

1 Introduction

Randomness is often described in one of two ways. On one view, it is a primitive feature of the world: outcomes are genuinely indeterminate. On the other, it is a deficit on the side of the observer: the world is orderly, but the observer lacks the information or computational power required to predict it. Both descriptions leave out an operational fact that scientific theories cannot avoid. A theory never receives the substrate for free. It packages some distinctions into stable objects, ignores others, and then calls the predictive remainder noise, chance, or stochasticity. Without specifying the layer, the packaging, and the budget that maintains it, claims about randomness remain underdetermined.

Within the Six Birds program, closure is organized by six primitive operations: *P1 operator rewrite*, *P2 constraints*, *P3 protocol holonomy / route mismatch*, *P4 staging*, *P5 packaging*, and *P6 accounting* [Tsiokos, 2026c]. In the present paper, P5 and P6 do most of the work. Packaging decides which substrate distinctions become objects at the layer, while accounting decides which such distinctions can actually be afforded, stabilized, and audited. P4 sets the scale and timescale at which closure is attempted, P2 specifies which distinctions are admissible, P3 captures variability induced by untracked route dependence, and P1 describes the interventions required when closure fails to descend. Randomness is not a seventh primitive. It is the variability that remains when these closure mechanics do not fully stabilize the distinctions that matter for prediction at the chosen layer.

Automorph Inc., Wilmington, DE, USA ioannis@automorph.io DOI: 10.5281/zenodo.18926433 © 2026 CC-BY 4.0
Preprint — not peer reviewed.

This paper develops that claim in a form that can be computed and tested. Our thesis is that randomness is the predictive residue of non-closure. Let X_t denote a micro process and let $\Pi : X \to Y$ be a packaging map, with $Y_t = \Pi(X_t)$ the packaged state observed at the layer. We define the *micro-closure deficit* at scale τ by

$$\mathsf{CD}_\tau(\Pi) := I(X_t; Y_{t+\tau} \mid Y_t).$$

This quantity measures exactly how much information about the next packaged state is still hidden inside the current macro-object. It yields the decomposition

$$H(Y_{t+\tau} \mid Y_t) = H(Y_{t+\tau} \mid X_t) + \mathsf{CD}_\tau(\Pi),$$

which separates intrinsic substrate uncertainty from randomness introduced by discarded distinctions. A macrostate looks random exactly to the extent that it has failed to become a sufficient statistic for its own future. When the substrate is deterministic, the closure deficit is the whole story. When the substrate itself is stochastic, observable layer-relative randomness decomposes into intrinsic noise and non-closure.

This viewpoint extends and sharpens several earlier strands of the Six Birds sequence. *Six Birds: Foundations of Emergence Calculus* defined packaging as an idempotent form of closure and distinguished it from novelty and directionality [Tsiokos, 2026c]. *To Notch a Stone with Six Birds* operationalized predictive non-closure through an order-1 versus order-2 log-loss gap, using that gap to motivate theory birth and lens refinement [Tsiokos, 2026a]. *To Create a Stone with Six Birds* supplied an explicit coarse-graining pipeline for induced macro kernels and route-mismatch diagnostics on stochastic substrates [Tsiokos, 2026b]. If *To Notch a Stone* asked when failure of closure forces a richer theory to be born, the present paper asks what that failure looks like inside the theory that still tries to run. Our answer is: unresolved closure is experienced as randomness.

The paper also places this Six Birds account in contact with existing external traditions. The compression–prediction trade-off studied by the information bottleneck and related representation-learning programs already treats predictive quality as something purchased under a representational budget [Geiger et al., 2015, Tishby et al., 1999]. Computational mechanics similarly isolates predictive sufficiency as the criterion that separates structure from randomness in a process [Shalizi and Crutchfield, 2001]. Our contribution is different in emphasis. We treat predictive residue explicitly as a *closure deficit* of a packaging, connect it to concrete coarse-graining diagnostics already native to Six Birds, and then carry the same logic into engineered one-wayness. In particular, hashing will appear not as a source of mysterious randomness, but as an especially sharp case of many-to-one packaging under feasibility constraints, in the same spirit that the random-oracle tradition uses idealized random functions to reason about one-way behavior [Bellare and Rogaway, 1993].

Three controlled demonstrations drive the paper. First, on finite Markov chains with known partitions, closure deficit vanishes on exactly lumpable constructions and rises under controlled within-fiber heterogeneity, while route mismatch tracks it closely. This gives a direct bridge between the exact quantity introduced here and the computable coarse-graining diagnostics already used in Six Birds and in classical lumpability analyses of Markov chains [Buchholz, 1994, Kemeny and Snell, 1976, Tsiokos, 2026b]. Second, in budget-limited prediction, increasing memory order buys back hidden distinctions that the current packaged state failed to stabilize, producing a concrete randomness–budget curve. Third, in a toy-safe hashing setting, inversion success follows the familiar $q/2^n$ scaling for high-entropy inputs at small success probabilities, while low-entropy inputs collapse the effect entirely. One-wayness and random-lookingness then appear as the same packaging-and-budget phenomenon: not entropy creation, but feasible irreversibility relative to bounded observers.

Our claims are deliberately narrower than the familiar metaphysical debate between ontic and epistemic randomness. We do not claim that all randomness is reducible to hidden microstructure, nor do we identify route mismatch with genuine irreversibility. We claim something operationally sharper: once a substrate, a packaging, and a maintenance budget are fixed, the non-intrinsic part of layer-relative randomness is measurable as the information about the next packaged state that remains hidden inside the current one. That claim is exact, falsifiable, and portable across domains.

All experiments, figure-generation code, and supporting formal files for this manuscript are tracked in the companion repository at https://github.com/ioannist/six-birds-randomness.

The rest of the paper proceeds as follows. Section 2 briefly recalls the Six Birds setting needed here. Section 3 defines closure deficit and states the exact decomposition. Section 4 connects the exact quantity to route mismatch and budget-limited predictors. Sections 5.1 to 5.3 present the three controlled demonstrations. Section 6 discusses scope, limitations, and the distinction between randomness, novelty, and arrow-of-time; Section 7 closes.

2 Six Birds Recap and Problem Setup

Six Birds treats higher-level description not as a free summary of a substrate, but as a maintained closure. The full framework organizes this closure by six primitives: P1 operator rewrite, P2 constraints, P3 protocol holonomy / route mismatch, P4 staging, P5 packaging, and P6 accounting [Tsiokos, 2026c]. In the present paper, the main roles are played by P5, P4, and P6. Packaging chooses which distinctions become objects at the layer. Staging chooses the scale and timescale at which those objects are expected to persist. Accounting records the fact that keeping those objects predictive requires memory, sensing, or computational expenditure.

We work throughout in a finite-state setting, which is already rich enough to capture the closure problem studied here and matches the controlled experiments later in the paper. The substrate is a stationary process X_t on a microstate space $\mathcal{X}$. When we need explicit dynamics, we take X_t to be a Markov chain with transition kernel P and stationary distribution π. A layer is specified by a packaging map $\Pi : \mathcal{X} \to \mathcal{Y}$ together with a staging parameter $\tau \in \{1, 2, \dots\}$. The packaged variable is

$$Y_t := \Pi(X_t). \tag{2.1}$$

All information-theoretic quantities in this paper are computed under the stationary joint law of $(X_t, Y_t, Y_{t+\tau})$, and all logarithms are natural, so entropies are measured in nats.

The predictive question is then immediate. If two microstates $x, x' \in \mathcal{X}$ share the same current package $\Pi(x) = \Pi(x')$, do they induce the same distribution of future packaged outcomes at scale τ? If they do, then the current package is already sufficient for the next packaged step. If they do not, then the current macro-object hides a micro-index that still matters for prediction. The present paper identifies that hidden predictive residue with randomness at the layer.

This finite-state specialization aligns directly with the operational coarse-graining pipeline used in earlier Six Birds work [Tsiokos, 2026b]. It also makes the later budgeted analysis precise. Once Π and τ are fixed, one may either keep the full microstate X_t, keep only the packaged state Y_t, or pay for intermediate side information through richer predictors. Section 3 first isolates the exact information lost in going from X_t to Y_t; Section 4 will then show how that loss is approximated in practice by route mismatch and limited-memory predictive gaps.

3 Closure Deficit and the Exact Decomposition

For each microstate $x \in \mathcal{X}$, define the packaged τ-step future distribution

$$p_x^{(\tau)}(y') := \Pr(Y_{t+\tau} = y' \mid X_t = x), \qquad y' \in \mathcal{Y}. \tag{3.1}$$

For each packaged state $y \in \mathcal{Y}$, define the packaged future conditioned only on the current package by

$$\bar{p}_y^{(\tau)}(y') := \Pr(Y_{t+\tau} = y' \mid Y_t = y). \tag{3.2}$$

Under the stationary micro law π, the latter is the fiber-average of the former:

$$\bar{p}_y^{(\tau)}(\cdot) = \sum_{x \in \Pi^{-1}(y)} \pi(x \mid y)\, p_x^{(\tau)}(\cdot). \tag{3.3}$$

Here

$$\pi(x \mid y) := \frac{\pi(x)}{\pi_Y(y)}, \qquad \pi_Y(y) := \sum_{x:\Pi(x)=y} \pi(x),$$

whenever $\pi_Y(y) > 0$; packaged states with $\pi_Y(y) = 0$ do not contribute under the stationary law. The issue of closure is exactly whether conditioning on Y_t alone loses predictive information relative to conditioning on X_t.

Definition 3.1 (Micro-closure deficit)**.** *Given a packaging $\Pi : \mathcal{X} \to \mathcal{Y}$ and a staging scale τ, the* micro-closure deficit *is*

$$\mathsf{CD}_\tau(\Pi) := I(X_t; Y_{t+\tau} \mid Y_t). \tag{3.4}$$

The conditioning on Y_t is essential. The quantity $\mathsf{CD}_\tau(\Pi)$ is not the absolute dependence of the next packaged state on the current microstate. It is the *extra* predictive information left in X_t once the current packaged state is already known. In other words, it measures exactly how much of the next packaged outcome remains hidden inside the current macro-object.

Proposition 3.2 (Exact decomposition)**.** *Under the stationary law of $(X_t, Y_t, Y_{t+\tau})$,*

$$H(Y_{t+\tau} \mid Y_t) = H(Y_{t+\tau} \mid X_t) + \mathsf{CD}_\tau(\Pi). \tag{3.5}$$

Proof. By the chain-rule identity for conditional mutual information,

$$I(X_t; Y_{t+\tau} \mid Y_t) = H(Y_{t+\tau} \mid Y_t) - H(Y_{t+\tau} \mid X_t, Y_t).$$

Since $Y_t = \Pi(X_t)$ is a deterministic function of X_t, conditioning on Y_t adds nothing once X_t is known, so

$$H(Y_{t+\tau} \mid X_t, Y_t) = H(Y_{t+\tau} \mid X_t).$$

Substituting into the first display gives (3.5). □

Figure 1 summarizes the decomposition schematically.

The two terms in (3.5) have different meanings. The term $H(Y_{t+\tau} \mid X_t)$ measures what remains unpredictable about the future packaged outcome even with exact access to the current microstate. It captures genuine substrate uncertainty propagated to the packaged level. The term $\mathsf{CD}_\tau(\Pi)$ measures something else: the unpredictability created by refusing or failing to retain micro distinctions that still matter once the current package is fixed. In deterministic substrates the intrinsic term vanishes, so all packaged randomness comes from non-closure.

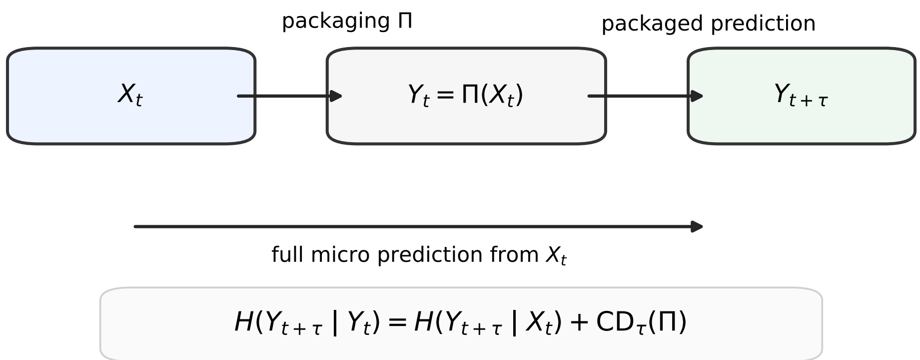

Figure 1: Schematic of the closure-deficit decomposition. Access to the current microstate X_t leaves only intrinsic packaged uncertainty $H(Y_{t+\tau} \mid X_t)$, while prediction from the current packaged state $Y_t = \Pi(X_t)$ incurs the additional deficit $\mathsf{CD}_\tau(\Pi)$ when the package is not yet sufficient for its own future.

Proposition 3.3 (Expected KL form)**.** *The micro-closure deficit can be written as*

$$\mathsf{CD}_\tau(\Pi) = \sum_{x\in\mathcal{X}} \pi(x)\, D_{\mathrm{KL}}(p_x^{(\tau)} \,\|\, \bar{p}_{\Pi(x)}^{(\tau)}), \tag{3.6}$$

where, in the finite case,

$$D_{\mathrm{KL}}(p\|q) := \sum_{y'\in\mathcal{Y}} p(y') \log \frac{p(y')}{q(y')},$$

with the convention $0 \log(0/q) = 0$*.*

Proof. Conditional mutual information is the expectation of the Kullback–Leibler divergence between the conditional law of $Y_{t+\tau}$ given (X_t, Y_t) and the conditional law of $Y_{t+\tau}$ given Y_t alone. Because Y_t is determined by X_t, the first law is exactly $p_x^{(\tau)}$ and the second is $\bar{p}_{\Pi(x)}^{(\tau)}$. Averaging these divergences under the stationary law π yields (3.6). □

Definition 3.4 (τ-closed packaging)**.** *A packaging* Π *is* τ-closed *if for every* $x, x' \in \mathcal{X}$ *with* $\Pi(x) = \Pi(x')$*, one has*

$$p_x^{(\tau)} = p_{x'}^{(\tau)} \quad \text{as distributions on } \mathcal{Y}. \tag{3.7}$$

Proposition 3.5 (Closure, sufficiency, and vanishing deficit)**.** *If* Π *is* τ*-closed, then* $\mathsf{CD}_\tau(\Pi) = 0$*. Conversely, if* $\pi(x) > 0$ *for every* $x \in \mathcal{X}$ *and* $\mathsf{CD}_\tau(\Pi) = 0$*, then* Π *is* τ*-closed.*

Proof. If Π is τ-closed, then all microstates inside the same fiber have the same packaged future distribution, so for every x one has $p_x^{(\tau)} = \bar{p}_{\Pi(x)}^{(\tau)}$. Every divergence in (3.6) is therefore zero, and thus $\mathsf{CD}_\tau(\Pi) = 0$.

Conversely, assume $\pi(x) > 0$ for all x and $\mathsf{CD}_\tau(\Pi) = 0$. By (3.6), a weighted sum of nonnegative KL divergences is zero. Since every weight $\pi(x)$ is positive, each divergence must itself be zero. Hence $p_x^{(\tau)} = \bar{p}_{\Pi(x)}^{(\tau)}$ for every x. If $\Pi(x) = \Pi(x')$, then both $p_x^{(\tau)}$ and $p_{x'}^{(\tau)}$ equal the same $\bar{p}_{\Pi(x)}^{(\tau)}$, so $p_x^{(\tau)} = p_{x'}^{(\tau)}$. Thus Π is τ-closed. □

Table 1: The exact quantity used in this paper and the main diagnostics or proxies used to approximate it in practice.

Quantity	Formal role	Operational reading
$\mathsf{CD}_\tau(\Pi)$	exact hidden predictive information after conditioning on Y_t	target object
$\mathrm{RM}_\tau^{\mathrm{lift}}(\Pi)$	stationary-weighted distributional mismatch between micro futures and induced macro rows	computable coarse-graining diagnostic
$\Delta_{\mathrm{pred}}^{(2,1)}$	extra held-out gain from one additional packaged-history step	macro-history proxy when X_t is unavailable
$\mathsf{Rand}_B(\Pi,\tau)$	best achievable predictive log loss under feasible class $\mathcal{Q}_B$	randomness–budget curve

The last proposition is the formal version of the paper's slogan. A packaged state looks random exactly to the extent that it has failed to become a sufficient statistic for its own packaged future. In finite Markov settings, τ-closedness at $\tau = 1$ is the classical lumpability condition stated in terms of packaged destination distributions; for $\tau > 1$ it is the same condition after staging by τ [Buchholz, 1994, Kemeny and Snell, 1976]. This is why closure deficit is the right exact object for the rest of the paper: it vanishes precisely when the current package already carries everything the layer needs for one staged step of prediction, and it becomes strictly positive exactly when hidden within-fiber distinctions remain unpaid for.

4 From Exact Quantity to Computable Diagnostics

The quantity $\mathsf{CD}_\tau(\Pi)$ is exact, but one rarely has direct access to the full stationary law of $(X_t, Y_t, Y_{t+\tau})$. In practice, the exact object must be approached through observables that live at different levels of accessibility. This section introduces three such bridges. The first is a route-mismatch diagnostic for coarse-grained Markov dynamics. The second is a predictive-gap proxy based on richer packaged history when the microstate itself is unavailable. The third is a budgeted log-loss formulation that turns randomness into a resource-relative curve. Table 1 summarizes the relation between the exact quantity and the main operational surrogates used in the paper.

Suppose an operational coarse-graining procedure supplies, for each current packaged state $y \in \mathcal{Y}$, an induced packaged row $\widehat{p}_y^{(\tau,\mathrm{lift})}$ intended to summarize the τ-step packaged future from that macrostate. Following the induced-kernel pipeline of Tsiokos, such rows are obtained by choosing a lift on each fiber and then evolving for τ steps before pushing forward again [Tsiokos, 2026b]. The associated stationary-weighted route mismatch is

$$\mathrm{RM}_\tau^{\mathrm{lift}}(\Pi) := \sum_{x\in\mathcal{X}} \pi(x) \left\| p_x^{(\tau)} - \widehat{p}_{\Pi(x)}^{(\tau,\mathrm{lift})} \right\|_1. \tag{4.1}$$

When the lift is stationary-conditional, one has $\widehat{p}_y^{(\tau,\mathrm{lift})} = \bar{p}_y^{(\tau)}$ for every y, and we write the resulting quantity as $\mathrm{RM}_\tau^{\mathrm{stat}}(\Pi)$.

Proposition 4.1 (Route mismatch lower bound)**.** *If the induced packaged rows coincide with the stationary-conditioned rows, then*

$$\mathsf{CD}_\tau(\Pi) \geq \frac{1}{2}\left(\mathrm{RM}_\tau^{\mathrm{stat}}(\Pi)\right)^2. \tag{4.2}$$

Proof. By Proposition 3.3,

$$\mathsf{CD}_\tau(\Pi) = \sum_{x\in\mathcal{X}} \pi(x)\, D_{\mathrm{KL}}(p_x^{(\tau)} \,\|\, \bar{p}_{\Pi(x)}^{(\tau)}).$$

Pinsker's inequality in natural units gives [Cover and Thomas, 2006]

$$D_{\mathrm{KL}}(p\|q) \geq \frac{1}{2}\|p-q\|_1^2.$$

Therefore

$$\mathsf{CD}_\tau(\Pi) \geq \frac{1}{2} \sum_{x\in\mathcal{X}} \pi(x)\, \|p_x^{(\tau)} - \bar{p}_{\Pi(x)}^{(\tau)}\|_1^2.$$

Applying Jensen's inequality to the convex map $u \mapsto u^2$ yields

$$\sum_{x\in\mathcal{X}} \pi(x)\, \|p_x^{(\tau)} - \bar{p}_{\Pi(x)}^{(\tau)}\|_1^2 \geq \left(\sum_{x\in\mathcal{X}} \pi(x)\, \|p_x^{(\tau)} - \bar{p}_{\Pi(x)}^{(\tau)}\|_1\right)^2,$$

which is exactly $\left(\mathrm{RM}_\tau^{\mathrm{stat}}(\Pi)\right)^2$. □

Thus the exact deficit controls any stationary-conditioned route-mismatch diagnostic. Operational route mismatch need not use the stationary-conditional lift, however. In particular, *To Create a Stone with Six Birds* uses induced macro kernels derived from explicit lift choices, including uniform-in-fiber lifts, as a practical audit of how badly macro rows fail to summarize the micro futures they stand for [Tsiokos, 2026b]. In the experiments below we report both a stationary version, which targets $\bar{p}_y^{(\tau)}$ directly, and a uniform version, which measures non-closure relative to a cheaper operational packaging choice.

When the microstate is inaccessible but a packaged sequence $y_{1:T}$ is observed, one can still ask whether the current package is sufficient for one more packaged step. Let NLL_1 and NLL_2 denote held-out average negative log-likelihood under first- and second-order Markov predictors on the packaged sequence. Define the predictive gap

$$\Delta_{\mathrm{pred}}^{(2,1)} := \mathrm{NLL}_1 - \mathrm{NLL}_2\,. \tag{4.3}$$

In the infinite-data idealization, with exact predictors of the stated orders on the same stationary source, this gap becomes

$$\Delta_{\mathrm{pred}}^{(2,1)} = I(Y_{t-1}; Y_{t+1} \mid Y_t).$$

A positive gap therefore indicates that one additional unit of packaged history still carries predictive information not stabilized into the current package. This is the predictive non-closure signal used operationally in *To Notch a Stone with Six Birds* [Tsiokos, 2026a]. It is not identical to the micro-closure deficit: it probes hidden predictive residue inside macro history rather than inside the current microstate. But when omitted micro distinctions persist through the packaged history, the predictive gap becomes a practical proxy for the same underlying failure of closure.

This viewpoint leads naturally to a budgeted notion of randomness. Let $\mathcal{Q}_B$ be a feasible predictor class available under budget B, and let $S_t^{(B)}$ denote the state summary accessible to that class. Define

$$\mathsf{Rand}_B(\Pi,\tau) := \inf_{Q\in\mathcal{Q}_B} \mathbb{E}\Big[-\log Q(Y_{t+\tau} \mid S_t^{(B)})\Big]\,. \tag{4.4}$$

If $\mathcal{Q}_B$ contains all conditional distributions on the same accessible state, then $\mathsf{Rand}_B(\Pi,\tau) = H(Y_{t+\tau} \mid S_t^{(B)})$. As the budget grows and the feasible class expands, $\mathsf{Rand}_B(\Pi,\tau)$ can only decrease.

In the simplest case studied later, $B = L$ is packaged memory order and $S_t^{(L)} = (Y_{t-L+1}, \dots, Y_t)$. The resulting randomness–budget curve is then the held-out log loss of the best order-L predictor. Extra memory lowers that curve exactly when the current package has failed to absorb distinctions that still matter for its future. At the top end of this scale, access to the exact microstate would reduce the best achievable loss to $H(Y_{t+\tau} \mid X_t)$, leaving only intrinsic substrate uncertainty.

Three distinctions are important before turning to data. First, closure deficit is not novelty. A layer can have $\mathsf{CD}_\tau(\Pi) > 0$ without forcing a theory extension or a new ontology. Second, closure deficit is not irreversibility. Arrow-of-time audits compare forward and reversed path measures and obey different monotonicity constraints under coarse-graining, whereas route mismatch and predictive gap only register unresolved predictive dependence [Esposito, 2012, Tsiokos, 2026c]. Hidden protocol phases or omitted state variables can therefore make a layer look random without certifying any genuine nonequilibrium drive. Third, route mismatch and predictive gap are diagnostics, not replacements for the exact object. They are useful precisely because $\mathsf{CD}_\tau(\Pi)$ is the right target and these observables track different operational shadows of it.

With these bridges in place, the remaining sections can move back and forth between the exact quantity and the observables without changing the underlying claim. Section 5.1 uses route mismatch to track known closure deficits in controlled Markov families. Section 5.2 measures Rand_B directly through held-out log loss under increasing memory order. Section 5.3 shows the same packaging-and-budget logic in a deterministic many-to-one map.

5 Controlled Demonstrations

5.1 Markov Coarse-Graining Benchmark

We begin with the controlled finite-state benchmark in which the exact object $\mathsf{CD}_\tau(\Pi)$ is available by construction. The benchmark consists of four families: exactly lumpable chains [Buchholz, 1994, Kemeny and Snell, 1976], perturbed-lumpable chains with heterogeneity parameter $\alpha \in \{0, 0.2, 0.5\}$, metastable block chains with escape parameter $p_{\mathrm{out}} \in \{0.02, 0.08, 0.20\}$, and hidden-type chains with within-fiber type strength in $\{0.1, 0.5, 0.9\}$. Across these 17 staged conditions we compute the exact closure deficit, the intrinsic term, the decomposition residual, and both uniform- and stationary-lift route mismatch.

The benchmark behaves exactly as Proposition 3.5 predicts on the closed cases. The exactly lumpable constructions collapse to machine zero at both staging depths: at $\tau = 1$, for example, the benchmark gives $\mathsf{CD}_1(\Pi) = 4.34 \times 10^{-17}$ and $\mathrm{RM}_1^{\mathrm{uniform}}(\Pi) = 5.29 \times 10^{-17}$. The decomposition identity is numerically stable throughout the sweep, with maximum absolute residual 2.87×10^{-16} over all rows. In this sense the benchmark serves as a calibration device: the exact object vanishes when it should vanish, and the numerical implementation does not introduce visible leakage.

Once within-fiber predictive heterogeneity is introduced, the deficit becomes strictly positive and the cheaper uniform-lift mismatch already tracks it closely. Over all benchmark rows, the Pearson correlation between $\mathrm{RM}^{\mathrm{uniform}}$ and CD is $r = 0.959$. The most extreme perturbed-lumpable condition, $\alpha = 0.5$ at $\tau = 1$, reaches $\mathsf{CD} = 0.0482$ with $\mathrm{RM}^{\mathrm{uniform}} = 0.2757$. The strongest hidden-type condition reaches $\mathsf{CD} = 0.0310$ with $\mathrm{RM}^{\mathrm{uniform}} = 0.2125$. By contrast, the metastable family produces smaller but still systematic non-closure, topping out at $\mathsf{CD} = 7.22 \times 10^{-4}$ and $\mathrm{RM}^{\mathrm{uniform}} = 2.30 \times 10^{-2}$ when $p_{\mathrm{out}} = 0.20$. Figure 2 shows the aggregate pattern. Although Proposition 4.1 is stated for the stationary-conditioned route mismatch, the cheaper operational mismatch already serves as a practical proxy for non-closure in this controlled setting.

The perturbed-lumpable family isolates the cleanest knob, because only the within-fiber heterogeneity parameter α is varied. At $\tau = 1$, raising α from 0.2 to 0.5 increases the exact closure deficit

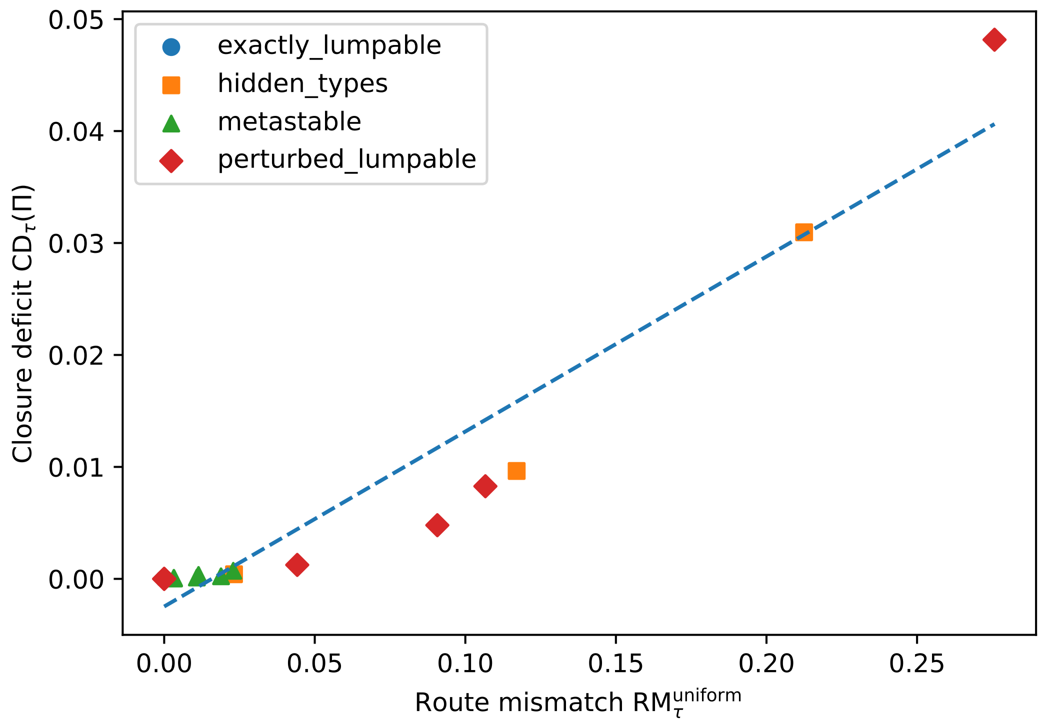

Figure 2: Uniform-lift route mismatch against exact closure deficit across the controlled Markov benchmark. Exactly lumpable cases remain at numerical zero, while controlled within-fiber heterogeneity and hidden types move points away from the origin in tandem.

from 0.00827 to 0.0482. At $\tau = 2$, the same perturbations remain visible but contract to 0.00125 and 0.00479, respectively. Figure 3 makes the staging effect explicit: longer staging partially absorbs the predictive consequences of hidden micro-variation, but it does not erase them. This is precisely the P4 point in Six Birds terms. Randomness is not fixed independently of scale; it is what remains unresolved at the scale and timescale the layer chooses to run.

The controlled benchmark therefore establishes the key empirical claim of the paper in the cleanest possible setting. Closure deficit vanishes on closed partitions, rises under controlled violations, and is tracked well by route-mismatch diagnostics that are much cheaper to compute than the exact object. The next subsection turns from controlled coarse-graining to the complementary question of what it costs to buy those hidden distinctions back once the current package has failed to stabilize them.

5.2 Budgeted Prediction

The controlled Markov benchmark established what unresolved closure is. The complementary question is what it costs to buy the missing distinctions back once the current package has already failed to stabilize them. To expose that accounting side directly, we fix the strongest hidden-type benchmark and measure the budgeted randomness curve Rand_B empirically using packaged predictors of increasing memory order.

The empirical curve is reported in the feasible-class sense of (4.4): at budget $B = L$, we plot the best held-out log loss available among predictors of order at most L. This matters because sparse higher-order contexts can make the raw loss of a particular order-L fit fluctuate upward in finite

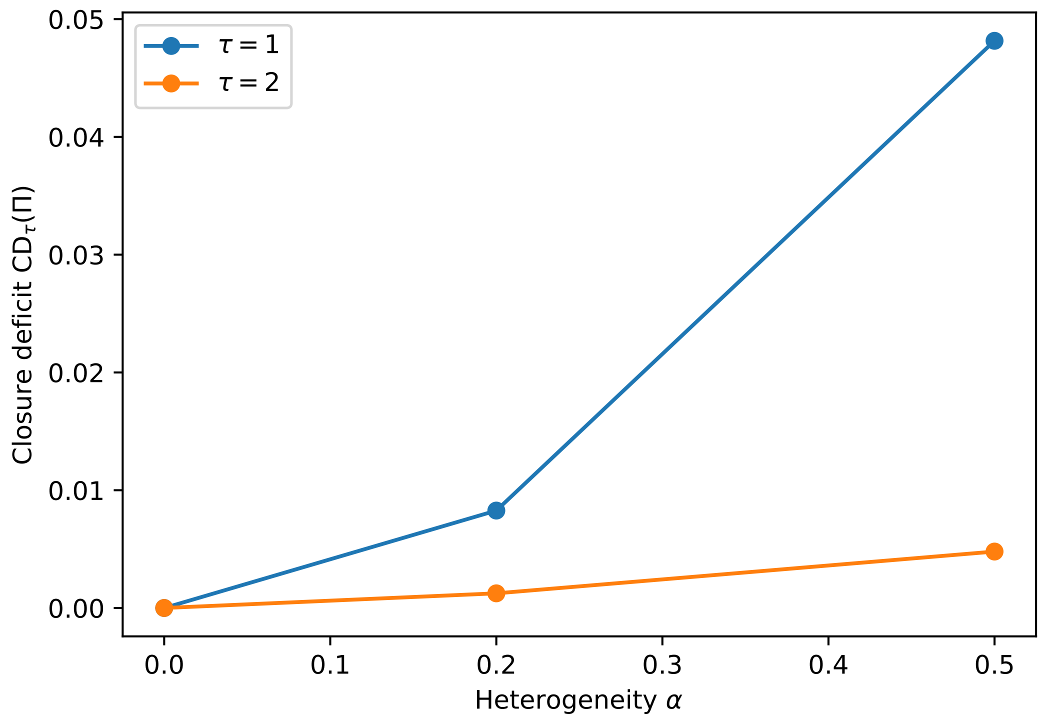

Figure 3: Exact closure deficit in the perturbed-lumpable family as a function of within-fiber heterogeneity α. Increasing heterogeneity raises the deficit, while larger staging depth τ reduces but does not eliminate it.

data, whereas the feasible-class quantity itself should never worsen when the admissible predictor class strictly expands.

In the frozen run, the order-0 baseline is 1.09565 nats. Allowing one packaged step of memory lowers the held-out loss to 1.04394, a gain of 5.17×10^{-2} nats. Allowing two packaged steps lowers it again to 1.03855, buying back an additional 5.38×10^{-3} nats. Beyond that point the feasible-class curve saturates through orders 3, 4, and 5, indicating that the hidden predictive residue visible to this sequence has been exhausted at the present sample size. The exact packaged first-order benchmark for the same process is $H(Y_{t+1} \mid Y_t) = 1.04953$, so the order-1 held-out estimate sits within 5.60×10^{-3} nats of the theoretical line, while the best order-≤ 2 predictor improves on that baseline by 1.10×10^{-2} nats. Figure 4 shows the resulting randomness–budget curve.

The point of this experiment is not that higher order is always better. The point is that randomness can be priced. Extra memory lowers predictive loss exactly when the current package has failed to absorb distinctions that still matter for its future. In Six Birds language, this is the P6 side of the story: what looks like chance at one budget becomes predictable at another because the layer has paid to carry more of its unresolved closure forward.

5.3 Hashing as Packaging and One-Wayness

Hashing provides an engineered version of the same logic in a deterministic many-to-one map. A truncated digest $H_n : \{0,1\}^m \to \{0,1\}^n$ does not create new entropy; it packages a large input space into a short stable record and leaves preimage identification to whatever budget the observer can afford [Rogaway and Shrimpton, 2004]. We therefore examine three input regimes on small safe truncations: a uniform high-entropy source, a low-entropy synthetic dictionary of size 256, and a fifty-fifty medium mixture of dictionary and out-of-dictionary inputs.

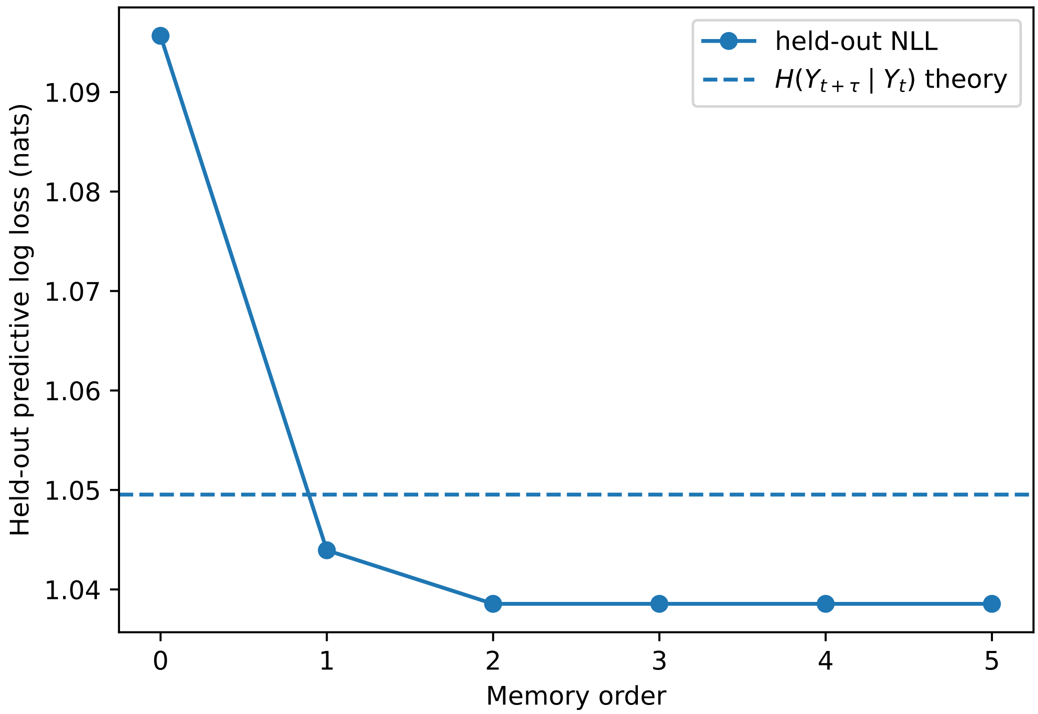

Figure 4: Held-out predictive log loss versus packaged memory order in the hidden-type benchmark. The horizontal line marks the exact packaged first-order uncertainty $H(Y_{t+1} \mid Y_t)$. The reported curve is the feasible-class envelope up to order L, so saturation indicates that additional packaged memory no longer buys back detectable predictive distinctions at this sample size.

In the uniform regime, the familiar one-wayness heuristic appears exactly where it should: in the small-success region $q/2^n \leq 0.25$, empirical inversion success tracks the line $q/2^n$ with mean absolute deviation 3.70×10^{-3}. Figure 5 plots empirical success against $q/2^n$ directly. For example, at $n = 16$ and budget $q = 256$, the empirical success rate is 6.67×10^{-3} against the simple baseline $256/2^{16} = 3.91 \times 10^{-3}$. In this regime, the digest is effectively random-looking relative to the bounded inversion budget, even though the map is fully deterministic.

That behavior collapses once the input distribution is small. At the same truncation $n = 16$ and the same budget $q = 256$, low-entropy dictionary inputs invert with probability 1.00, while the medium mixture already reaches 0.463. The difference is not in the hash function but in how much preimage uncertainty the input source actually contains. Hashing does not manufacture entropy; it can only hide distinctions that were present to begin with.

Simple random-lookingness tests tell the same story. Using 3000 sampled digests and 16-bit truncation for the collision audit, the uniform regime has collision ratio 0.816 relative to the ideal random-output expectation and is not grossly flagged by simple byte- or bit-level tests. The medium mixture and low-entropy regimes, by contrast, jump to collision ratios 66.24 and 255.30, with distinct-value fractions 0.569 and 0.0853, respectively. Figure 6 shows the collision-ratio separation on a log scale. A bounded observer therefore distinguishes these regimes without sophisticated cryptanalysis: high-entropy inputs look random enough, low-entropy inputs do not, and medium mixtures fall in between.

The hashing experiment is deliberately toy-sized, but its conceptual role is exact. One-wayness and random-lookingness do not require a primitive source of randomness. They arise when a deterministic packaging map is observed through a bounded budget and supplied with enough input entropy

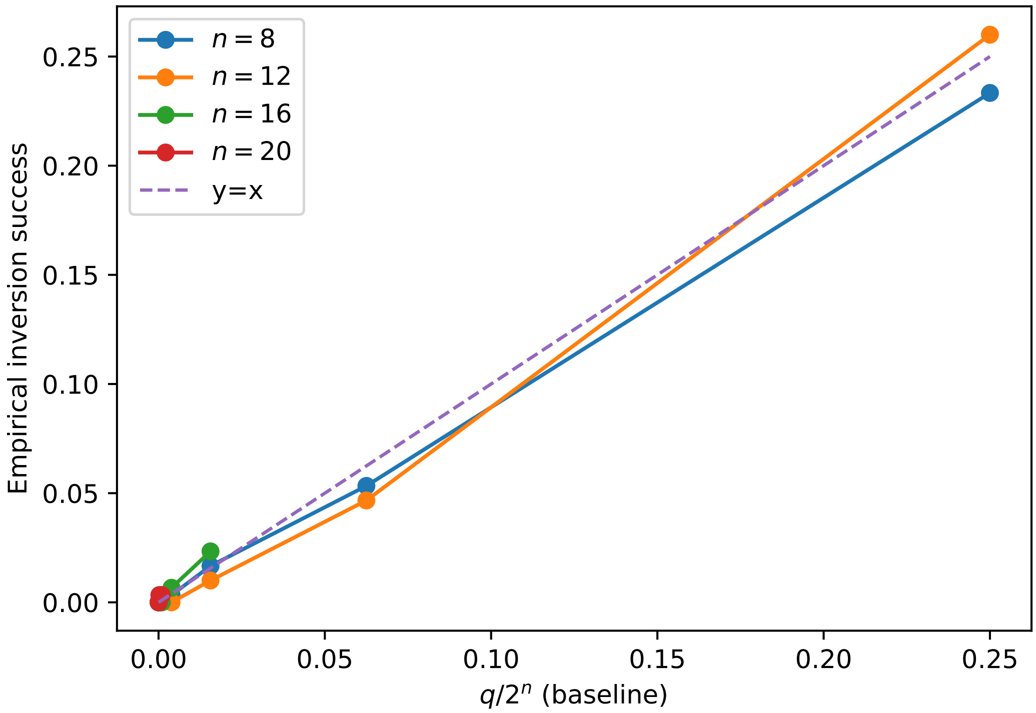

Figure 5: Uniform-input inversion success in the low-success regime plotted against the baseline $q/2^n$. Across truncation levels, empirical success remains close to the diagonal, matching the expected feasibility-limited scaling of one-way search.

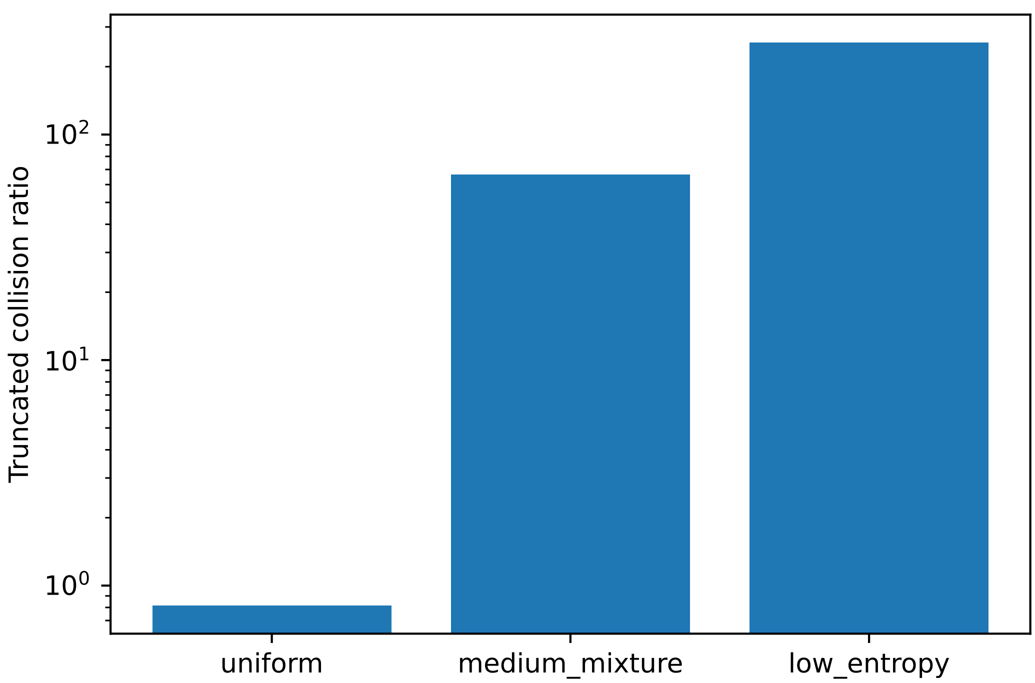

Figure 6: Collision-ratio separation for 16-bit truncated digests under three input regimes. The vertical axis is logarithmic. High-entropy uniform inputs remain near the ideal random-output expectation, while medium-mixture and low-entropy inputs become trivially distinguishable.

to make unpaid distinctions expensive to reconstruct. This is the same closure-and-accounting mechanism seen earlier in coarse-grained dynamics, now displayed in an engineered many-to-one map.

6 Discussion and Scope

The paper's claim can now be stated compactly. Once a substrate, a packaging, and a staging scale are fixed, observable packaged randomness decomposes into two parts: intrinsic substrate uncertainty and closure deficit. The first term is what remains unpredictable even with exact access to the current microstate. The second is the information about the next packaged state that remains hidden inside the current one because the layer has not stabilized all the distinctions that still matter for prediction. In the controlled Markov benchmark, this deficit vanishes on closed partitions and rises under controlled violations; in budgeted prediction, it is partially bought back by extra memory; and in hashing, it appears as feasibility-limited irreversibility rather than entropy creation. Across these domains, the same explanatory pattern repeats: what a layer calls noise is often unresolved predictive structure.

One advantage of this account is that it separates questions that are often blurred together. Closure deficit is not novelty. A layer may have $\mathsf{CD}_\tau(\Pi) > 0$ and still admit a perfectly serviceable ontology; the deficit only says that the current package is not sufficient for one staged step of prediction. Nor is closure deficit the same as genuine irreversibility. Path-space asymmetry and entropy-production audits obey different monotonicity constraints under coarse observation, whereas route mismatch and predictive gaps only register unresolved dependence [Esposito, 2012, Tsiokos, 2026c]. Hidden protocol phases, omitted variables, or cheap lifts can therefore make a layer look random without certifying any genuine nonequilibrium drive. The present paper is about the predictive residue of non-closure, not about novelty or arrows of time.

A second advantage is that the account makes randomness explicitly budget-relative without reducing it to a vague appeal to ignorance. Budget matters because records and distinctions are costly to maintain. The curve in Figure 4 is therefore not merely a machine-learning performance plot. It is a ledger: increasing memory order lowers the best achievable loss only when the current package has failed to absorb distinctions that still matter for its future. In that sense, randomness is what prediction looks like when closure has been budgeted too cheaply.

The hashing demonstration sharpens the same lesson in an engineered setting. A one-way hash is a deliberately extreme packaging: huge preimage fibers are compressed into a short stable record. But the experiment shows that the appearance of randomness depends as much on the source as on the map. High-entropy inputs produce the familiar small-success $q/2^n$ inversion law and pass simple random-lookingness checks; low-entropy inputs collapse both effects immediately. One-wayness is thus not metaphysical irreversibility. It is packaging observed through a bounded budget, supplied with enough entropy that unpaid distinctions remain expensive to reconstruct.

The auxiliary clustering experiment in Appendix A is useful precisely because it does *not* support a simplistic monotonic story. Increasing the number of learned clusters there does not reduce the empirical closure deficit; in the frozen run it increases from 0.0782 at $k = 2$ to 0.1579 at $k = 8$, while held-out packaged prediction worsens at the same time. The lesson is not that the account fails, but that representational budget is not the same as good packaging. More bins do not help when the learned partition is misaligned with the dynamics. Six Birds does not say that refinement is always better; it says that refinement only pays when it closes the right distinctions at the right staging scale.

Several limitations should be kept explicit. The exact treatment here is finite-state and stationary. That is sufficient for the controlled claims we make, but continuous-state or strongly nonstationary settings will require estimators and approximation schemes rather than exact sums. The route-mismatch bound in Proposition 4.1 is stated for the stationary-conditioned lift, while many operational pipelines use cheaper lifts for practical reasons. The predictive-gap proxy likewise detects unresolved dependence in packaged history, not literally the micro-closure deficit. These are not objections to the framework; they are reminders that diagnostics should be interpreted as shadows of the exact object rather than substitutes for it.

What the paper does provide is a portable operational answer to a question that is usually posed metaphysically. Randomness need not be treated as a primitive ingredient of explanation at the layer where it is observed. Once the substrate, packaging, staging, and budget are specified, the non-intrinsic part of that randomness is measurable as the information about the next packaged state that remains hidden inside the current one. You usually do not remove such randomness outright; you relocate it, price it, or decide not to pay for it.

7 Conclusion

We have argued that a layer looks random exactly to the extent that its current packaged state has failed to become a sufficient statistic for its own packaged future. The exact quantity capturing that failure is the micro-closure deficit

$$\mathsf{CD}_\tau(\Pi) = I(X_t; Y_{t+\tau} \mid Y_t),$$

which yields the decomposition

$$H(Y_{t+\tau} \mid Y_t) = H(Y_{t+\tau} \mid X_t) + \mathsf{CD}_\tau(\Pi).$$

This separates intrinsic substrate uncertainty from randomness introduced by discarded distinctions. The controlled Markov benchmark showed that the deficit vanishes on closed partitions and rises under controlled violations, while route mismatch tracks it closely. Budget-limited prediction showed that extra memory buys back hidden distinctions and lowers predictive loss. Toy hashing showed the same mechanism in a deterministic many-to-one map: one-wayness and random-lookingness emerge under bounded budgets and sufficient input entropy, not from entropy creation by the hash itself.

Within Six Birds, this gives randomness a precise place. It is not a new primitive alongside P1–P6, and it is not interchangeable with novelty or irreversibility. It is the predictive residue of non-closure under a chosen packaging, staging scale, and maintenance budget. That residue can be measured exactly in controlled settings and approximated operationally when only coarser observables are available.

The broader consequence is methodological. Debates about whether randomness is intrinsic or merely epistemic often collapse because the relevant layer has not been specified. Once the layer is specified, much of the question becomes concrete: which distinctions were stabilized into objects, which were left unresolved, and what would it cost to recover them? That is the sense in which randomness becomes an accounting problem. It is what remains after a theory has decided what to make real and what not to pay for.

A Auxiliary Clustering Result

This appendix records the representation-learning result that we treat as auxiliary rather than core evidence [Geiger et al., 2015, Tishby et al., 1999]. The experiment asks whether increasing representational budget, measured by the number of learned clusters k, automatically reduces predictive

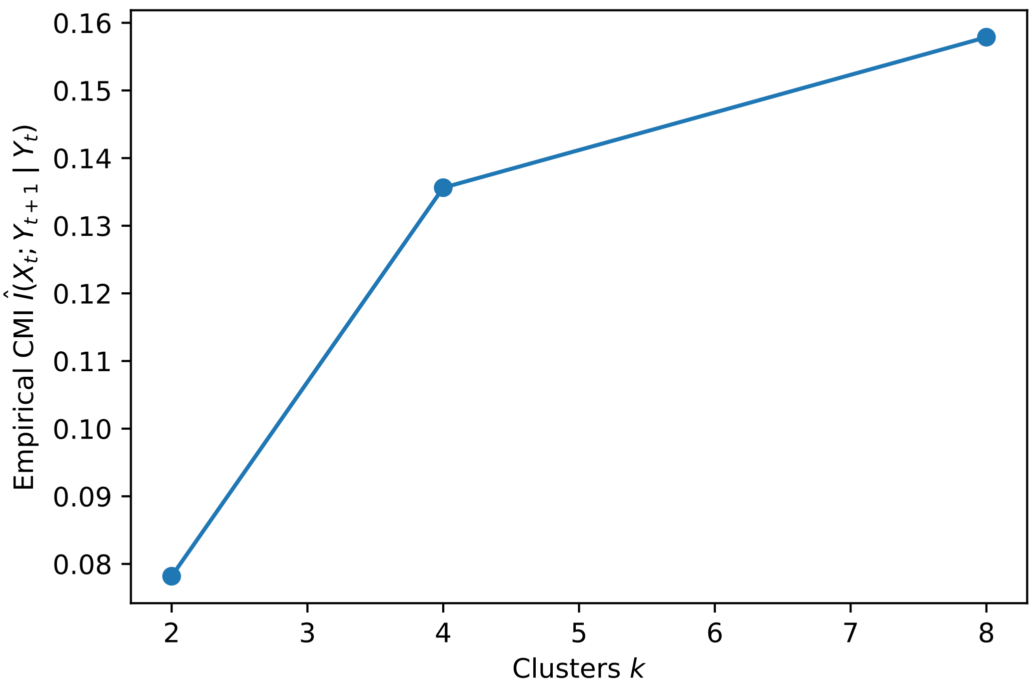

Figure 7: Auxiliary clustering result: raw empirical closure deficit versus learned cluster count k. Increasing representational granularity does not reduce non-closure in this run; it increases it.

non-closure when packaging is learned from observations rather than supplied by construction. It does not.

We generate observation sequences from the hidden-type process used earlier and learn k-means partitions on those observations for $k \in \{2, 4, 8\}$ [MacQueen, 1967]. For each learned packaging we then evaluate the resulting packaged sequence by held-out first- and second-order predictive log loss and by the empirical conditional mutual information

$$\widehat{I}(X_t; Y_{t+1} \mid Y_t),$$

computed from the joint sequence of hidden microstates and learned packaged labels. In this appendix we report the *raw* empirical conditional mutual information, not the normalized version used only for diagnostics during development.

The resulting trend is a useful caution. As k increases from 2 to 8, the empirical closure deficit rises from 0.0782 to 0.1356 and then to 0.1579 rather than falling. Held-out first-order predictive loss rises in the same direction, from 0.626 to 1.307 and then to 2.005. Second-order predictive loss likewise worsens, from 0.628 to 1.318 and then to 2.837. Figures 7 and 8 summarize the pattern.

The lesson is not that the closure-deficit account fails outside controlled partitions. The lesson is that more representational bins are not automatically the same thing as better packaging. A learned partition only helps if it closes distinctions that matter for the staged dynamics. If the partition instead slices the observations in a way that is misaligned with the predictive structure, additional representational budget can make the packaged process harder rather than easier to predict. This is why the result is auxiliary: it is informative about misaligned packaging, but it is not needed for the paper's main evidentiary arc.

For the same reason, we do not use the exploratory neural bottleneck runs as evidence. In those runs the learned code usage collapsed below the nominal category count, so they do not provide a

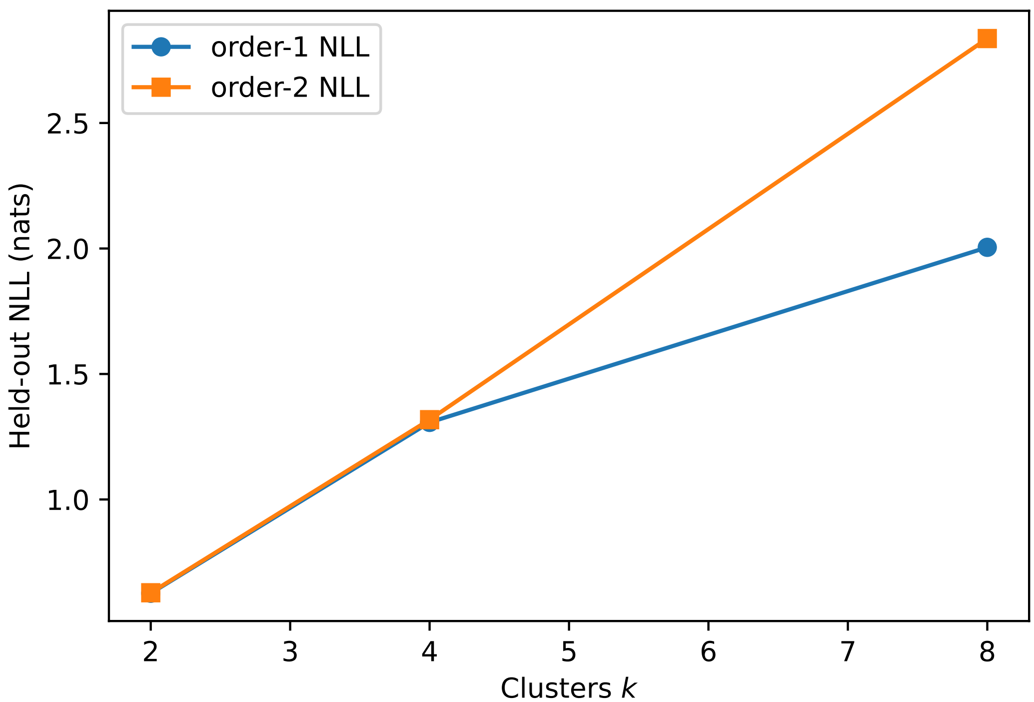

Figure 8: Auxiliary clustering result: held-out packaged predictive loss versus learned cluster count k. Both first- and second-order losses worsen as the learned partition becomes more finely split.

clean same-budget comparison to the clustering baseline. The controlled Markov, budgeted-prediction, and hashing demonstrations are therefore the proper basis of the main claims.

B Lean KL Bridge

The main text uses the expected-Kullback–Leibler form of closure deficit in Proposition 3.3. As a small formal appendix, the companion repository at https://github.com/ioannist/six-birds-randomness includes a machine-checked Lean development that bridges probability-level KL divergence to the expected log-likelihood ratio form used in the finite arguments [de Moura et al., 2015, The mathlib Community, 2020]. The formal artifact is intentionally modest. It does not attempt a full mechanization of conditional mutual information or of the complete closure-deficit theory. Instead it certifies the KL step on which the expected-divergence reading rests.

For probability measures μ and ν with $\mu \ll \nu$, the formal development proves the identity

$$(\mathrm{klDiv}(\mu, \nu))^{\mathrm{toReal}} = \int \mathrm{llr}_{\mu,\nu} \, d\mu,$$

that is, the real-valued Kullback–Leibler divergence equals the integral of the log-likelihood ratio under μ. From that bridge it derives the nonnegativity statement

$$0 \leq \int \mathrm{llr}_{\mu,\nu} \, d\mu.$$

The development also records the zero-divergence characterization for probability measures, namely that KL divergence vanishes exactly when the two measures are equal.

Conceptually, this appendix matters because the paper's finite-state formula

$$\mathsf{CD}_\tau(\Pi) = \sum_x \pi(x)\, D_{\mathrm{KL}}(p_x^{(\tau)} \,\|\, \bar{p}_{\Pi(x)}^{(\tau)})$$

is an instance of this more general expected-divergence pattern. The Lean artifact therefore does not formalize the entire manuscript, but it does machine-check the bridge between divergence and expected log-likelihood ratio that underwrites the KL reading of hidden predictive residue.

The tracked Lean files in the repository are the smoke module and the KL bridge module. Their role in the present paper is evidentiary rather than expository: they certify a core algebraic step, while the manuscript itself remains focused on the conceptual and empirical account of randomness as predictive non-closure.

References

Mihir Bellare and Phillip Rogaway. Random oracles are practical: A paradigm for designing efficient protocols. In *Proceedings of the 1st ACM Conference on Computer and Communications Security*, pages 62–73, 1993. doi: 10.1145/168588.168596.

Peter Buchholz. Exact and ordinary lumpability in finite markov chains. *Journal of Applied Probability*, 31(1):59–75, 1994. doi: 10.2307/3215235.

Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley-Interscience, Hoboken, NJ, 2 edition, 2006. ISBN 9780471241959. doi: 10.1002/047174882X.

Leonardo de Moura, Soonho Kong, Jeremy Avigad, Floris van Doorn, and Jakob von Raumer. The lean theorem prover (system description). In *Automated Deduction – CADE-25*, volume 9195 of *Lecture Notes in Computer Science*, pages 378–388. Springer, 2015. doi: 10.1007/978-3-319-21401-6_26.

Massimiliano Esposito. Stochastic thermodynamics under coarse graining. *Physical Review E*, 85(4):041125, 2012. doi: 10.1103/PhysRevE.85.041125.

Bernhard C. Geiger, Tatjana Petrov, Gernot Kubin, and Heinz Koeppl. Optimal kullback–leibler aggregation via information bottleneck. *IEEE Transactions on Automatic Control*, 60(4):1010–1022, 2015. doi: 10.1109/TAC.2014.2364971.

John G. Kemeny and J. Laurie Snell. *Finite Markov Chains*. Springer-Verlag, New York, 1976.

J. B. MacQueen. Some methods for classification and analysis of multivariate observations. In Lucien M. Le Cam and Jerzy Neyman, editors, *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Theory of Statistics*, pages 281–297, Berkeley, CA, 1967. University of California Press.

Phillip Rogaway and Thomas Shrimpton. Cryptographic hash-function basics: Definitions, implications, and separations for preimage resistance, second-preimage resistance, and collision resistance. In Bimal K. Roy and Willi Meier, editors, *Fast Software Encryption: 11th International Workshop, FSE 2004, Delhi, India, February 5–7, 2004, Revised Papers*, volume 3017 of *Lecture Notes in Computer Science*, pages 371–388. Springer, 2004. doi: 10.1007/978-3-540-25937-4_24.

Cosma Rohilla Shalizi and James P. Crutchfield. Computational mechanics: Pattern and prediction, structure and simplicity. *Journal of Statistical Physics*, 104(3–4):817–879, 2001. doi: 10.1023/A:1010388907793.

The mathlib Community. The lean mathematical library. In *Proceedings of the 9th ACM SIGPLAN International Conference on Certified Programs and Proofs*, pages 367–381, New York, NY, 2020. Association for Computing Machinery. doi: 10.1145/3372885.3373824.

Naftali Tishby, Fernando C. Pereira, and William Bialek. The information bottleneck method. In *Proceedings of the 37th Annual Allerton Conference on Communication, Control and Computing*, pages 368–377, 1999.

Ioannis Tsiokos. To notch a stone with six birds: Time as a theory artifact of order, measure, and arrow, 2026a. URL https://doi.org/10.5281/zenodo.18495363. Preprint.

Ioannis Tsiokos. To create a stone with six birds: Emergent geometric and thermodynamic regimes from a minimal stochastic substrate, 2026b. URL https://doi.org/10.5281/zenodo.18838994. Preprint.

Ioannis Tsiokos. Six birds: Foundations of emergence calculus, 2026c. URL https://doi.org/10.5281/zenodo.18365949. Preprint.