Credal Rate–Distortion Theory

Nicolò Bonacorsi, Michele Caprio, Francesca Meneghello, and Francesco Restuccia, *Senior Member, IEEE*

Abstract

We develop a rate–distortion theory for lossy compression when each source symbol is associated with a credal set of plausible downstream laws. Existing universal and minimax formulations account for uncertainty in the source or distortion model but typically retain a point-valued probabilistic description for each input. Credal-set metrics, on the other hand, compare set-valued uncertainty models without specifying how several such sets should be represented when their inputs share a codeword. Under support-only semantics, we assign each codeword the convex hull of its constituent credal sets. We introduce a directed free-energy replacement distortion and prove that the distortion of a compression cell equals its worst pairwise constituent distortion. Consequently, one-shot codebook size and minimum output entropy are characterized by the chromatic number and chromatic entropy of an incompatibility graph. For product credal families, coordinatewise observation yields disjunctive graph powers, with a fixed-length exponent determined by the fractional chromatic number and an expected-prefix rate given by Körner graph entropy, while bounded-window observation yields non-tensorizing block graphs and a regularized expected-prefix rate. Under unrestricted observation and every distortion tolerance below one, the asymptotic costs reduce to those of transmitting the credal equality classes. We characterize the exact graph-saturation window and bound its size. A three-symbol construction separates the coordinatewise, two-coordinate, and unrestricted expected-prefix rates as 0, $2/3$, and $\log_2 3$ bits per symbol. Experiments on the Street View House Numbers dataset illustrate the difference between credal-preserving compression and point-summary clustering.

Index Terms

Credal sets, epistemic uncertainty, graph coloring, graph entropy, imprecise probability, rate–distortion theory, zero-error source coding.

I. Introduction

TRADITIONALLY, rate–distortion theory characterizes the minimum rate required to satisfy a prescribed reconstruction criterion under a precisely specified probabilistic model [2]–[6]. While such a model captures *aleatoric uncertainty*, namely the intrinsic randomness of the source, it assumes that the underlying probability laws themselves are known. By contrast, in many learning and inference problems these laws cannot be reliably identified due to limited data, ambiguous or noisy labels, calibration error or model misspecification [8]–[10]. This gives rise to *epistemic uncertainty*, which is uncertainty about the probability model itself and not randomness under a known model.

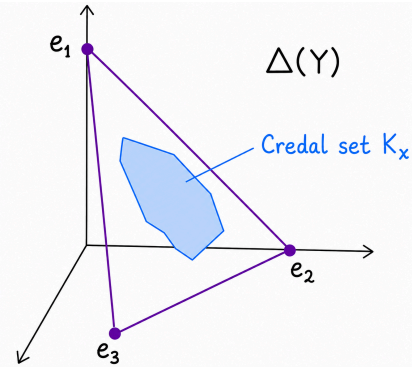

Fig. 1. A credal set K_x within the probability simplex $\Delta(Y)$ for $Y = \{1, 2, 3\}$. The vertices e_1, e_2, e_3 are the standard basis probability vectors, each representing certainty in one outcome; for example, $e_1 = (1, 0, 0)$. Their convex hull forms the full simplex, while the blue region K_x is a convex subset containing the plausible probability distributions associated with input x.

Credal sets – see Fig. 1 – provide a natural representation of epistemic uncertainty by collecting those laws into a closed convex family of distributions [11], [12], so that each input carries a set of plausible downstream laws. As such, *how should several such inputs be represented when they share a codeword, without discarding their unresolved uncertainty*? Existing universal and minimax formulations account for uncertainty in the source or distortion model but still retain point-valued input descriptions [13]–[15]. Conversely, while credal-set metrics compare set-valued uncertainty models [16], [17], they do not prescribe how multiple credal sets should be combined under compression or what coding law results from that aggregation.

In this paper, we address this gap through support-only aggregation. When several inputs are mapped to the same codeword, the decoder knows the corresponding compression cell but not which input in that cell occurred. The source prior and the encoder determine nominal posterior weights within the cell; our fidelity semantics deliberately does not use those weights to contract the decoded uncertainty. Instead, the representation must remain valid under arbitrary reweightings of the constituent inputs. We therefore assign to each codeword the convex hull of its constituent credal sets, namely the smallest credal set containing them all. The source prior determines the probability, and hence the coding cost, of each cell, whereas the credal sets determine the uncertainty that must be retained. Fig. 2 contrasts this construction with point-based merging.

N. Bonacorsi is with Northeastern University, United States (email: n.bonacorsi@northeastern.edu).
M. Caprio is with the University of Warwick, Coventry, UK (email: Michele.Caprio@warwick.ac.uk).
F. Meneghello is with Northeastern University, Boston, MA, USA (email: fr.meneghello@northeastern.edu).
F. Restuccia is with Northeastern University, Boston, MA, USA (email: f.restuccia@northeastern.edu)
A six-page preliminary version of part of this work is available in [1] and will be presented in part at IEEE Information Theory Workshop (IEEE ITW'26).

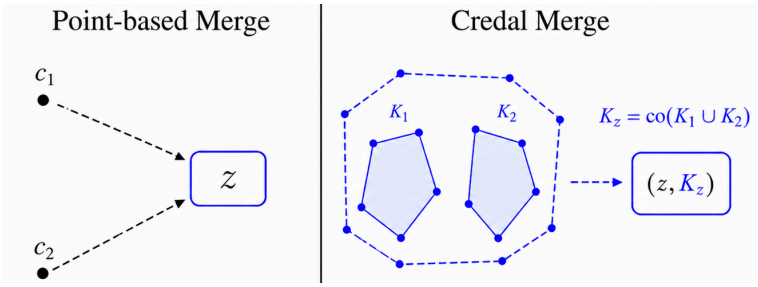

Fig. 2. Point-based vs credal merging. Left: two point representatives c_1, c_2 share a codeword z. Right: under support-only semantics, merging two credal likelihoods produces the aggregate $K_z = \mathrm{co}(K_1 \cup K_2)$, shown by the dashed outer hull, and the decoded object is (z, K_z). Fidelity must control replacement of each constituent set by this entire aggregate.

A second modeling choice is the class of downstream tests used to assess whether compression preserves the relevant information. At blocklength n, an observer may inspect a single coordinate, any subset of at most k coordinates or the entire output block. These choices induce different notions of fidelity. Two product credal sets may appear compatible when examined coordinatewise yet become distinguishable once a test couples multiple coordinates, i.e., increasing the observation window can reveal joint incompatibilities that are invisible to smaller windows. Consequently, the observer scale determines which inputs may share a codeword and must be built directly into the compression model. We show a simple example that demonstrates why the observer scale matters. The three-symbol construction of Section XI gives a simple illustration. At the chosen tolerance, the blocks $x^2 = (0, 1)$ and $u^2 = (1, 2)$ are compatible under coordinatewise observation because each coordinatewise comparison is individually safe. A two-coordinate observer, however, sees the two discrepancies jointly and declares the pair incompatible.

To formalize fidelity, we introduce a directed free-energy distortion that quantifies the additional downstream ambiguity introduced when a credal set K is replaced by another set L. The resulting directed distortion vanishes if and only if $L \subseteq K$, so a zero-cost replacement introduces no law that was not already plausible under K. Under convex-hull aggregation, the distortion of an entire compression cell reduces to the largest pairwise distortion among its constituent credal sets. So, at any given tolerance, feasible compression cells are exactly the independent sets of an induced incompatibility graph. This reduction connects credal compression to graph coloring and functional compression [18]–[21], while retaining a crucial distinction between the number of codewords and their usage.

We accordingly focus our attention on two key issues. In fixed-length coding, every codeword has the same length, so performance is determined by the minimum number of admissible codewords, equivalently the chromatic number of the incompatibility graph. In expected-prefix coding, codeword lengths may adapt to their probabilities, so performance is determined by the minimum entropy of the encoder output, i.e., the graph's chromatic entropy [22]. The resulting fidelity criterion differs from Shannon's usual expected per-letter distortion: the tolerance imposes a hard compatibility constraint on every compression cell, while the source prior determines the cost of the admissible cells. The two objectives generally differ under a nonuniform source prior and need not be attained by the same coloring.

Summary of Impact. The framework separates two roles that are usually coupled in probabilistic compression: the source prior determines the coding cost of a compression cell, whereas its credal sets determine the uncertainty that must remain admissible after compression. Under support-only aggregation, this uncertainty is retained through the convex hull of the constituent credal sets. The resulting hard-fidelity problem admits exact graph-theoretic characterizations and exposes how the coding law changes with the observation scale.

Summary of Novel Contributions. The main contributions of this work are described as follows:

1. ***One-shot credal compression:*** We prove that, under convex-hull aggregation, the distortion of a compression cell equals the largest pairwise distortion among its constituent credal sets. Feasible cells are therefore exactly the independent sets of the incompatibility graph. Hence the minimum one-shot codebook size and minimum output entropy are the graph's chromatic number and chromatic entropy, respectively. Under the same support-only semantics, encoder randomization improves neither objective;

2. ***Observer-dependent asymptotic rates:*** For product credal families, coordinatewise observation produces disjunctive powers of the one-shot graph. The fixed-length exponent is the base-two logarithm of its fractional chromatic number, while the expected-prefix rate is given by Körner graph entropy. Observation of at most k coordinates produces block graphs that need not tensorize; we prove that the corresponding expected-prefix rate exists and admits a regularized graph-entropy characterization. Under unrestricted observation and any distortion tolerance below one, the fixed-length exponent is the base-two logarithm of the number of credal equality classes, while the expected-prefix rate is the entropy of their induced distribution;

3. ***Graph saturation and separation of observer scales:*** We characterize the smallest observation window for which the bounded-window and unrestricted incompatibility graphs coincide at every blocklength, and derive an explicit finite upper bound in terms of pairwise credal separation. We also construct a three-symbol example whose coordinatewise, two-coordinate, and unrestricted expected-prefix rates under the uniform source are respectively 0, $2/3$, and $\log_2 3$ bits per source symbol. The intermediate value follows from an exact finite-block extremal theorem that also classifies all maximizing families;

4. ***Experimental evaluation:*** We evaluate the one-shot framework on sixty predictive credal sets from the Street View House Numbers dataset (SVHN) [23]. At a representative operating point, a minimum coloring uses fourteen codewords. With the

same number of clusters, a barycentric minimax baseline achieves a smaller point-summary radius but places 18.6% of within-cluster pairs above the credal incompatibility threshold, whereas the credal coloring has zero such violations by construction. The comparison illustrates the difference between point-summary and credal-preserving fidelity at a matched codebook size.

Paper organization. Section II reviews related work and Section III introduces the model. Sections IV–VI develop the one-shot theory. Sections VII–X analyze the coordinatewise, bounded-window and unrestricted block problems. Section IX-A proves finite-block non-attainment, Section XI establishes the exact intermediate rate and Section XII presents the numerical study.

II. Related Work

Information bottleneck and distributional clustering. The information bottleneck (IB) compresses an input X while preserving information about a relevance variable Y [24]. Inputs are compared through their conditional laws $p(y \mid x)$, and the compression–relevance tradeoff is measured using mutual information. Agglomerative IB implements this principle through prior-dependent hard merges [25]. *Our setting instead associates each input with a credal set K_x and not with a single conditional law*. When several inputs share a codeword, the decoder retains the convex hull of the union of their credal sets. This support-only aggregation rule leads to a one-shot graph-coloring problem [26].

Robust and universal lossy coding. Universal and minimax rate–distortion formulations seek codes that remain effective under an unknown source distribution or distortion measure [13], [14]. Distributionally robust formulations similarly optimize over ambiguity sets of source, channel or noise laws [15]. In these approaches, uncertainty concerns a global model governing the coding problem. Our uncertainty is instead indexed by the source symbol: observing x identifies a specific credal set K_x, and compression determines which symbols may share an aggregate credal representation. The source prior determines the cost of the resulting partition, while the credal sets determine its feasibility.

Credal sets and risk-based discrepancies. Credal sets, lower previsions and upper previsions provide general models of imprecise probability [11], [12]. Existing divergences quantify separation or relative informativeness between credal sets [16], while upper-expectation discrepancies compare uncertainty models through classes of test functions [17]. These works provide tools for comparing credal sets but do not specify how several sets should be aggregated when their associated inputs share a codeword. Our directed distortion is designed for this replacement problem: it measures the additional downstream ambiguity introduced when one credal set is replaced by another, and its zero set is determined by inclusion rather than equality. For singleton credal sets, its variational form is related to the log-Laplace envelope used in entropic risk [27]. For product credal sets, the test class determines whether replacement is assessed coordinatewise, over bounded windows or over the full block.

Graph entropy and functional compression. Graph coloring is key to zero-error source coding, where a confusability graph specifies which source symbols must remain distinguishable [18], [26]. For disjunctive graph powers, graph entropy characterizes the asymptotic expected-prefix rate [19]. Related graph formulations arise in functional compression, where the decoder must recover a prescribed function from compressed observations and side information [20], [21]. *In our setting, adjacency is induced by violation of a credal replacement constraint and not from failure to recover a fixed function.* Coordinate-wise observation produces disjunctive powers of a one-shot incompatibility graph and bounded-window observation produces block graphs that need not be powers of any single graph. This lack of tensorization gives rise to the regularized intermediate rate studied here.

Maximal-distortion graph and hypergraph coding. A closely related hard-fidelity literature studies source and functional compression under maximal distortion. Distortion-characteristic graphs can encode which source symbols may share a reconstruction [45], while more general maximal-distortion functional-compression problems can require characteristic hypergraphs because admissibility of an entire source cell need not be determined by its pairs [47]. Our setting has a different source of uncertainty: each source symbol carries a credal set, and compression replaces the credal sets associated with a cell by a common aggregate. The relevant structural result is therefore that, under support-only aggregation and the replacement distortion considered here, whole-cell admissibility reduces exactly to pairwise compatibility. Theorem IV.9 establishes this graph reduction.

Finite-blocklength and coloring objectives. Finite-blocklength lossy source coding typically controls expected distortion or excess-distortion probability relative to an asymptotic rate–distortion limit [28]. Once the credal incompatibility graph is fixed, our one-shot problem instead imposes a hard compatibility constraint: no adjacent pair may share a codeword. The minimum codebook size is therefore the chromatic number, while the minimum output entropy is the chromatic entropy. The two objectives may be attained by different colorings [22], [48]. Accordingly, the empirical study reports the entropy of a minimum-color solution but does not identify it with chromatic entropy.

III. System Model

Our model separates the fidelity constraint from the coding objective. The credal likelihoods determine which inputs may be grouped into the same compression cell, while the source prior determines how frequently each cell is used. Let X and Y be finite nonempty input and downstream alphabets.

For any finite nonempty set E, define

$$\Delta(E) := \Big\{ p = (p(a))_{a\in E} \in [0,1]^E : \sum_{a\in E} p(a) = 1 \Big\}. \tag{1}$$

For a set A in a finite-dimensional real vector space, $\mathrm{co}(A)$ denotes its convex hull. For $p \in \Delta(E)$, let

$$\mathrm{supp}(p) := \{a \in E : p(a) > 0\}.$$

For a finite graph G, $V(G)$ and $E(G)$ denote its vertex and edge sets, respectively.

A *credal set* on E is a nonempty compact convex subset of $\Delta(E)$, and $\mathscr{K}(E)$ denotes the collection of all such sets. Since $\Delta(E)$ is compact, compactness is equivalent to closedness. We write $\mathbb{N} = \{1, 2, \ldots\}$, $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$, and $[n] = \{1, \ldots, n\}$; $\mathbf{1}_A$ denotes the indicator of an event A, and $\mathbf{1}$ denotes the constant-one function on the ambient finite alphabet. Each input $x \in X$ is associated with a *credal likelihood* $K_x \in \mathscr{K}(Y)$, representing the plausible conditional laws of the downstream quantity given x. The source prior $\pi \in \Delta(X)$ determines the probability of each compression cell. Whenever no confusion can arise, we also use X to denote an X-valued source random variable with law π. It does not select a particular law from K_x or reduce the uncertainty retained after aggregation. Unless stated otherwise, the fidelity constraint is imposed on every input in X, including inputs with zero prior probability; such inputs may therefore affect fixed-length feasibility even though they do not contribute to entropy under π.

Cells. A deterministic encoder $e : X \to Z$ maps each input to a codeword in a finite alphabet Z. Thus, every active codeword $z \in e(X)$ identifies the compression cell $C_z = e^{-1}(z)$. Once the encoder retains only z, the decoder knows that the original input belonged to C_z but not which $x \in C_z$ occurred. The source prior and the encoder determine nominal posterior weights within C_z; under the support-only convention, these weights are deliberately not used to contract the uncertainty retained by the codeword. Hence the codeword must remain valid under arbitrary reweightings of the constituent inputs, i.e.,

$$K_{C_z} := \mathrm{co}\Big(\bigcup_{x \in C_z} K_x \Big). \tag{2}$$

This is the smallest credal set, under set inclusion, that contains every credal likelihood associated with the cell. For a cell C_z satisfying $\sum_{x \in C_z} \pi(x) > 0$, if the nominal posterior weights induced by π were trusted for fidelity, one could instead use the generally smaller family

$$\left\{ \sum_{x \in C_z} \mathbb{P}(X = x \mid X \in C_z) p_x : p_x \in K_x \text{ for every } x \in C_z \right\}. \tag{3}$$

That choice corresponds to a different operational assumption. We adopt (2): the source prior determines the probability, and hence the coding cost, of each cell, but it does not restrict the uncertainty retained within that cell. Fig. 2 contrasts point-based merging with the full ambiguity set preserved by a credal merge. The one-shot problem is then to partition X into compression cells that minimize either the number of codewords or their prior-weighted description length, while ensuring that replacing each K_x by its cell aggregate K_{C_z} satisfies the fidelity constraint. The replacement geometry developed next turns this requirement into a graph-based coding law.

Observers. For a block $x^n \in X^n$, we model downstream uncertainty through the independent product of the coordinate credal sets. Fidelity then depends on how much of the downstream block Y^n an observer is allowed to inspect. We consider three nested observer classes:

(i) coordinate-local observers, which inspect a single coordinate;
(ii) bounded-window observers, which inspect any subset of at most k coordinates;
(iii) global observers, which may inspect the entire block.

These classes capture progressively richer interactions across coordinates and therefore induce increasingly stringent fidelity constraints. We characterize the coding laws associated with each regime, summarized in Table I, as well as the smallest window for which bounded-window and global incompatibility coincide at every blocklength.

TABLE I
OBSERVER REGIMES AND CORRESPONDING CODING LAWS

Observer	**Block incompatibility structure**	**Asymptotic coding law**
Coordinate-local	Disjunctive powers of the one-shot incompatibility graph	Single-letter fixed-length and expected-prefix laws
Bounded k-window	Finite-window block graphs that need not tensorize	Regularized expected-prefix law
Global	Unrestricted full-block incompatibility graphs	Equality-class fixed-length and expected-prefix laws below the distortion ceiling
Sufficiently large finite window	Exact blockwise agreement with unrestricted incompatibility	Exact saturation to the global regime

IV. Replacement Distortion for Credal Sets

We measure the cost of replacing one credal set by another through a directed distortion built from free-energy envelopes. For a finite nonempty set E, let $f : E \to \mathbb{R}$, $p \in \Delta(E)$, and $K \in \mathscr{K}(E)$. Define

$$\Lambda_p(f) := \log \sum_{a\in E} p(a)e^{f(a)}, \qquad \Psi_K(f) := \sup_{p\in K} \Lambda_p(f). \tag{4}$$

Here, $\Lambda_p(f)$ is the log-Laplace transform of the score f under p, while $\Psi_K(f)$ is its worst-case value over all distributions in the credal set K. Thus, Ψ_K captures the largest downstream score consistent with the uncertainty represented by K. Let $\operatorname{osc}(f) = \max_{a\in E} f(a) - \min_{a\in E} f(a)$. Since $\Psi_K(f + c) = c + \Psi_K(f)$ for every constant c, comparisons between credal sets are invariant to additive shifts of f and depend only on its oscillation. We therefore restrict attention to the normalized test class $\mathcal{F}(E) = \{f : E \to \mathbb{R} : \min f = 0,\ \max f \le 1\}$, and write $\mathcal{F} = \mathcal{F}(Y)$ throughout the one-shot analysis. This normalization removes irrelevant offsets, bounds the score range and places all replacement distortions in $[0, 1]$.

Shannon information quantities are measured in bits (i.e., base-two logarithms). By contrast, all unsubscripted logarithms appearing in free-energy and relative-entropy expressions are natural. For finite-valued random variables U and V, we write

$$H(U) := -\sum_u \mathbb{P}(U = u) \log_2 \mathbb{P}(U = u), \qquad I(U; V) := H(U) - H(U \mid V), \tag{5}$$

with the convention $0 \log_2 0 := 0$, where $H(U \mid V)$ is the usual conditional entropy. For a probability vector ρ, $H(\rho)$ denotes the entropy of a random variable with law ρ. The dependence of $\Psi_K(f)$ on the credal set K can be expressed through the support function of K evaluated at e^f. This motivates the following convex-geometric quantity:

Definition IV.1. For any finite nonempty set E, any function $g : E \to \mathbb{R}$, and any credal set $K \in \mathscr{K}(E)$, let us define the support envelope by

$$S_K(g) := \sup_{p\in K} \langle p, g\rangle = \sup_{p\in K} \sum_{a\in E} p(a)g(a). \tag{6}$$

Proposition IV.2. *For every $K \in \mathscr{K}(Y)$ and every $f : Y \to \mathbb{R}$, $\Psi_K(f) = \log S_K(e^f)$, where e^f denotes the pointwise exponential of f.*

Proof. For every $p \in K$, $\Lambda_p(f) = \log \sum_y p(y)e^{f(y)} = \log\langle p, e^f\rangle$. By monotonicity of the logarithm, it follows that $\Psi_K(f) = \log \sup_{p\in K}\langle p, e^f\rangle = \log S_K(e^f)$. □

The logarithmic form also matches the product structure: support envelopes multiply under independent products, and Theorem A.3 shows that the logarithm is essentially the unique increasing transform that makes them additive. A complementary interpretation follows from convex duality. For $r, p \in \Delta(Y)$, define the relative entropy

$$\mathrm{KL}(r\|p) := \sum_{y:r(y)>0} r(y) \log \frac{r(y)}{p(y)}, \tag{7}$$

with $\mathrm{KL}(r\|p) = +\infty$ whenever $r(y) > 0 = p(y)$ for some y. Let $I_K(r) := \inf_{p\in K} \mathrm{KL}(r\|p)$ denote the minimum relative-entropy cost of representing r by a distribution in K. Appendix B shows that

$$\Psi_K(f) = \sup_{r\in\Delta(Y)} \big\{\langle r, f\rangle - I_K(r)\big\}. \tag{8}$$

Thus, the free-energy envelope is supported both by the product-additivity requirement and by an equivalent convex-dual representation. Since replacing K by L is harmful only when it increases the largest downstream score, the resulting distortion is naturally directed.

Definition IV.3. For $K, L \in \mathscr{K}(Y)$, define the directed and symmetrized distortions by

$$\begin{aligned} d_+(K, L) &:= \sup_{f\in\mathcal{F}} \big(\Psi_L(f) - \Psi_K(f)\big), \\ d(K, L) &:= \max\{d_+(K, L), d_+(L, K)\}. \end{aligned} \tag{9}$$

Since $f \in \mathcal{F}$ takes values in $[0, 1]$, $1 \le \sum_y p(y)e^{f(y)} \le e$ and hence $0 \le \Lambda_p(f), \Psi_K(f) \le 1$. Therefore,

$$0 \le d_+(K, L) \le 1, \qquad 0 \le d(K, L) \le 1. \tag{10}$$

For $K \in \mathscr{K}(Y)$, define

$$\operatorname{supp}(K) := \bigcup_{p\in K} \operatorname{supp}(p). \tag{11}$$

At the opposite endpoint,

$$d_+(K, L) = 1 \quad \Longleftrightarrow \quad \exists\, q \in L : \operatorname{supp}(q) \cap \operatorname{supp}(K) = \varnothing. \tag{12}$$

Indeed, compactness implies that the supremum defining d_+ is attained. Since every normalized test satisfies $0 \le \Psi_K(f) \le 1$, equality $d_+(K,L) = 1$ requires $\Psi_L(f) = 1$ and $\Psi_K(f) = 0$ for some $f \in \mathcal{F}$. The first equality yields a $q \in L$ supported on the level set $\{f = 1\}$, whereas the second forces every $p \in K$ to be supported on $\{f = 0\}$. Conversely, if such a q exists, choosing $f = 1$ on $\mathrm{supp}(q)$ and $f = 0$ on $\mathrm{supp}(K)$ gives $\Psi_L(f) - \Psi_K(f) = 1$.

Thus, zero directed distortion should mean that replacing K by L introduces no new law, i.e., $L \subseteq K$.

Theorem IV.4. *For $K, L \in \mathscr{K}(Y)$,*

$$d_+(K,L) = 0 \quad \Longleftrightarrow \quad L \subseteq K. \tag{13}$$

Proof. If $L \subseteq K$, then $\Psi_L(f) \le \Psi_K(f)$ for every $f \in \mathcal{F}$, so $d_+(K,L) = 0$.

Conversely, suppose $d_+(K,L) = 0$ but $L \not\subseteq K$, and choose $q_0 \in L \setminus K$. Strong separation of the compact convex set K in $\mathbb{R}^Y$ [29] gives $h : Y \to \mathbb{R}$ and $\eta > 0$ such that

$$\langle q_0, h\rangle \ge \sup_{p\in K} \langle p, h\rangle + \eta. \tag{14}$$

Set $\widetilde{h} = h - \min h \ge 0$. Translation by a constant preserves the separation because all probability vectors have mass one, so

$$\langle q_0, \widetilde{h}\rangle - \sup_{p\in K} \langle p, \widetilde{h}\rangle \ge \eta. \tag{15}$$

Strict separation makes $M := \max \widetilde{h} > 0$. With $\alpha = (e-1)/M$, $g = \mathbf{1} + \alpha\widetilde{h}$ satisfies $1 \le g \le e$ and $f := \log g \in \mathcal{F}$. Moreover,

$$\begin{aligned} S_L(g) - S_K(g) &\ge \alpha\Big(\langle q_0, \widetilde{h}\rangle - \sup_{p\in K} \langle p, \widetilde{h}\rangle\Big) \\ &\ge \alpha\eta > 0. \end{aligned} \tag{16}$$

Proposition IV.2 then yields $\Psi_L(f) > \Psi_K(f)$, a contradiction. Thus $L \subseteq K$. □

The directed quantity is the operational replacement penalty. Its symmetrization supplies the ordinary geometry needed to compare two inputs and define incompatibility edges.

Corollary IV.5. *The map $d : \mathscr{K}(Y) \times \mathscr{K}(Y) \to [0,1]$ is a metric.*

Proof. Nonnegativity, symmetry, and $d(K,K) = 0$ are immediate. For the triangle inequality, fix $K, L, M \in \mathscr{K}(Y)$. For every $f \in \mathcal{F}$,

$$\begin{aligned} \Psi_M(f) - \Psi_K(f) &= \big(\Psi_M(f) - \Psi_L(f)\big) + \big(\Psi_L(f) - \Psi_K(f)\big) \\ &\le d_+(L,M) + d_+(K,L). \end{aligned} \tag{17}$$

Taking the supremum shows that d_+ is subadditive. Hence

$$\begin{aligned} d(K,M) &= \max\{d_+(K,M), d_+(M,K)\} \\ &\le \max\{d_+(K,L) + d_+(L,M), d_+(M,L) + d_+(L,K)\} \\ &\le \max\{d_+(K,L), d_+(L,K)\} + \max\{d_+(L,M), d_+(M,L)\} \\ &= d(K,L) + d(L,M). \end{aligned}$$

If $d(K,L) = 0$, Theorem IV.4 gives both $L \subseteq K$ and $K \subseteq L$, hence $K = L$. □

A. Aggregation

For the family $(K_x)_{x\in X}$, the support-only hull acts on free-energy envelopes as a pointwise maximum, reducing cluster distortion to its worst pairwise value.

Theorem IV.6. *Let $K_1, \dots, K_m \in \mathscr{K}(Y)$ and $K_* := \mathrm{co}(\bigcup_{i=1}^m K_i)$. Then for every $f : Y \to \mathbb{R}, \Psi_{K_*}(f) = \max_{1\le i\le m} \Psi_{K_i}(f)$.*

Proof. Set $g = e^f$. Proposition IV.2 reduces the claim to $S_{K_*}(g) = \max_i S_{K_i}(g)$. One inequality follows from $K_i \subseteq K_*$. Conversely, write any $p \in K_*$ as $p = \sum_{j=1}^n \alpha_j p_j$, where $p_j \in K_{i_j}$, $\alpha_j \ge 0$, and $\sum_j \alpha_j = 1$. Then

$$\langle p, g\rangle = \sum_{j=1}^n \alpha_j \langle p_j, g\rangle \le \max_{1\le j\le n} \langle p_j, g\rangle \le \max_{1\le i\le m} S_{K_i}(g). \tag{18}$$

Taking the supremum gives the reverse inequality. □

Corollary IV.7. *For any $K, L \in \mathscr{K}(Y)$, $\Psi_{\mathrm{co}(K\cup L)} = \max\{\Psi_K, \Psi_L\}$ as functions on $\mathbb{R}^Y$. Consequently,*

$$d_+\big(K, \mathrm{co}(K \cup L)\big) = d_+(K,L), \text{ and } d_+\big(L, \mathrm{co}(K \cup L)\big) = d_+(L,K). \tag{19}$$

Proof. Theorem IV.6 with $m = 2$ gives, for every $f \in \mathcal{F}$,

$$\Psi_{\mathrm{co}(K\cup L)}(f) - \Psi_K(f) = \max\{0, \Psi_L(f) - \Psi_K(f)\}. \tag{20}$$

Because $0 \in \mathcal{F}$, the supremum of the second argument of the maximum is nonnegative. Taking suprema proves the first identity; exchanging K and L proves the second. □

For nonempty $C \subseteq X$, let $K_C := \mathrm{co}(\bigcup_{x\in C} K_x)$. Equivalently, using convexity of the individual sets,

$$K_C = \left\{ \sum_{x\in C} \alpha_x p_x : \alpha_x \geq 0,\ \sum_{x\in C} \alpha_x = 1,\ p_x \in K_x \right\}. \tag{21}$$

For an encoder fiber, K_C is the decoded credal likelihood. The prior prices the codeword but does not restrict the weights $(\alpha_x)_{x\in C}$ used for fidelity.

Definition IV.8. For nonempty $C \subseteq X$, let $\mathsf{D}(C) := \sup_{x\in C} d_+(K_x, K_C)$. If $e : X \to Z$ is deterministic with $C_z = e^{-1}(z)$, let $\mathsf{D}(e) := \sup_{z\in e(X)} \mathsf{D}(C_z)$. The encoder e is *ε-faithful* if $\mathsf{D}(e) \leq \varepsilon$.

Theorem IV.9. *For every nonempty subset $C \subseteq X$,*

$$\mathsf{D}(C) = \max_{x,x'\in C} d_+(K_x, K_{x'}). \tag{22}$$

Although the decoder replaces every member of a cell by a single convex hull, no higher-order feasibility constraint remains: the worst ordered pair already determines the full cell distortion.

Proof. Theorem IV.6 gives $\Psi_{K_C}(f) = \max_{x'\in C} \Psi_{K_{x'}}(f)$. Thus, for fixed $x \in C$,

$$d_+(K_x, K_C) = \sup_{f\in\mathcal{F}} \big(\Psi_{K_C}(f) - \Psi_{K_x}(f)\big) = \sup_{f\in\mathcal{F}} \max_{x'\in C} \big(\Psi_{K_{x'}}(f) - \Psi_{K_x}(f)\big).$$

Because C is finite, the supremum and maximum commute, giving $d_+(K_x, K_C) = \max_{x'\in C} d_+(K_x, K_{x'})$. Maximizing over x proves the claim. □

Corollary IV.10. *A nonempty subset $C \subseteq X$ satisfies $\mathsf{D}(C) \leq \varepsilon$ if and only if*

$$d(K_x, K_{x'}) \leq \varepsilon \qquad \textit{for all } x, x' \in C.$$

Proof. By Theorem IV.9, $\mathsf{D}(C) \leq \varepsilon$ iff $d_+(K_x, K_{x'}) \leq \varepsilon$ for all ordered pairs in $C \times C$, equivalently

$$\max\{d_+(K_x, K_{x'}), d_+(K_{x'}, K_x)\} \leq \varepsilon \qquad \text{for all } x, x' \in C, \tag{23}$$

which is the stated condition. □

Since $K_C \supseteq K_x$, $d_+(K_x, K_C)$ measures the ambiguity introduced by the merge. Thus, at zero tolerance, a cell can contain only identical credal sets.

V. One-Shot Coding

The pairwise characterization of cell distortion turns one-shot feasibility into a graph problem. For $\varepsilon \in [0, 1]$, define the *ε-incompatibility graph* G_ε on vertex set X, with an edge $x \sim x'$ whenever $x \neq x'$ and $d(K_x, K_{x'}) > \varepsilon$. Thus, two adjacent inputs cannot belong to the same ε-faithful compression cell. For a finite graph G, let $\chi(G)$ denote its chromatic number and $\alpha(G)$ its independence number. For any subset A of the vertex set, $G[A]$ denotes the subgraph induced by A. Let $M^*(\varepsilon)$ denote the minimum number of active codewords among all deterministic encoders $e : X \to Z$ satisfying $\mathsf{D}(e) \leq \varepsilon$.

Theorem V.1. *For every $\varepsilon \in [0, 1]$, $M^*(\varepsilon) = \chi(G_\varepsilon)$. As such, the optimal one-shot fixed-length binary description length is $\mathsf{L}^*_{\mathrm{fl}}(\varepsilon) = \lceil \log_2 \chi(G_\varepsilon) \rceil$.*

Proof. If e is ε-faithful, Corollary IV.10 makes every fiber $e^{-1}(z)$ independent in G_ε. Coloring by codeword therefore gives $\chi(G_\varepsilon) \leq |e(X)|$, and hence $\chi(G_\varepsilon) \leq M^*(\varepsilon)$. Conversely, every color class of a proper coloring is independent, so Corollary IV.10 makes the coloring an ε-faithful encoder. A minimum coloring gives $M^*(\varepsilon) \leq \chi(G_\varepsilon)$. The fixed-length formula follows immediately. □

At $\varepsilon = 0$, two inputs are compatible if and only if their credal sets are identical. Hence, G_0 is complete multipartite, with parts $\{x : K_x = K\}$, and $M^*(0)$ equals the number of distinct credal sets. More generally, increasing the distortion tolerance can only remove incompatibility edges: if $\varepsilon_1 \leq \varepsilon_2$, then $G_{\varepsilon_2} \subseteq G_{\varepsilon_1}$. Thus, $M^*(\varepsilon)$ is a nonincreasing step function whose changes occur only at pairwise credal distances. Finally, although support-only semantics also allows stochastic encoders, randomization cannot reduce either the number of active codewords or the output entropy.

A. Entropy and prefix codes

The source prior converts codebook size into a prior-weighted coding cost. Minimum color count and minimum color entropy are generally different objectives: the former depends only on the number of cells, while the latter also accounts for how probability mass is distributed across them. As a result, an entropy-optimal coloring may use more colors than a minimum coloring. Let $\pi \in \Delta(X)$ be the source prior. For a deterministic encoder $e : X \to Z$ and $X \sim \pi$, define the encoder output $Z_e := e(X)$. Its distribution is $\pi_e(z) = \mathbb{P}(Z_e = z) = \sum_{x \in e^{-1}(z)} \pi(x)$ for $z \in e(X)$, and its entropy is $H(Z_e) = -\sum_z \pi_e(z) \log_2 \pi_e(z)$.

Definition V.2. The *prior-weighted credal compression entropy* at distortion level ε is

$$\mathsf{H}^*_\pi(\varepsilon) := \inf\big\{H(Z_e) : e : X \to Z \text{ deterministic and } \mathsf{D}(e) \le \varepsilon\big\}. \tag{24}$$

The *chromatic entropy* of the probabilistic graph (G_ε, π) is [22]

$$H_\chi(G_\varepsilon, \pi) := \inf\big\{H(c(X)) : c \text{ is a proper coloring of } G_\varepsilon\big\}. \tag{25}$$

Theorem V.3. *For every prior $\pi \in \Delta(X)$ and every $\varepsilon \in [0, 1]$, $\mathsf{H}^*_\pi(\varepsilon) = H_\chi(G_\varepsilon, \pi)$.*

Proof. Every ε-faithful deterministic encoder is a proper coloring by Theorem V.1, with the same output and color entropy. Conversely, Corollary IV.10 makes every proper coloring ε-faithful. The two feasible classes, and thus their infima, coincide. □

Theorem V.4 (Support randomization)**.** *Let $Q = (Q(z \mid x))$ be a stochastic encoder with finite output alphabet Z. For every active codeword let*

$$C_z(Q) := \{x \in X : Q(z \mid x) > 0\}, \text{ and } K_z^Q := \mathrm{co}\Big(\bigcup_{x \in C_z(Q)} K_x\Big), \tag{26}$$

and call Q support-ε-faithful if

$$\max_{z : C_z(Q) \neq \varnothing} \max_{x \in C_z(Q)} d_+(K_x, K_z^Q) \le \varepsilon. \tag{27}$$

Then the smallest number of active codewords among support-ε-faithful stochastic encoders equals $\chi(G_\varepsilon)$. Moreover, for every prior π, with $X \sim \pi$ and Z_Q distributed according to the channel Q,

$$\inf_{Q \text{ support-}\varepsilon\text{-faithful}} H(Z_Q) = H_\chi(G_\varepsilon, \pi) = \mathsf{H}^*_\pi(\varepsilon). \tag{28}$$

Proof. If Q is support-ε-faithful, then every support fiber $C_z(Q)$ is an independent set of G_ε by Corollary IV.10. For each x, select any codeword z such that $Q(z \mid x) > 0$. This produces a deterministic proper coloring using no more active codewords than Q. The reverse inequality follows immediately because deterministic encoders are a special case of stochastic encoders. For the entropy objective, let $A_x = \{z : Q(z \mid x) > 0\}$ and define $\mathcal{Q}_A = \{\widetilde{Q} : \widetilde{Q}(\cdot \mid x) \in \Delta(A_x) \text{ for all } x\}$. Every $\widetilde{Q} \in \mathcal{Q}_A$ remains support-ε-faithful because its support fibers can only shrink. The output entropy is concave in $\widetilde{Q}$, and $\mathcal{Q}_A$ is a compact polytope, so a minimum is attained at an extreme point. Since every extreme point is deterministic row by row, there exists an ε-faithful deterministic encoder with output entropy no larger than that of Q. Theorem V.3 then gives the claimed identity. □

Binary prefix codes may assign the empty string when the output alphabet is a singleton, so codeword lengths are in $\mathbb{N}_0$.

Corollary V.5. *Let $\mathsf{L}^*_\pi(\varepsilon)$ denote the minimum expected binary prefix-code length obtained by first choosing a deterministic ε-faithful encoder and then assigning a prefix-free binary code to its output alphabet. Then*

$$H_\chi(G_\varepsilon, \pi) \le \mathsf{L}^*_\pi(\varepsilon) \le H_\chi(G_\varepsilon, \pi) + 1. \tag{29}$$

Proof. For any deterministic ε-faithful encoder e, Shannon's source-coding bound gives

$$H(Z_e) \le \mathbb{E}[\ell(Z_e)] \le H(Z_e) + 1.$$

Taking the infimum over all such encoders and applying Theorem V.3 proves the result. □

VI. An Example: Binary Intervals

We next illustrate the one-shot theory for binary interval credal sets $I[a, b] = \{p \in \Delta(\{0, 1\}) : p(1) \in [a, b]\}$, with $0 \le a \le b \le 1$. Every normalized binary test has the form $(0, t)$ or $(t, 0)$ for some $t \in [0, 1]$. Evaluating the corresponding support envelopes and using monotonicity in t yields

$$d_+\big(I[a, b], I[c, d]\big) = \max\left\{\log \frac{1 + d(e-1)}{1 + b(e-1)}, \log \frac{1 + (1-c)(e-1)}{1 + (1-a)(e-1)}, 0\right\}. \tag{30}$$

As expected from Theorem IV.4, this quantity is zero exactly when $[c, d] \subseteq [a, b]$. Let us consider the five credal sets

$$K_1 = I[0.90, 1.00], \quad K_2 = I[0.65, 0.85], \quad K_3 = I[0.40, 0.60], K_4 = I[0.15, 0.35], K_5 = I[0.00, 0.10]. \tag{31}$$

Equation (30) directly determines the incompatibility graphs in Fig. 3. All graph decisions use the unrounded distances. In particular, $d(K_1, K_3) = d(K_3, K_5) \approx 0.549948 < 0.55$; numerical values displayed or discussed to two decimal places should therefore not be interpreted as the values used for thresholding. At $\varepsilon = 0.40$, an optimal partition is $\{K_1, K_2\}, \{K_3\}, \{K_4, K_5\}$, while at $\varepsilon = 0.55$ it becomes $\{K_1, K_2, K_3\}, \{K_4, K_5\}$. Thus, increasing the fidelity tolerance removes incompatibility edges and reduces the required codebook from three codewords to two.

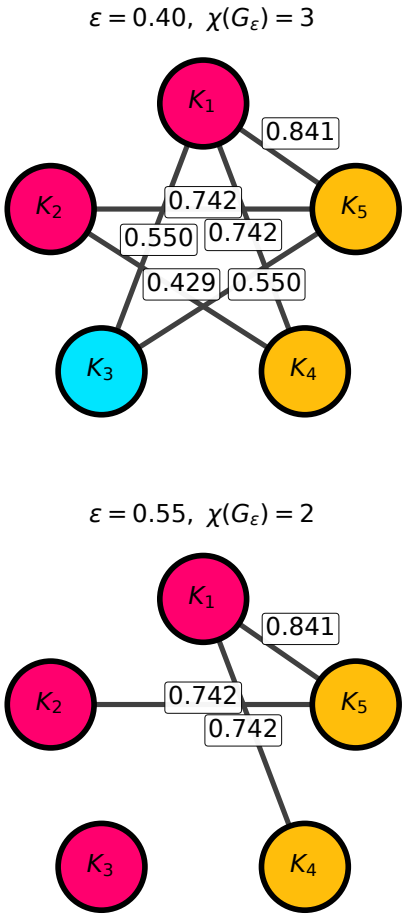

Fig. 3. Incompatibility graphs for five binary interval credal sets at $\varepsilon = 0.40$ and $\varepsilon = 0.55$. An edge joins two sets whose symmetric replacement distance exceeds the threshold. The displayed colorings attain chromatic numbers 3 and 2, respectively.

For a simple classification interpretation, let 1 denote *cat* and 0 denote *dog*. Set $K_a = I[1,1]$, $K_b = I[0.3, 0.8]$, and $K_c = I[0,0]$, corresponding respectively to a clear cat, an ambiguous example, and a clear dog. Direct substitution gives

$$\begin{aligned} d(K_a, K_b) &= \log \frac{3+7e}{10} \approx 0.789728, \\ d(K_b, K_c) &= \log \frac{1+4e}{5} \approx 0.864840, \\ d(K_a, K_c) &= 1. \end{aligned} \tag{32}$$

Hence, for $\varepsilon \in [\log((3+7e)/10), \log((1+4e)/5))$, a and b may share a compression cell, while the pairs (b, c) and (a, c) may not. Among the two natural one-bit partitions, only $\{a, b\} \mid \{c\}$ is ε-faithful. ***The example shows that an ambiguous input can be merged with a compatible clear input while retaining its uncertainty instead of reducing it to a point prediction.***

VII. Local Tests

Let us now consider product credal families when the downstream observer inspects one coordinate at a time. In this regime, block incompatibility inherits a simple product structure and is governed by disjunctive, or OR, powers of the one-shot incompatibility graph. Proofs are collected in Appendix D.

Definition VII.1 (Independent product credal set)**.** Let E_1 and E_2 be finite nonempty sets and let $K_i \in \mathscr{K}(E_i)$ for $i = 1, 2$. Their independent product is $K_1 \otimes K_2 := \mathrm{co}\{p_1 \otimes p_2 : p_1 \in K_1,\ p_2 \in K_2\} \subseteq \Delta(E_1 \times E_2)$, where $(p_1 \otimes p_2)(a_1, a_2) := p_1(a_1)p_2(a_2)$. Iterated products are defined associatively. This construction is the strong extension of the marginal credal sets.

Definition VII.2 (Disjunctive graph product)**.** For finite graphs G_1 on V_1 and G_2 on V_2, their disjunctive, or OR, product $G_1 \vee G_2$ is the graph on $V_1 \times V_2$ in which $(v_1, v_2) \sim (u_1, u_2)$ if and only if $v_1 \sim u_1$ in G_1 or $v_2 \sim u_2$ in G_2. We write $G^{\vee n}$ for the n-fold OR power.

For a block $x^n = (x_1, \dots, x_n) \in X^n$, define $K_{x^n}^{\otimes} = K_{x_1} \otimes \cdots \otimes K_{x_n} \subseteq \Delta(Y^n)$, equivalently,

$$K_{x^n}^{\otimes} := \mathrm{co}\Big\{p_1 \otimes \cdots \otimes p_n : p_i \in K_{x_i}, 1 \le i \le n\Big\}. \tag{33}$$

This strong extension generally differs from the independent natural extension [30]; see Appendix A. The choice is deliberate: each extreme law is a product of admissible coordinate laws, while convexification represents uncertainty over which such product law governs the block. Consequently, the construction does not assert independence under every distribution in $K_{x^n}^{\otimes}$; mixtures of product laws may themselves be dependent. For any nonempty $C \subseteq X^n$, we similarly define the aggregate $K_C^{\otimes} = \mathrm{co}(\bigcup_{u^n \in C} K_{u^n}^{\otimes})$. To model coordinatewise observation, for $1 \le i \le n$ and $f \in \mathcal{F}$, let $(\Gamma_i f)(y_1, \dots, y_n) = f(y_i)$ and define $\mathcal{F}_n^{\mathrm{loc}} = \{\Gamma_i f : 1 \le i \le n,\ f \in \mathcal{F}\}$. Thus, every admissible test depends on at most one coordinate of the downstream block. For $A, B \in \mathscr{K}(Y^n)$, define

$$d_{+,n}^{\mathrm{loc}}(A, B) := \sup_{F \in \mathcal{F}_n^{\mathrm{loc}}} \big(\Psi_B(F) - \Psi_A(F)\big), \qquad d_n^{\mathrm{loc}}(A, B) := \max\{d_{+,n}^{\mathrm{loc}}(A, B), d_{+,n}^{\mathrm{loc}}(B, A)\}. \tag{34}$$

Since a local test depends on only one coordinate, all remaining tensor factors integrate to one. As such, the resulting block distortion reduces to the largest one-shot distortion across coordinates. Importantly, this simplification comes from the observer restriction, not from the product credal model itself, i.e., $K_{x^n}^{\otimes}$ still contains joint laws on Y^n. Once tests are allowed to couple several coordinates, new incompatibilities can appear even when every coordinatewise comparison is individually compatible.

Theorem VII.3 (Local tensorization)**.** *The following statements hold.*

(i) *For every* $K_1, \dots, K_n, L_1, \dots, L_n \in \mathscr{K}(Y)$,

$$d_{+,n}^{\mathrm{loc}}(K_1 \otimes \cdots \otimes K_n, L_1 \otimes \cdots \otimes L_n) = \max_{1 \le i \le n} d_+(K_i, L_i).$$

Consequently,

$$d_n^{\mathrm{loc}}(K_1 \otimes \cdots \otimes K_n, L_1 \otimes \cdots \otimes L_n) = \max_{1\le i\le n} d(K_i, L_i). \tag{35}$$

(ii) *For every nonempty cluster $C \subseteq X^n$ and every $x^n \in C$,*

$$d_{+,n}^{\mathrm{loc}}(K_{x^n}^{\otimes}, K_C^{\otimes}) = \max_{u^n\in C} \max_{1\le i\le n} d_+(K_{x_i}, K_{u_i}).$$

Hence, with $\mathsf{D}_n^{\mathrm{loc}}(C) := \sup_{x^n\in C} d_{+,n}^{\mathrm{loc}}(K_{x^n}^{\otimes}, K_C^{\otimes})$,

$$\mathsf{D}_n^{\mathrm{loc}}(C) = \max_{x^n,u^n\in C} \max_{1\le i\le n} d_+(K_{x_i}, K_{u_i}). \tag{36}$$

(iii) *For every nonempty $C \subseteq X^n$ and every $\varepsilon \in [0,1]$, the following are equivalent:*

(a) $\mathsf{D}_n^{\mathrm{loc}}(C) \le \varepsilon$;

(b) *for all $x^n, u^n \in C$, $\max_{1\le i\le n} d(K_{x_i}, K_{u_i}) \le \varepsilon$;*

(c) *C is an independent set of the OR power $G_\varepsilon^{\vee n}$.*

Remark VII.4. *The local restriction is essential: coupled tests can create an incompatibility even when every coordinate is safe. Independent products ensure that the coordinate marginals remain the symbol-indexed credal sets.*

For deterministic $e_n : X^n \to Z_n$ with fibers $C_z = e_n^{-1}(z)$, let $\mathsf{D}_n^{\mathrm{loc}}(e_n) = \sup_{z\in e_n(X^n)} \mathsf{D}_n^{\mathrm{loc}}(C_z)$. The minimum number of codewords required for local fidelity is

$$M_{n,\mathrm{loc}}^*(\varepsilon) := \min\Big\{ |e_n(X^n)| :\ e_n : X^n \to Z_n \text{ deterministic}, \mathsf{D}_n^{\mathrm{loc}}(e_n) \le \varepsilon \Big\}. \tag{37}$$

Corollary VII.5. *For every $n \ge 1$ and every $\varepsilon \in [0,1]$, $M_{n,\mathrm{loc}}^*(\varepsilon) = \chi(G_\varepsilon^{\vee n})$.*

Thus, $n^{-1}\lceil \log_2 M_{n,\mathrm{loc}}^*(\varepsilon)\rceil$ is the blocklength-n local fixed-length cost per source symbol. For a finite graph G, let $\chi_f(G)$ denote its fractional chromatic number, defined by

$$\chi_f(G) := \min\left\{ \sum_I \lambda_I : \lambda_I \ge 0,\ \sum_{I\ni v} \lambda_I \ge 1 \text{ for every } v \in V(G) \right\}, \tag{38}$$

where the weights are indexed by the independent sets I of G.

Theorem VII.6 (Local fixed-length exponent)**.** *For every $\varepsilon \in [0,1]$,*

$$\lim_{n\to\infty} \frac{1}{n} \left\lceil \log_2 M_{n,\mathrm{loc}}^*(\varepsilon) \right\rceil = \log_2 \chi_f(G_\varepsilon). \tag{39}$$

Proof. For every finite graph G, the fractional chromatic number satisfies the classical disjunctive-power identity [35]

$$\chi_f(G) = \lim_{n\to\infty} \chi(G^{\vee n})^{1/n}. \tag{40}$$

Applying this with $G = G_\varepsilon$ and using Corollary VII.5 gives

$$\lim_{n\to\infty} \frac{1}{n} \log_2 M_{n,\mathrm{loc}}^*(\varepsilon) = \log_2 \chi_f(G_\varepsilon). \tag{41}$$

The outer ceiling changes the normalized quantity by at most $1/n$, and therefore does not affect the limit. □

The prior-weighted output-entropy objective introduced next is different in general. Let $\pi \in \Delta(X)$ and let $X^n \sim \pi^{\otimes n}$.

Definition VII.7. The minimum local block output entropy at level ε is

$$\mathsf{H}_{\pi,n}^{*,\mathrm{loc}}(\varepsilon) := \inf\Big\{ H(e_n(X^n)) :\ e_n : X^n \to Z_n, e_n \text{ deterministic}, \mathsf{D}_n^{\mathrm{loc}}(e_n) \le \varepsilon \Big\}. \tag{42}$$

Proposition VII.8. *For every $n \ge 1$ and every $\varepsilon \in [0,1]$, $\mathsf{H}_{\pi,n}^{*,\mathrm{loc}}(\varepsilon) = H_\chi(G_\varepsilon^{\vee n}, \pi^{\otimes n})$.*

Recall the graph-entropy functional. Let $\mathcal{I}(G)$ denote the independent sets of a finite graph G on vertex set X. For $I \subseteq X$, let $\mathbf{1}_I \in \{0,1\}^X$ be its incidence vector. Define

$$\mathrm{VP}(G) := \mathrm{co}\{\mathbf{1}_I : I \in \mathcal{I}(G)\} \subseteq [0,1]^X. \tag{43}$$

This is the vertex-packing polytope of G. For $\pi \in \Delta(X)$, its graph entropy is

$$H_G(\pi) := \min_{a\in \mathrm{VP}(G)} \sum_{x:\ \pi(x)>0} \pi(x) \log_2 \frac{1}{a_x}. \tag{44}$$

The objective is interpreted as $+\infty$ if $a_x = 0$ for some x with $\pi(x) > 0$.

Theorem VII.9 (Local entropy rate)**.** *For every prior $\pi \in \Delta(X)$ and every $\varepsilon \in [0,1]$,*

$$\lim_{n\to\infty} \frac{1}{n} \mathsf{H}_{\pi,n}^{*,\mathrm{loc}}(\varepsilon) = H_{G_\varepsilon}(\pi), \qquad \textit{or equivalently,} \qquad \lim_{n\to\infty} \frac{1}{n} H_\chi(G_\varepsilon^{\vee n}, \pi^{\otimes n}) = H_{G_\varepsilon}(\pi). \tag{45}$$

A. Prefix coding

Let $\mathsf{L}^*_{\pi,n}(\varepsilon)$ be the minimum expected length of a binary prefix code with a locally ε-faithful deterministic block encoder and lengths in $\mathbb{N}_0$; thus the empty word is available for a singleton output alphabet.

Proposition VII.10 (Prefix-code sandwich)**.** *For every $n \geq 1$,*

$$H_\chi(G_\varepsilon^{\vee n}, \pi^{\otimes n}) \leq \mathsf{L}^*_{\pi,n}(\varepsilon) \leq H_\chi(G_\varepsilon^{\vee n}, \pi^{\otimes n}) + 1. \tag{46}$$

Consequently,

$$\lim_{n\to\infty} \frac{1}{n} \mathsf{L}^*_{\pi,n}(\varepsilon) = H_{G_\varepsilon}(\pi). \tag{47}$$

Proof. A locally faithful encoder is a proper coloring of $G_\varepsilon^{\vee n}$. For each induced output Z, Shannon's prefix-code bounds place the optimal expected binary prefix length between $H(Z)$ and $H(Z)+1$. Minimizing over proper colorings proves the sandwich. Divide by n and use Theorem VII.9. □

Let $\mathsf{L}^{*,\mathrm{rand}}_{\pi,n}(\varepsilon)$ denote the minimum expected binary prefix length over stochastic block encoders Q_n whose nonempty support fibers are independent sets of $G_\varepsilon^{\vee n}$, again allowing the empty word for a singleton active alphabet.

Theorem VII.11 (Local Shannon coding theorem)**.** *For every $\pi \in \Delta(X)$ and $\varepsilon \in [0,1]$,*

$$\lim_{n\to\infty} \frac{1}{n} \mathsf{L}^*_{\pi,n}(\varepsilon) = \lim_{n\to\infty} \frac{1}{n} \mathsf{L}^{*,\mathrm{rand}}_{\pi,n}(\varepsilon) = H_{G_\varepsilon}(\pi). \tag{48}$$

Equivalently, every expected-prefix rate strictly above $H_{G_\varepsilon}(\pi)$ is achievable for all sufficiently large blocklengths, whereas every sequence of locally ε-faithful prefix codes satisfies

$$\liminf_{n\to\infty} \frac{1}{n} \mathbb{E}[\ell_n] \geq H_{G_\varepsilon}(\pi). \tag{49}$$

The same converse holds for support-locally faithful stochastic encoders. Moreover,

$$H_{G_\varepsilon}(\pi) = \min I(X;W), \tag{50}$$

where W ranges over independent-set-valued random variables satisfying $X \in W$ almost surely.

Proof. Proposition VII.10 and Theorem VII.9 establish the deterministic limit and the corresponding operational statements. Let us now consider a stochastic encoder Q_n with a fixed prefix code. For each x^n, select a shortest codeword in the support of $Q_n(\cdot \mid x^n)$. The resulting deterministic encoder has fibers contained in the corresponding support fibers of Q_n and therefore remains faithful. Moreover, its codeword length is no larger, pointwise, than the conditional expected length under Q_n. Hence randomization cannot improve the optimal expected prefix length at any blocklength. The mutual-information expression then follows from the classical single-letter characterization of graph entropy [18], [19]. □

VIII. Window Tests

Coordinatewise tests can miss incompatibilities that emerge only through interactions among multiple coordinates. To capture such effects, we next allow the observer to inspect arbitrary subsets of up to k coordinates, with unrestricted full-block observation as the limiting case.

Definition VIII.1. For $n \geq 1$, define the class of unrestricted block tests

$$\mathcal{F}_n^{\mathrm{glob}} := \{F : Y^n \to \mathbb{R} : \min F = 0,\ \max F \leq 1\}. \tag{51}$$

For $A, B \in \mathscr{K}(Y^n)$, define

$$d^{\mathrm{glob}}_{+,n}(A,B) := \sup_{F \in \mathcal{F}_n^{\mathrm{glob}}} \big(\Psi_B(F) - \Psi_A(F)\big), \text{ and } d^{\mathrm{glob}}_n(A,B) := \max\{d^{\mathrm{glob}}_{+,n}(A,B), d^{\mathrm{glob}}_{+,n}(B,A)\}. \tag{52}$$

For a nonempty cell $C \subseteq X^n$, define its unrestricted distortion as $\mathsf{D}^{\mathrm{glob}}_n(C) = \sup_{x^n \in C} d^{\mathrm{glob}}_{+,n}(K^{\otimes}_{x^n}, K^{\otimes}_C)$. Since X^n can itself be viewed as a finite input alphabet, the one-shot construction applies directly at blocklength n, yielding the corresponding unrestricted incompatibility graph and coding objectives.

Definition VIII.2. For $n \geq 1$ and $\varepsilon \in [0,1]$, the global block incompatibility graph $G^{\mathrm{glob}}_{n,\varepsilon}$ has vertex set X^n and joins distinct x^n, u^n iff $d^{\mathrm{glob}}_n(K^{\otimes}_{x^n}, K^{\otimes}_{u^n}) > \varepsilon$. For deterministic $e_n : X^n \to Z_n$ with fibers $C_z = e_n^{-1}(z)$, set $\mathsf{D}^{\mathrm{glob}}_n(e_n) = \sup_{z \in e_n(X^n)} \mathsf{D}^{\mathrm{glob}}_n(C_z)$. Let

$$M^*_{n,\mathrm{glob}}(\varepsilon) := \min\big\{|e_n(X^n)| : e_n : X^n \to Z_n \text{ deterministic}, \mathsf{D}^{\mathrm{glob}}_n(e_n) \leq \varepsilon\big\}. \tag{53}$$

If $X^n \sim \pi^{\otimes n}$, define the minimum global block output entropy by

$$\mathsf{H}^{*,\mathrm{glob}}_{\pi,n}(\varepsilon) := \inf\big\{H(e_n(X^n)) : e_n : X^n \to Z_n \text{ deterministic}, \mathsf{D}^{\mathrm{glob}}_n(e_n) \leq \varepsilon\big\}. \tag{54}$$

To interpolate between coordinatewise and global observation, we restrict each test to depend on at most k coordinates. The selected coordinates need not be contiguous, so k captures interaction order rather than geometric locality. This leaves k as the sole parameter controlling the observer scale.

Definition VIII.3 (Free-window k-local tests)**.** Let us set $n \geq 1$ and $k \geq 1$ and let us define

$$\mathcal{F}_n^{(k)} := \Big\{ F : Y^n \to \mathbb{R} : \min F = 0,\ \max F \leq 1,\ \exists S \subseteq [n] \text{ with } 1 \leq |S| \leq \min\{k, n\},\ F = F_S \circ \mathrm{pr}_S \Big\}, \tag{55}$$

where $\mathrm{pr}_S : Y^n \to Y^S$ is the canonical coordinate projection and $F_S : Y^S \to \mathbb{R}$. For $x^n \in X^n$, let $x_S = (x_i)_{i \in S}$, ordered increasingly. For $A, B \in \mathscr{K}(Y^n)$ define

$$d_{+,n}^{(k)}(A, B) := \sup_{F \in \mathcal{F}_n^{(k)}} \big(\Psi_B(F) - \Psi_A(F) \big), \text{ and } d_n^{(k)}(A, B) := \max\{ d_{+,n}^{(k)}(A, B), d_{+,n}^{(k)}(B, A) \}. \tag{56}$$

The k-local incompatibility graph on X^n is

$$G_{n,\varepsilon}^{(k)} : \qquad x^n \sim u^n \quad \Longleftrightarrow \quad d_n^{(k)}(K_{x^n}^{\otimes}, K_{u^n}^{\otimes}) > \varepsilon. \tag{57}$$

The graph $G_{n,\varepsilon}^{\mathrm{glob}}$ corresponds to unrestricted tests.

A test supported on S integrates out every coordinate outside S, reducing the block optimization to the largest unrestricted window distortion.

Lemma VIII.4 (Window reduction)**.** *For every $n, k \geq 1$ and every $x^n, u^n \in X^n$,*

$$d_{+,n}^{(k)}\big(K_{x^n}^{\otimes}, K_{u^n}^{\otimes}\big) = \max_{\substack{S \subseteq [n] \\ 1 \leq |S| \leq \min\{k,n\}}} d_{+,|S|}^{\mathrm{glob}}\big(K_{x_S}^{\otimes}, K_{u_S}^{\otimes}\big). \tag{58}$$

Consequently,

$$d_n^{(k)}\big(K_{x^n}^{\otimes}, K_{u^n}^{\otimes}\big) = \max_{\substack{S \subseteq [n] \\ 1 \leq |S| \leq \min\{k,n\}}} d_{|S|}^{\mathrm{glob}}\big(K_{x_S}^{\otimes}, K_{u_S}^{\otimes}\big). \tag{59}$$

Proof. Fix $S \subseteq [n]$ and $F = F_S \circ \mathrm{pr}_S$. For every generating product measure $p_1 \otimes \cdots \otimes p_n$ with $p_i \in K_{x_i}$,

$$\sum_{y^n \in Y^n} \Big(\prod_{i=1}^n p_i(y_i) \Big) e^{F_S(y_S)} = \sum_{y_S \in Y^S} \Big(\prod_{i \in S} p_i(y_i) \Big) e^{F_S(y_S)}, \tag{60}$$

because all coordinates outside S integrate to one. Taking suprema and using the convex-hull definition gives

$$\Psi_{K_{x^n}^{\otimes}}(F_S \circ \mathrm{pr}_S) = \Psi_{K_{x_S}^{\otimes}}(F_S), \qquad \Psi_{K_{u^n}^{\otimes}}(F_S \circ \mathrm{pr}_S) = \Psi_{K_{u_S}^{\otimes}}(F_S). \tag{61}$$

Taking the supremum over normalized F_S and the maximum over admissible windows proves the directed identity. Apply it in both directions for the symmetric identity. □

For nonempty $C \subseteq X^n$, let $\mathsf{D}_n^{(k)}(C) = \max_{x^n \in C} d_{+,n}^{(k)}(K_{x^n}^{\otimes}, K_C^{\otimes})$. For deterministic $e_n : X^n \to Z_n$ with fibers $C_z = e_n^{-1}(z)$, set $\mathsf{D}_n^{(k)}(e_n) = \max_{z \in e_n(X^n)} \mathsf{D}_n^{(k)}(C_z)$.

Proposition VIII.5 (Window coding law)**.** *For every nonempty $C \subseteq X^n$,*

$$\mathsf{D}_n^{(k)}(C) = \max_{x^n, u^n \in C} d_{+,n}^{(k)}\big(K_{x^n}^{\otimes}, K_{u^n}^{\otimes}\big). \tag{62}$$

Consequently, a deterministic encoder satisfies $\mathsf{D}_n^{(k)}(e_n) \leq \varepsilon$ if and only if it is a proper coloring of $G_{n,\varepsilon}^{(k)}$. In particular, the minimum codebook size and output entropy are respectively $\chi\big(G_{n,\varepsilon}^{(k)}\big)$ and $H_\chi\big(G_{n,\varepsilon}^{(k)}, \pi^{\otimes n}\big)$.

Proof. Theorem IV.6 gives $\Psi_{K_C^{\otimes}}(F) = \max_{u^n \in C} \Psi_{K_{u^n}^{\otimes}}(F)$. For fixed $x^n \in C$, interchange the finite maximum with the supremum over $F \in \mathcal{F}_n^{(k)}$ to obtain

$$d_{+,n}^{(k)}\big(K_{x^n}^{\otimes}, K_C^{\otimes}\big) = \max_{u^n \in C} d_{+,n}^{(k)}\big(K_{x^n}^{\otimes}, K_{u^n}^{\otimes}\big). \tag{63}$$

Maximizing over x^n proves the first assertion. Because both pair orientations occur, $\mathsf{D}_n^{(k)}(C) \leq \varepsilon$ iff $d_n^{(k)}(K_{x^n}^{\otimes}, K_{u^n}^{\otimes}) \leq \varepsilon$ for all $x^n, u^n \in C$. Thus faithful clusters are precisely the independent sets of $G_{n,\varepsilon}^{(k)}$, proving the coding identities. □

Corollary VIII.6 (Monotonicity in the observer window)**.** *For fixed n, x^n, u^n, the quantities*

$$d_n^{(1)}\big(K_{x^n}^{\otimes}, K_{u^n}^{\otimes}\big) \leq d_n^{(2)}\big(K_{x^n}^{\otimes}, K_{u^n}^{\otimes}\big) \leq \cdots \leq d_n^{(n)}\big(K_{x^n}^{\otimes}, K_{u^n}^{\otimes}\big) = d_n^{\mathrm{glob}}\big(K_{x^n}^{\otimes}, K_{u^n}^{\otimes}\big) \tag{64}$$

are nondecreasing. Hence the graphs satisfy

$$G_{n,\varepsilon}^{(1)} \subseteq G_{n,\varepsilon}^{(2)} \subseteq \cdots \subseteq G_{n,\varepsilon}^{(n)} = G_{n,\varepsilon}^{\mathrm{glob}}. \tag{65}$$

For every $k \geq n$, $d_n^{(k)} = d_n^{\mathrm{glob}}$ and $G_{n,\varepsilon}^{(k)} = G_{n,\varepsilon}^{\mathrm{glob}}$. Consequently, for fixed n, ε, and π, both the minimum codebook size

$$\chi\big(G_{n,\varepsilon}^{(k)}\big)$$

and the minimum output entropy

$$H_\chi\big(G_{n,\varepsilon}^{(k)}, \pi^{\otimes n}\big)$$

are nondecreasing in k.

Proof. The test classes are nested, or equivalently the maxima in Lemma VIII.4 are taken over larger collections of windows as k grows. If $k \geq n$, every normalized global test belongs to $\mathcal{F}_n^{(k)}$ by taking $S = [n]$, while the reverse inclusion is automatic. Hence $\mathcal{F}_n^{(k)} = \mathcal{F}_n^{\mathrm{glob}}$ for every $k \geq n$. □

Corollary VIII.7 (Recovery of the one-coordinate local theorem)**.** *For $k = 1$,*

$$d_n^{(1)}\big(K_{x^n}^{\otimes}, K_{u^n}^{\otimes}\big) = \max_{1 \leq i \leq n} d(K_{x_i}, K_{u_i}). \tag{66}$$

Therefore $G_{n,\varepsilon}^{(1)} = G_\varepsilon^{\vee n}$, where G_ε has edge $x \sim u$ iff $d(K_x, K_u) > \varepsilon$.

Proof. Apply Lemma VIII.4 with $|S| = 1$. □

IX. Entropy-Rate Regularization

We first fix the operational terminology. At blocklength n, define

$$r_{n,k}^{\mathrm{fl}}(\varepsilon) := \frac{1}{n}\left\lceil \log_2 \chi\big(G_{n,\varepsilon}^{(k)}\big) \right\rceil, \text{ and } r_{n,k}^{\mathrm{ent}}(\varepsilon, \pi) := \frac{1}{n} H_\chi\big(G_{n,\varepsilon}^{(k)}, \pi^{\otimes n}\big). \tag{67}$$

The first quantity is the fixed-length binary cost per source symbol and does not depend on π. The second is the minimum output entropy per source symbol and generally does. The minimum expected binary prefix length per source symbol differs from $r_{n,k}^{\mathrm{ent}}$ by at most $1/n$. We use $R_k(\varepsilon, \pi)$ exclusively for the limiting entropy rate, equivalently the limiting expected-prefix rate. The coordinatewise and unrestricted fixed-length exponents are treated separately; for a general fixed window $k > 1$, we do not establish existence of a fixed-length asymptotic limit. For coordinatewise tests, the block graphs are disjunctive powers of the one-shot graph, yielding a single-letter Körner graph-entropy rate. For $k > 1$, this product structure generally disappears: the block graphs need not be powers of a single graph. Nevertheless, the asymptotic entropy rate still exists and admits a regularized characterization.

As such, the main challenge is not the finite-block coding law (which remains chromatic entropy) but the absence of a graph product linking different blocklengths. To establish the limit, we pass from chromatic entropy to graph entropy, exploit its product properties, and then control the resulting gap back to proper colorings. For a finite probabilistic graph (G, P), let $H_G(P)$ denote Körner graph entropy and $H_\chi(G, P)$ the minimum entropy over proper colorings. Chromatic entropy has the direct coding interpretation, while graph entropy provides the structure needed to prove rate existence. The following bound shows that their gap is sublinear in blocklength.

Lemma IX.1 (Entropy-coloring gap)**.** *Let G be a finite graph and let P be a distribution on its vertices. Entropies in this lemma are measured in bits. Then*

$$H_G(P) \leq H_\chi(G, P) \leq H_G(P) + \log_2\big(H_G(P) + 1\big) + 4. \tag{68}$$

In particular, if the vertex set has size at most M, then

$$0 \leq H_\chi(G, P) - H_G(P) \leq \log_2(\log_2 M + 1) + 4. \tag{69}$$

Proof. The lower bound is standard [18], [19]. For the upper bound, use the classical characterization [18], [19] $H_G(P) = \inf I(X; W)$, where $X \sim P$, W is independent-set valued, and $X \in W$ almost surely. Choose W^* with $I(X; W^*) = H_G(P)$. The strong functional representation lemma [31] gives $Z \perp X$ and a deterministic g such that $W^* = g(X, Z)$ in distribution and

$$H(W^* \mid Z) \leq I(X; W^*) + \log_2\big(I(X; W^*) + 1\big) + 4. \tag{70}$$

For fixed z, let $c_z(x) = g(x, z)$. Since W^* is an independent set containing X almost surely and Z is independent of X, after discarding a null set of z's, each map c_z assigns every x in the support of P to an independent set containing x. On that support, every color class $\{x : c_z(x) = I\}$ is contained in the independent set I. Extend c_z to vertices x with $P(x) = 0$ by assigning each such vertex a fresh singleton color. The extended map is a proper coloring of all of G, and the extension does not change the distribution or entropy of $c_z(X)$. Since conditional entropy is the average of $H(c_z(X))$ over z, some value z_0 outside the null set satisfies $H(c_{z_0}(X)) \leq H(c_Z(X) \mid Z)$. Thus $H_\chi(G, P) \leq H(c_{z_0}(X)) \leq H(W^* \mid Z)$, which gives the upper bound. Finally, if $|V(G)| \leq M$, then $H_G(P) \leq H(P) \leq \log_2 M$, which gives the stated consequence. □

Since $G_{n,\varepsilon}^{(k)}$ has at most $|X|^n$ vertices, the gap is $O(\log n) = o(n)$. Superadditivity of graph entropy under concatenation then yields the limit.

Theorem IX.2 (Regularized entropy rate)**.** *Fix $k \geq 1$, $\varepsilon \in [0,1]$, and $\pi \in \Delta(X)$. Then the limit*

$$R_k(\varepsilon, \pi) := \lim_{n\to\infty} \frac{1}{n} H_\chi(G_{n,\varepsilon}^{(k)}, \pi^{\otimes n}) \tag{71}$$

exists and is given by the regularized graph-entropy formula

$$R_k(\varepsilon, \pi) = \sup_{n\geq 1} \frac{1}{n} H_{G_{n,\varepsilon}^{(k)}}(\pi^{\otimes n}). \tag{72}$$

Proof. Set $b_n = H_{G_{n,\varepsilon}^{(k)}}(\pi^{\otimes n})$ and $c_n = H_\chi(G_{n,\varepsilon}^{(k)}, \pi^{\otimes n})$. The sequence (b_n) is superadditive. If two blocks of lengths m and n are concatenated, any k-local incompatibility in one block or in the second block remains a k-local incompatibility in the concatenated block. Hence

$$G_{m+n,\varepsilon}^{(k)} \supseteq G_{m,\varepsilon}^{(k)} \vee G_{n,\varepsilon}^{(k)}, \tag{73}$$

where $\vee$ denotes the disjunctive, or OR, product of graphs. Graph entropy is monotone under adding edges and additive under the OR product with product distributions [18], [19]. Therefore

$$b_{m+n} = H_{G_{m+n,\varepsilon}^{(k)}}(\pi^{\otimes(m+n)}) \geq H_{G_{m,\varepsilon}^{(k)} \vee G_{n,\varepsilon}^{(k)}}(\pi^{\otimes m} \otimes \pi^{\otimes n}) = H_{G_{m,\varepsilon}^{(k)}}(\pi^{\otimes m}) + H_{G_{n,\varepsilon}^{(k)}}(\pi^{\otimes n}) = b_m + b_n. \tag{74}$$

Since $0 \leq b_n \leq H(\pi^{\otimes n}) \leq n \log_2 |X|$, Fekete's lemma [35] gives

$$\lim_{n\to\infty} \frac{b_n}{n} = \sup_{n\geq 1} \frac{b_n}{n}. \tag{75}$$

By Lemma IX.1,

$$0 \leq c_n - b_n \leq \log_2(n \log_2 |X| + 1) + 4 = o(n). \tag{76}$$

Therefore c_n/n has the same limit as b_n/n, and the claimed formula follows. □

Interpretation. For $k = 1$, the block graphs tensorize as OR powers of a single one-shot incompatibility graph, so the asymptotic rate reduces to a single-letter Körner graph entropy. For $k > 1$, no single graph need generate the block graphs across all blocklengths. Nevertheless, Theorem IX.2 shows that the operational entropy rate, equivalently the expected-prefix rate, still exists and is exactly the regularized graph-entropy quantity. The regularization is essential in general: as shown next, its supremum need not be attained at any finite blocklength.

A. Strict Regularization and Non-Attainment

The regularized formula need not reduce to any finite blocklength: an interaction may be invisible at one shot yet produce a positive entropy rate and therefore a positive expected-prefix rate.

Proposition IX.3 (Finite-block non-attainment)**.** *There exist two singleton credal sets, a window of size two, a distortion threshold, and a Bernoulli source such that the one-shot incompatibility graph is empty, the limiting intermediate entropy rate is positive, and the supremum in Theorem IX.2 is not attained at any finite blocklength.*

Proof. See Appendix G. □

The following collapse theorem states the two global operational limits; its proof is in Appendix F.

Theorem IX.4 (Global fixed-length and entropy laws)**.** *Let $\mathcal{C} = \{K_x : x \in X\} \subseteq \mathscr{K}(Y)$ be the set of distinct one-shot credal sets, $m = |\mathcal{C}|$, $\kappa(x) = K_x$, and $\bar{\pi} = \pi \circ \kappa^{-1}$. For blocks, write $\kappa^n(x^n) = (\kappa(x_1), \ldots, \kappa(x_n)) \in \mathcal{C}^n$. Fix $\varepsilon \in [0,1)$. Then there exist an integer $r_\varepsilon \geq 1$ and a polynomially growing function*

$$B_\varepsilon(n) := \sum_{j=0}^{\min\{r_\varepsilon - 1, n\}} \binom{n}{j} (m-1)^j. \tag{77}$$

such that every independent set of $G_{n,\varepsilon}^{\mathrm{glob}}$ has image under κ^n of cardinality at most $B_\varepsilon(n)$. Consequently, the fixed-length codebook size satisfies

$$\frac{m^n}{B_\varepsilon(n)} \leq M_{n,\mathrm{glob}}^*(\varepsilon) \leq m^n, \tag{78}$$

and its exponent is

$$\lim_{n\to\infty} \frac{1}{n} \log_2 M_{n,\mathrm{glob}}^*(\varepsilon) = \log_2 m. \tag{79}$$

Moreover, the minimum output entropy satisfies

$$nH(\bar{\pi}) - \log_2 B_\varepsilon(n) \le \mathsf{H}^{*,\mathrm{glob}}_{\pi,n}(\varepsilon) \le nH(\bar{\pi}), \tag{80}$$

and therefore its normalized entropy rate is

$$\lim_{n\to\infty} \frac{1}{n} \mathsf{H}^{*,\mathrm{glob}}_{\pi,n}(\varepsilon) = H(\bar{\pi}). \tag{81}$$

The two limits have different operational meanings. The first is the unrestricted fixed-length codebook exponent, $\log_2 m$; the second is the unrestricted entropy rate, equivalently the expected-prefix rate, $H(\bar{\pi})$. They are equal only when $\bar{\pi}$ is uniform on all m equality classes. In particular, the second limit must not be read as a fixed-length coding theorem. The restriction $\varepsilon < 1$ is essential: normalized distortion is at most one, so at $\varepsilon = 1$ every incompatibility graph is empty.

Corollary IX.5 (Expected-prefix interpretation)**.** *Let $\mathsf{L}^{*,(k),\mathrm{pfx}}_{\pi,n}(\varepsilon)$ be the minimum expected binary prefix length over all deterministic k-local ε-faithful block encoders $e_n : X^n \to Z_n$ and all prefix-free binary codes assigned to the active output alphabet $e_n(X^n)$. Then*

$$\lim_{n\to\infty} \frac{1}{n} \mathsf{L}^{*,(k),\mathrm{pfx}}_{\pi,n}(\varepsilon) = R_k(\varepsilon, \pi). \tag{82}$$

Proof. For every deterministic k-local ε-faithful encoder with output Z,

$$H(Z) \le \mathbb{E}[\ell(Z)] \le H(Z) + 1.$$

Minimizing over faithful encoders, dividing by n, and applying Theorem IX.2 proves the claim. □

X. Graph-Saturation Window

We now determine when bounded tests detect every unrestricted incompatibility by representing block pairs through finite profiles. A block pair has an untruncated ordered-pair profile

$$\mathsf{P}(x^n, u^n)_{ab} := \#\{i \in [n] : x_i = a,\ u_i = b\}, \qquad a, b \in X. \tag{83}$$

This records the multiplicity of each ordered symbol pair. A size-k window cannot distinguish counts above k.

Definition X.1 (Truncated pair profile)**.** For $x^n, u^n \in X^n$, define

$$T_k(x^n, u^n)_{ab} := \min\big\{k, \mathsf{P}(x^n, u^n)_{ab}\big\}, \qquad a, b \in X. \tag{84}$$

Thus $T_k(x^n, u^n) \in \{0, 1, \dots, k\}^{X\times X}$, with the coordinatewise order and truncated addition $(s \oplus t)_{ab} = \min\{k, s_{ab} + t_{ab}\}$.

Proposition X.2 (Finite-profile dependence)**.** *For fixed k and ε, adjacency in $G^{(k)}_{n,\varepsilon}$ depends only on $T_k(x^n, u^n)$. Specifically, there is an upward-closed $\mathcal{B}_{k,\varepsilon} \subseteq \{0, 1, \dots, k\}^{X\times X}$ such that*

$$x^n \sim u^n \text{ in } G^{(k)}_{n,\varepsilon} \quad \Longleftrightarrow \quad T_k(x^n, u^n) \in \mathcal{B}_{k,\varepsilon}. \tag{85}$$

Proof. By Lemma VIII.4, adjacency is equivalent to an $S \subseteq [n]$ with $|S| \le k$ such that

$$d^{\mathrm{glob}}_{|S|}\big(K^{\otimes}_{x_S}, K^{\otimes}_{u_S}\big) > \varepsilon. \tag{86}$$

The count vector $h \in \{0, 1, \dots, k\}^{X\times X}$ with $\sum_{a,b} h_{ab} = |S|$ encodes the ordered multiset of pairs on S. Up to coordinate permutation, h determines both product credal sets, and global distortion is permutation invariant. Taking the upward closure of all bad h of mass at most k gives $\mathcal{B}_{k,\varepsilon}$ and the stated equivalence. □

Removing truncation gives the global window. Let $e_{ab} \in \mathbb{N}_0^{X\times X}$ be the unit profile at (a, b). For nonzero $h \in \mathbb{N}_0^{X\times X}$, let

$$\|h\|_1 := \sum_{a,b} h_{ab}.$$

Choose any sequence

$$\big((a_j, b_j)\big)_{j=1}^{\|h\|_1}$$

containing exactly h_{ab} copies of each ordered pair (a, b), and define

$$A_h := \bigotimes_{j=1}^{\|h\|_1} K_{a_j}, \qquad B_h := \bigotimes_{j=1}^{\|h\|_1} K_{b_j}. \tag{87}$$

The same ordering is used in both products; different realizing sequences differ only by a common coordinate permutation and therefore yield the same global distortion. Define the global bad-profile set

$$\mathcal{B}^{\mathrm{glob}}_\varepsilon := \big\{h \in \mathbb{N}_0^{X\times X} \setminus \{0\} : d^{\mathrm{glob}}_{\|h\|_1}(A_h, B_h) > \varepsilon\big\}. \tag{88}$$

This set is coordinatewise upward closed.

Lemma X.3 (Upward closure)**.** *If $h \in \mathcal{B}_\varepsilon^{\mathrm{glob}}$ and $h' \geq h$ coordinatewise, then $h' \in \mathcal{B}_\varepsilon^{\mathrm{glob}}$.*

Proof. Lift a global test for h to h' by ignoring the added coordinates. Neither directed nor symmetric global distortion can decrease. □

By Dickson's lemma [32], every upward-closed subset of $\mathbb{N}_0^d$ has finitely many minimal elements. Let $\mathrm{Min}(\mathcal{B}_\varepsilon^{\mathrm{glob}})$ be this antichain and define

$$k_{\min}(\varepsilon) := \begin{cases} 1, & \mathcal{B}_\varepsilon^{\mathrm{glob}} = \varnothing, \\ \max\limits_{h \in \mathrm{Min}(\mathcal{B}_\varepsilon^{\mathrm{glob}})} \|h\|_1, & \mathcal{B}_\varepsilon^{\mathrm{glob}} \neq \varnothing. \end{cases} \tag{89}$$

If bad profiles exist, $k_{\min}(\varepsilon)$ is the largest size of a minimal witness. Any smaller window misses at least one such witness, while every global incompatibility contains one of size at most $k_{\min}(\varepsilon)$. If no bad profiles exist, all global graphs are empty and we set $k_{\min}(\varepsilon) = 1$. Minimality is crucial, since a large bad profile may contain a much smaller bad subprofile and therefore need not determine the required window size. The minimal profiles remove this redundancy and identify the irreducible witnesses of global incompatibility.

Theorem X.4 (Exact graph-saturation window)**.** *For every $\varepsilon \in [0,1]$, the smallest positive integer k such that*

$$G_{n,\varepsilon}^{(k)} = G_{n,\varepsilon}^{\mathrm{glob}} \qquad \textit{for every } n \geq 1 \tag{90}$$

is $k_{\min}(\varepsilon)$. Equivalently,

$$G_{n,\varepsilon}^{(k)} = G_{n,\varepsilon}^{\mathrm{glob}} \ \textit{ for all } n \quad \Longleftrightarrow \quad k \geq k_{\min}(\varepsilon). \tag{91}$$

Here $G_{n,\varepsilon}^{(k)}$ is defined for every $n, k \geq 1$ as in Definition VIII.3, so that windows have size at most $\min\{k, n\}$.

Proof. See Appendix D. □

Notice that this graph-level threshold is uniform over all blocklengths. At and beyond it, both the chromatic numbers and the prior-weighted chromatic entropies agree with their unrestricted counterparts at every n. Equality of the asymptotic entropy rates alone does not imply graph saturation.

Corollary X.5 (Entropy-rate endpoints)**.** *For $k = 1$, $R_1(\varepsilon, \pi) = H_{G_\varepsilon}(\pi)$, the Körner graph entropy of the one-shot incompatibility graph. For the unrestricted observer, define*

$$R_{\mathrm{glob}}(\varepsilon, \pi) := \lim_{n \to \infty} \frac{1}{n} \mathsf{H}_{\pi,n}^{*,\mathrm{glob}}(\varepsilon). \tag{92}$$

This limit exists: Theorem IX.4 covers $\varepsilon < 1$, while at $\varepsilon = 1$ every global incompatibility graph is empty. If $k \geq k_{\min}(\varepsilon)$, then

$$R_k(\varepsilon, \pi) = R_{\mathrm{glob}}(\varepsilon, \pi). \tag{93}$$

For every fixed $\varepsilon < 1$, the unrestricted endpoint is

$$R_{\mathrm{glob}}(\varepsilon, \pi) = H(\bar{\pi}), \tag{94}$$

where $\bar{\pi}$ is the pushforward of π under $x \mapsto K_x$, or equivalently the distribution of the one-shot credal equality class of X. At $\varepsilon = 1$, $R_1(1, \pi) = R_k(1, \pi) = R_{\mathrm{glob}}(1, \pi) = 0$ for every k; the equality with $H(\bar{\pi})$ is asserted only for $\varepsilon < 1$.

Proof. The case $k = 1$ is Corollary VIII.7 together with the classical OR-power graph-entropy theorem. If $k \geq k_{\min}(\varepsilon)$, Theorem X.4 makes the bounded-window and global graphs identical at every blocklength, so their entropy rates, and hence their expected-prefix rates, coincide. The stated global values follow from Theorem IX.4 for $\varepsilon < 1$ and from the distortion ceiling at $\varepsilon = 1$. □

The converse need not hold at the entropy-rate level. For example, if π is supported on a single input, then $R_k(\varepsilon, \pi) = R_{\mathrm{glob}}(\varepsilon, \pi) = 0$ for every k, even if the corresponding block graphs have not yet saturated. Thus, $k_{\min}(\varepsilon)$ is the smallest window that guarantees graph equality at every blocklength, not necessarily the smallest window that yields the same entropy rate for a particular prior. Minimal bad profiles establish that such a saturation window is finite, but they do not give an explicit numerical bound. To obtain one, we exploit repeated strict noninclusions between credal sets. Under repeated independent products, distinct credal sets become increasingly distinguishable: a separating score induces a threshold event whose probability approaches one under the replacing family and zero under the original family. This concentration argument yields an explicit bound on the window size required for global incompatibility.

Theorem X.6 (Quantitative global saturation)**.** *Let $K, L \in \mathscr{K}(Y)$ with $L \not\subseteq K$. Then*

$$d_{+,r}^{\mathrm{glob}}\big(K^{\otimes r}, L^{\otimes r}\big) \geq \log\big(1 + (e-1)(1 - e^{-cr})\big) - \log\big(1 + (e-1)e^{-cr}\big) \tag{95}$$

for every $r \geq 1$ and for some constant $c = c(K, L) > 0$. Consequently,

$$d_{+,r}^{\mathrm{glob}}\big(K^{\otimes r}, L^{\otimes r}\big) \longrightarrow 1. \tag{96}$$

More generally, for every distinct $K, L \in \mathscr{K}(Y)$, the symmetric global distortion converges to 1. In particular, for every $\varepsilon < 1$ there is a finite $r_{K,L}(\varepsilon)$ after which the corresponding directed distortion exceeds ε.

The directed conclusion requires $L \not\subseteq K$: if $L \subseteq K$, then $L^{\otimes r} \subseteq K^{\otimes r}$ and the distortion is zero. The normalized ceiling similarly forces $\varepsilon < 1$ in finite-window consequences.

Proof. Choose $q \in L \setminus K$. Strict separation gives $h : Y \to \mathbb{R}$, so $\sup_{p \in K} \langle p, h\rangle < \langle q, h\rangle$. Rescale so that $0 \leq h \leq 1$, and set

$$m_K := \sup_{p \in K} \langle p, h\rangle, \qquad \eta := \frac{\langle q, h\rangle - m_K}{2} > 0, \qquad t := m_K + \eta. \tag{97}$$

For $r \geq 1$, let

$$E_r := \left\{ y^r : \frac{1}{r} \sum_{j=1}^{r} h(y_j) \geq t \right\}, \tag{98}$$

and set $F_r = \mathbf{1}_{E_r} \in \mathcal{F}(Y^r)$. Hoeffding's inequality [33], applied under $q^{\otimes r} \in L^{\otimes r}$, gives

$$q^{\otimes r}(E_r) \geq 1 - e^{-2\eta^2 r}. \tag{99}$$

For any product $p_1 \otimes \cdots \otimes p_r$ with $p_j \in K$, the variables $h(Y_j)$ are independent, lie in $[0, 1]$, and have means at most m_K. Hence

$$(p_1 \otimes \cdots \otimes p_r)(E_r) \leq e^{-2\eta^2 r}. \tag{100}$$

Convexity extends the bound to all of $K^{\otimes r}$. Hence $\Psi_{L^{\otimes r}}(F_r) \geq \log\big(1 + (e-1)(1 - e^{-2\eta^2 r})\big)$, whereas $\Psi_{K^{\otimes r}}(F_r) \leq \log\big(1 + (e-1)e^{-2\eta^2 r}\big)$. Subtracting proves the bound with $c = 2\eta^2$. It tends to the distortion ceiling one, proving the directed limit. The symmetric limit follows in the orientation for which noninclusion holds. □

The saturation estimate also bounds the graph-saturation window. For $(a, b) \in X^2$, define the symmetric separation margin

$$\gamma_{ab} := \max\left\{ \max_{f : Y \to [0,1]} \big(S_{K_b}(f) - S_{K_a}(f)\big), \max_{f : Y \to [0,1]} \big(S_{K_a}(f) - S_{K_b}(f)\big) \right\}. \tag{101}$$

If $K_a \neq K_b$, then $\gamma_{ab} > 0$ by strong separation. For $\varepsilon \in [0, 1)$ and $K_a \neq K_b$, set

$$r_{ab}(\varepsilon) := 1 + \left\lfloor \frac{2}{\gamma_{ab}^2} \log \frac{(e-1)(1+e^{\varepsilon})}{e - e^{\varepsilon}} \right\rfloor. \tag{102}$$

As such, $r_{ab}(\varepsilon)$ repetitions suffice for global incompatibility at threshold ε.

Corollary X.7 (Effective graph-saturation bound)**.** *Let*

$$\mathcal{P} := \{(a, b) \in X^2 : K_a \neq K_b\}. \tag{103}$$

For every $\varepsilon \in [0, 1)$, if $\mathcal{P} = \varnothing$, then $k_{\min}(\varepsilon) = 1$. Otherwise,

$$k_{\min}(\varepsilon) \leq \max\left\{ \max_{(a,b) \in \mathcal{P}} r_{ab}(\varepsilon), \sum_{(a,b) \in \mathcal{P}} \big(r_{ab}(\varepsilon) - 1\big) \right\} \leq \sum_{(a,b) \in \mathcal{P}} r_{ab}(\varepsilon). \tag{104}$$

In particular, a bound valid without separating the degenerate case is

$$k_{\min}(\varepsilon) \leq \max\left\{ 1, \sum_{(a,b) : K_a \neq K_b} r_{ab}(\varepsilon) \right\}. \tag{105}$$

Proof. Fix $(a, b) \in \mathcal{P}$. Suppose the first maximum defining γ_{ab} is attained from K_a to K_b. Choose a maximizing $f : Y \to [0, 1]$ and $q \in K_b$ with $\langle q, f\rangle = S_{K_b}(f)$. In the proof of Theorem X.6, this gives $\eta = \gamma_{ab}/2$ and $c = \gamma_{ab}^2/2$. If the reverse direction attains the maximum, exchange K_a and K_b. Since the global distortion is symmetric, in either case

$$d_r^{\mathrm{glob}}(K_a^{\otimes r}, K_b^{\otimes r}) \geq \log\big(1 + (e-1)(1 - e^{-cr})\big) - \log\big(1 + (e-1)e^{-cr}\big). \tag{106}$$

The right-hand side exceeds ε iff $e^{-cr} < (e - e^{\varepsilon})/((e-1)(1+e^{\varepsilon}))$. The strict inequality holds for every $r \geq r_{ab}(\varepsilon)$ by (102). Consequently,

$$r_{ab}(\varepsilon) e_{ab} \in \mathcal{B}_{\varepsilon}^{\mathrm{glob}}. \tag{107}$$

Let $h \in \mathrm{Min}(\mathcal{B}_\varepsilon^{\mathrm{glob}})$. If $h_{ab} > 0$ with $K_a = K_b$, Lemma E.1 allows that coordinate to be removed without changing distortion, contradicting minimality. Hence $h_{ab} = 0$ for such pairs. If $h_{ab} \geq r_{ab}(\varepsilon)$ for some $(a,b) \in \mathcal{P}$, then (107) gives a bad profile $r_{ab}(\varepsilon)e_{ab} \leq h$. Minimality forces $h = r_{ab}(\varepsilon)e_{ab}$, hence $\|h\|_1 = r_{ab}(\varepsilon)$. Otherwise, $h_{ab} \leq r_{ab}(\varepsilon) - 1$ for every $(a,b) \in \mathcal{P}$, so

$$\|h\|_1 \leq \sum_{(a,b)\in\mathcal{P}} \big(r_{ab}(\varepsilon) - 1\big). \tag{108}$$

Taking the maximum over all minimal bad profiles proves (104). If $\mathcal{P} = \varnothing$, all symbol-indexed credal sets coincide, every global incompatibility graph is empty, and the defining convention gives $k_{\min}(\varepsilon) = 1$. □

The bound also gives an explicit scaling with the distortion tolerance. If $\mathcal{P} \neq \varnothing$, define

$$\gamma_* := \min_{(a,b)\in\mathcal{P}} \gamma_{ab} > 0. \tag{109}$$

Then

$$k_{\min}(\varepsilon) \leq |\mathcal{P}| \left(1 + \frac{2}{\gamma_*^2} \log \frac{(e-1)(1+e^\varepsilon)}{e - e^\varepsilon}\right). \tag{110}$$

Consequently, for a fixed finite credal family,

$$k_{\min}(\varepsilon) = O\left(\log \frac{1}{1-\varepsilon}\right), \qquad \varepsilon \uparrow 1. \tag{111}$$

For a finite nonempty set E and $p, q \in \Delta(E)$, write

$$\mathrm{TV}(p,q) = \frac{1}{2} \sum_{a\in E} |p(a) - q(a)|.$$

Remark X.8. *For $K, L \in \mathscr{K}(Y)$, the directed separation margin appearing above has the equivalent form*

$$\max_{f:Y\to[0,1]} \big(S_L(f) - S_K(f)\big) = \max_{q\in L} \min_{p\in K} \mathrm{TV}(q,p). \tag{112}$$

This follows from finite-dimensional minimax duality [34]. Hence the constants in Corollary X.7 are geometric and explicitly computable; for finitely generated credal polytopes, the inner problems are linear programs and it is enough to evaluate generators of the outer polytope.

Thus, any one-shot noninclusion and $\varepsilon < 1$ yield a bad repeated profile and an explicit bound on the largest minimal witness.

XI. Exact Intermediate Entropy Rate

We now construct a three-symbol singleton family whose two-coordinate entropy rate, equivalently its expected-prefix rate, lies strictly between the corresponding local and unrestricted rates. The argument has four steps. First, finite credal certificates identify the induced window-two graph as the staircase graph. Second, separator families give the upper bound $R_\star(U) \leq 2/3$. Third, an exact extremal result on the balanced type class yields the matching lower bound $R_\star(U) \geq 2/3$. Together, these establish the exact intermediate rate $R_\star(U) = 2/3$. The detailed combinatorial proofs are given in Appendix H.

For $q \in [0,1]$, write $h(q) = -q\log_2 q - (1-q)\log_2(1-q)$, with $0\log_2 0 = 0$.

A. Certified staircase instance

The parameters make every one-coordinate profile safe and only two oriented two-coordinate profiles bad, isolating an intermediate obstruction. Let $X = \{0,1,2\}$ and $Y = \{1,2,3\}$. Define singleton credal sets $K_i = \{p_i\}$ by

$$p_0 = \left(\frac{1}{2}, 0, \frac{1}{2}\right), \quad p_1 = \left(\frac{1}{10}, \frac{1}{3}, \frac{17}{30}\right), \quad p_2 = \left(\frac{31}{50}, \frac{11}{50}, \frac{4}{25}\right), \tag{113}$$

and set $\varepsilon_0 = 17/25$. For $x \in X^n$, define

$$Z_x := \{i : x_i = 0\}, \qquad O_x := \{i : x_i = 1\}, \qquad T_x := \{i : x_i = 2\}. \tag{114}$$

Let Γ_n^{st} be the graph on X^n in which x and u are adjacent if either

$$Z_x \cap O_u \neq \varnothing \quad \text{and} \quad O_x \cap T_u \neq \varnothing, \text{ or } O_x \cap Z_u \neq \varnothing \quad \text{and} \quad T_x \cap O_u \neq \varnothing. \tag{115}$$

Equivalently, the bad-profile ideal is generated upward by $e_{01} + e_{12}$ and $e_{10} + e_{21}$.

Lemma XI.1 (Certified antichain)**.** *For the singleton family above and threshold $\varepsilon_0 = 17/25$, every one-shot profile is safe. Among the weight-two multisets of off-diagonal ordered-pair types, the only bad oriented profiles are $e_{01} + e_{12}$ and $e_{10} + e_{21}$. Every other profile has symmetric distance below $0.670187 < 17/25$, whereas the two bad profiles have symmetric distance above $0.695273 > 17/25$.*

Proof. The finite-event calculation and complete interval table are given in Appendix H-A. □

Corollary XI.2 (Identification of the window-two graph)**.** *For every blocklength n, $G^{(2)}_{n,\varepsilon_0} = \Gamma^{\mathrm{st}}_n$.*

Proof. One-shot profiles are safe by Lemma XI.1. By Lemma E.1, adding a diagonal coordinate preserves that distance. Fully off-diagonal windows are bad only for the two listed generators, exactly the adjacency rule of Γ^{st}_n. □

The graph family Γ^{st}_n therefore determines the window-two behavior. Define the associated entropy rate

$$R_\star(\pi) := \lim_{n\to\infty} \frac{1}{n} \min\{H(c(X^n)) : c \text{ properly colors } \Gamma^{\mathrm{st}}_n\}, X^n \sim \pi^{\otimes n}. \tag{116}$$

By Corollary XI.2 and Theorem IX.2 applied at $k = 2$, the limit exists, equals $\sup_{n\geq 1} \frac{1}{n} H_{\Gamma^{\mathrm{st}}_n}(\pi^{\otimes n})$, and satisfies $R_2(\varepsilon_0, \pi) = R_\star(\pi)$. Notice that the numerical verification fixes the credal graph, after which the entropy-rate calculation is entirely combinatorial and does not depend on the coordinates of p_0, p_1, p_2.

B. Separator families

Large independent sets follow by separating positions occupied by adjacent symbols. For $A \subseteq [n]$, let $\mathcal{I}^{12}_A = \{x \in X^n : O_x \subseteq A,\ T_x \subseteq A^c\}$. If $x, u \in \mathcal{I}^{12}_A$, then $O_x, O_u \subseteq A$ and $T_x, T_u \subseteq A^c$. Hence, $O_x \cap T_u = \varnothing$ and $T_x \cap O_u = \varnothing$. Both generators require one of these intersections, so $\mathcal{I}^{12}_A$ is independent. Likewise, $\mathcal{I}^{01}_A = \{x \in X^n : Z_x \subseteq A,\ O_x \subseteq A^c\}$ is independent.

Proposition XI.3 (Separator upper bound)**.** *For the uniform law U on X, $R_\star(U) \leq \frac{2}{3}$.*

C. Exact extremum

A matching lower bound controls independent sets on the balanced type class, the dominant uniform-source type. The key step is a two-fiber freezing property. Let $\mathcal{T}_m = \{x \in \{0,1,2\}^{3m} : |Z_x| = |O_x| = |T_x| = m\}$. Let $\Gamma^{\mathrm{st}}[\mathcal{T}_m]$ denote the subgraph of $\Gamma^{\mathrm{st}}_{3m}$ induced by $\mathcal{T}_m$. For $A \subseteq [3m]$, the separator star $\mathcal{I}^{12}_A \cap \mathcal{T}_m$ has size

$$\binom{|A|}{m}\binom{3m-|A|}{m}. \tag{117}$$

Let $\mathsf{S}_m = \binom{\lfloor 3m/2 \rfloor}{m}\binom{\lceil 3m/2 \rceil}{m}$.

Lemma XI.4 (Two-fiber freezing)**.** *Let $I \subseteq \mathcal{T}_m$ be independent in $\Gamma^{\mathrm{st}}[\mathcal{T}_m]$. For an occupied O-set S, write $I_S := \{x \in I : O_x = S\}$. If S and R are distinct occupied O-sets and $A = S \setminus R$ and $B = R \setminus S$, then one of these alternatives holds:*

$$T_x \cap B = \varnothing \quad \forall x \in I_S, \qquad T_u \cap A = \varnothing \quad \forall u \in I_R, \tag{118}$$

$$B \subseteq T_x \quad \forall x \in I_S, \qquad A \subseteq T_u \quad \forall u \in I_R. \tag{119}$$

Theorem XI.5 (Exact staircase extremum)**.** *For every $m \geq 1$,*

$$\alpha(\Gamma^{\mathrm{st}}[\mathcal{T}_m]) = \mathsf{S}_m = \binom{\lfloor 3m/2 \rfloor}{m}\binom{\lceil 3m/2 \rceil}{m}. \tag{120}$$

Consequently,

$$\alpha(\Gamma^{\mathrm{st}}[\mathcal{T}_m]) = \Theta\left(\frac{2^{3mh(1/3)}}{m}\right). \tag{121}$$

All families attaining this value have the structure below.

Theorem XI.6 (Classification of staircase extremizers)**.** *Let $I \subseteq \mathcal{T}_m$ be independent in $\Gamma^{\mathrm{st}}[\mathcal{T}_m]$. Then $|I| = \mathsf{S}_m$ if and only if there is a set $A \subseteq [3m]$ with $|A| \in \{\lfloor 3m/2 \rfloor, \lceil 3m/2 \rceil\}$ such that either $I = \mathcal{I}^{12}_A \cap \mathcal{T}_m$ or $I = \mathcal{I}^{01}_A \cap \mathcal{T}_m$. Thus the maximum independent families are separator stars, up to coordinate permutations and the symbol reversal $0 \leftrightarrow 2$.*

Remark XI.7 (Stability caveat)**.** *The classification completes the Erdős–Ko–Rado-type equality case [36]–[39]. A stability statement toward only the balanced extremizers is false. Indeed, let $m = 2q$ and choose $A \subseteq [6q]$ with $|A| = 3q + 1$. Then*

$$\frac{|\mathcal{I}^{12}_A \cap \mathcal{T}_{2q}|}{\mathsf{S}_{2q}} = \frac{3q+1}{3q+3} \longrightarrow 1, \tag{122}$$

although its largest intersection with a balanced separator star of the same orientation is only a fraction $(q+1)/(3q+1) \to 1/3$ of its size; intersections with balanced stars of the opposite orientation are exponentially smaller. A meaningful stability theorem must therefore allow slightly unbalanced separator stars.

D. Exact entropy rate

The separator construction and balanced type-class extremum give matching entropy-rate bounds.

Proposition XI.8 (Balanced type-class bridge)**.** *The staircase extremal theorem implies* $R_\star(U) \geq \log_2 3 - h(1/3) = \frac{2}{3}$.

Theorem XI.9 (Exact intermediate entropy rate)**.** *For the singleton credal family above, threshold* $\varepsilon_0 = 17/25$*, window* $k = 2$*, and uniform source* U *on* $X = \{0, 1, 2\}$*,* $R_2(\varepsilon_0, U) = R_\star(U) = \frac{2}{3}$*. Moreover,* $0 = H_{G_{\varepsilon_0}}(U) < \frac{2}{3} < \log_2 3$.

Proof. Corollary XI.2 and Theorem IX.2 give $R_2(\varepsilon_0, U) = R_\star(U)$. Propositions XI.3 and XI.8 give the matching bounds $R_\star(U) \leq 2/3$ and $R_\star(U) \geq 2/3$. The one-shot graph is empty, while the three distinct singleton sets have unrestricted entropy rate, equivalently expected-prefix rate, $H(U) = \log_2 3$ by Corollary X.5. □

Notice that this theorem makes no claim about the fixed-length chromatic exponent for the window-two graph family.

Remark XI.10. *In Theorem G.1, the regularized entropy rate equals the global entropy-rate endpoint; here it is strictly smaller.*

Products of independent sets may require exponentially many admissible pieces, obstructing the single-letter product argument.

Theorem XI.11 (No subexponential product refinement)**.** *For the singleton credal family above and every* $m \geq 1$*, there are independent* $C_m, D_m \subseteq X^m$ *in* $G^{(2)}_{m,\varepsilon_0}$ *such that every partition of* $C_m \times D_m \subseteq X^{2m}$ *into independent sets of* $G^{(2)}_{2m,\varepsilon_0}$ *has at least* $\binom{m}{\lfloor m/2 \rfloor}$ *nonempty parts. Since this number is* $2^{m-o(m)}$*, even for credal-realizable finite-window graphs the Cartesian product of two independent families can require exponentially many independent refinements.*

Proof. See Appendix H-G. □

This obstruction explains why the bounded-window problem cannot in general be recovered from a simple product argument. For an OR-product graph, the Cartesian product of two independent sets is again independent. Here, concatenation can create new cross-block incompatibilities, and the theorem shows that repairing them may require exponentially many independent pieces. Thus, even a subexponential-refinement surrogate for exact tensorization fails in general.

XII. Experimental Result on SVHN Dataset

We examine the one-shot construction on predictive credal sets from the SVHN experiment of [40]. We select the first $N = 60$ samples in the experimental SVHN predictive array. For the SVHN experiment, let $Y = \{0, \ldots, 9\}$ denote the digit-label alphabet. For each sample (image) i, five ensemble members $m \in \{1, \ldots, 5\}$ evaluate three views $v \in \{1, 2, 3\}$, yielding credal likelihood

$$K_i = \mathrm{co}\{p_{i,v,m} : v \in \{1, 2, 3\},\ m \in \{1, \ldots, 5\}\} \subseteq \Delta(Y). \tag{123}$$

Hence, each empirical credal set K_i is generated by 15 distributions (5 ensembles for each of the 3 views). Let $\phi : \{1, \ldots, 15\} \to \{1, 2, 3\} \times \{1, \ldots, 5\}$ be any bijection. For each fixed i, we set $p_{i,r} := p_{i,v,m}$ if $\phi(r) = (v, m)$. For a target generator $q \in \Delta(Y)$, let $\boldsymbol{h} = (h_y)_{y=0}^9 \in \mathbb{R}^{10}$ and $s \in \mathbb{R}$. The exponential of its directed generator-to-set distance is the optimum of

$$\begin{aligned} \max_{\boldsymbol{h}, s} \quad & \sum_{y=0}^{9} q(y)\, h_y \\ \text{s.t.} \quad & \text{(i)} \quad \sum_{y=0}^{9} p_{i,r}(y)\, h_y \leq 1, \quad r \in \{1, \ldots, 15\}, \\ & \text{(ii)} \quad s \leq h_y \leq e \cdot s, \qquad y \in \{0, \ldots, 9\}, \\ & \text{(iii)} \quad h_y \geq 0,\ s \geq 0, \qquad y \in \{0, \ldots, 9\}. \end{aligned} \tag{124}$$

Let $V(q; K_i)$ denote the optimal value of the linear program in (124) when K_i is the source credal set. Then

$$\widehat{d}_+(K_i, K_j) := \max_{\substack{v \in \{1,2,3\} \\ m \in \{1, \ldots, 5\}}} \log V(p_{j,v,m}; K_i), \tag{125}$$

and

$$\widehat{d}_{ij} := \max\{\widehat{d}_+(K_i, K_j),\ \widehat{d}_+(K_j, K_i)\}. \tag{126}$$

Let $\widehat{d} = (\widehat{d}_{ij})_{i,j \in [N]}$ denote the resulting symmetric distance matrix. The graph $\widehat{G}_\varepsilon$ on $[N]$ joins distinct i, j iff $\widehat{d}_{ij} > \varepsilon$. For reporting, let U_N denote the uniform distribution on $[N]$, and let $c_\varepsilon : [N] \to [\chi(\widehat{G}_\varepsilon)]$ be the reported minimum-color coloring. If $n_r := |\{i : c_\varepsilon(i) = r\}|$, the reported-color entropy is

$$H_{\mathrm{rep}}(\varepsilon) := -\sum_{r : n_r > 0} \frac{n_r}{N} \log_2 \frac{n_r}{N}. \tag{127}$$

This is the entropy of the returned minimum-color solution under U_N; it is not asserted to equal the chromatic entropy.

Let $y_i \in \{0, \dots, 9\}$ denote the ground-truth digit label of input i. The purity of the reported coloring is

$$\mathrm{Purity}(\varepsilon) := \frac{1}{N} \sum_{r} \max_{a \in \{0,\dots,9\}} |\{i : c_\varepsilon(i) = r,\ y_i = a\}|. \tag{128}$$

We also report the edge density

$$\rho_\varepsilon := \frac{2|E(\widehat{G}_\varepsilon)|}{N(N-1)} \tag{129}$$

and the codeword-count reduction factor $N/\chi(\widehat{G}_\varepsilon)$, relative to assigning a distinct codeword to every input. We use a monotone reparameterization of the distortion tolerance $\tau := -\log_{10}(1-\varepsilon)$ to spread operating points near $\varepsilon = 1$ on the horizontal axis in the following plots.

A. Numerical certification

The reported distance matrix $\widehat{d}$ is computed using the HiGHS backend and the standard solver settings. As a numerical stability check, we recompute all 54,000 linear programs using the HiGHS dual-simplex solver, with primal and dual feasibility tolerances both set to 10^{-10} and a per-LP time limit of 30 seconds.

Let $\widehat{d}^{\,\mathrm{strict}} = (\widehat{d}^{\,\mathrm{strict}}_{ij})$ denote the resulting symmetric distance matrix. For each reported threshold ε, define

$$g_\varepsilon := \min_{1 \le i < j \le N} \left| \widehat{d}^{\,\mathrm{strict}}_{ij} - \varepsilon \right|. \tag{130}$$

TABLE II
SVHN OPERATING POINTS

ε	τ	$\chi(\widehat{G}_\varepsilon)$	N/χ	$H_{\mathrm{rep}}(\varepsilon)$	Purity(ε)	g_ε
0.959367	1.391	18	3.333	4.052	0.767	6.324×10^{-4}
0.988745	1.949	14	4.286	3.702	0.783	1.991×10^{-7}
0.996122	2.411	12	5.000	3.387	0.733	9.117×10^{-6}
0.998508	2.826	10	6.000	3.121	0.700	7.499×10^{-7}
0.999339	3.180	8	7.500	2.876	0.650	2.390×10^{-7}
0.999873	3.897	6	10.000	2.347	0.550	1.809×10^{-8}

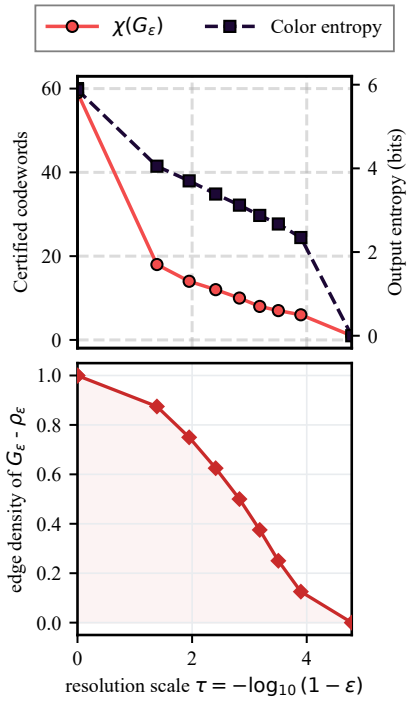

Fig. 4. Credal coloring for $N = 60$ SVHN predictive credal sets. Top: chromatic number and entropy of the reported minimum-color coloring. Bottom: edge density. The horizontal coordinate is $\tau = -\log_{10}(1-\varepsilon)$.

Table II reports the selected operating points, while Fig. 4 summarizes how the resulting incompatibility graph changes with the distortion tolerance.

In the top panel of Fig. 4, the red curve reports the chromatic number $\chi(\widehat{G}_\varepsilon)$, i.e., the minimum number of codewords required by the credal compatibility constraint. Complete-graph, complement-matching, clique, or exhaustive degree-of-saturation (DSATUR) searches [46] certify the exact chromatic number at each reported threshold. The exhaustive DSATUR procedure is therefore used as an exact chromatic-number solver rather than as a heuristic coloring procedure. The dashed curve reports $H_{\mathrm{rep}}(\varepsilon)$, the output entropy of the returned minimum-color coloring; this quantity is not claimed to equal the chromatic entropy $H_\chi(\widehat{G}_\varepsilon, U_N)$.

The bottom panel reports the edge density ρ_ε of $\widehat{G}_\varepsilon$. As ε increases, or equivalently as $\tau = -\log_{10}(1-\varepsilon)$ increases, incompatibility edges are removed, so ρ_ε decreases and the graph becomes progressively sparser. Correspondingly, the exact chromatic number decreases, showing that fewer codewords are required at larger distortion tolerances.

We use $\varepsilon = 0.988745$ as a representative operating point for the matched-cardinality comparison with the barycentric baseline below. At this threshold, the computed graph has $\chi = 14$, corresponding to a codeword-count reduction factor of $60/14 = 4.286$, with purity 0.783.

We separate numerical distance accuracy from combinatorial graph verification. Across all $\binom{N}{2} = 1770$ off-diagonal pairwise distances,

$$\max_{1 \le i < j \le N} \left| \widehat{d}^{\,\mathrm{strict}}_{ij} - \widehat{d}_{ij} \right| \le 1.45 \times 10^{-10}, \tag{131}$$

whereas across the six reported thresholds,

$$\min_{\varepsilon} g_\varepsilon = 1.81 \times 10^{-8}. \tag{132}$$

The stricter-tolerance recomputation moves no pairwise distance by more than 1.45×10^{-10}, while the closest any distance comes to any reported threshold is $g_\varepsilon \ge 1.81 \times 10^{-8}$, i.e., more than two orders of magnitude larger. Thus, the recomputation leaves every comparison $\widehat{d}_{ij} > \varepsilon$ unchanged, and hence produces the same incompatibility graph at all six thresholds.

The distance recomputation and the coloring checks address different numerical issues: the former verifies stability of the graph under stricter solver tolerances, while the latter establishes the reported chromatic number once the graph is fixed. These computations do not provide interval enclosures of every ideal real-arithmetic distance.

B. Barycentric baseline

To separate credal replacement from point reconstruction, we use the barycenter $\bar{p}_i = 15^{-1}\sum_{v=1}^{3}\sum_{m=1}^{5} p_{i,v,m}$ and matrix $D_{ij}^{\mathrm{TV}} = \mathrm{TV}(\bar{p}_i, \bar{p}_j)$. At each threshold ε, we set $k = \chi(\widehat{G}_\varepsilon)$ and solve

$$\min_{\substack{S\subseteq[N]\\|S|=k}} \max_{i\in[N]} \min_{j\in S} D_{ij}^{\mathrm{TV}}. \tag{133}$$

The implementation searches the sorted pairwise radii. A binary set-cover mixed-integer linear program (MILP) tests each candidate: every point must lie within the radius of a selected center, with exactly k centers. The MILPs are solved with the COIN-OR Branch-and-Cut (CBC) solver [7], using zero relative and absolute optimality gaps; only solutions returned with status `Optimal` are retained.

For any partition $\mathcal{C} = \{C_1, \dots, C_k\}$ of $[N]$, define its within-cluster credal violation rate at threshold ε by

$$V_\varepsilon(\mathcal{C}) := \frac{\sum_{r=1}^{k} \left|\{\{i,j\} \subseteq C_r : \widehat{d}_{ij} > \varepsilon\}\right|}{\sum_{r=1}^{k} \binom{|C_r|}{2}}, \tag{134}$$

whenever the denominator is positive, and set $V_\varepsilon(\mathcal{C}) = 0$ otherwise. Thus $V_\varepsilon(\mathcal{C})$ is the fraction of unordered within-cluster pairs that are edges of $\widehat{G}_\varepsilon$. For a partition $\mathcal{C} = \{C_1, \dots, C_k\}$ of $[N]$, define its barycentric medoid radius by

$$R_{\mathrm{TV}}(\mathcal{C}) := \max_{1\le r\le k} \min_{j\in C_r} \max_{i\in C_r} D_{ij}^{\mathrm{TV}}. \tag{135}$$

Let $\mathcal{C}_\varepsilon^{\mathrm{col}}$ denote the partition induced by the reported minimum coloring c_ε. For the barycentric baseline, let $S_\varepsilon \subseteq [N]$ denote the selected optimal k-center set, with $k = \chi(\widehat{G}_\varepsilon)$. Each selected center is assigned to its own cluster, and every remaining input $i \notin S_\varepsilon$ is assigned to

$$j^*(i) := \min \arg\min_{j\in S_\varepsilon} D_{ij}^{\mathrm{TV}}, \tag{136}$$

where the outer minimum specifies the smallest-index tie-breaking rule. The resulting nonempty assignment cells define $\mathcal{C}_\varepsilon^{\mathrm{bar}}$.

Fig. 5 compares the two partitions at a matched number of codewords or clusters, $k = \chi(\widehat{G}_\varepsilon)$. The coloring is optimized for credal compatibility, whereas the barycentric k-center solution is optimized for point-summary total-variation radius. At the representative operating point $\varepsilon = 0.988745$, for which $k = 14$, $R_{\mathrm{TV}}(\mathcal{C}_\varepsilon^{\mathrm{col}}) = 0.9629$, and $R_{\mathrm{TV}}(\mathcal{C}_\varepsilon^{\mathrm{bar}}) = 0.2990$, while $V_\varepsilon(\mathcal{C}_\varepsilon^{\mathrm{col}}) = 0$ and $V_\varepsilon(\mathcal{C}_\varepsilon^{\mathrm{bar}}) = 0.186$. Thus the barycentric solution attains the smaller point-summary radius, whereas the credal coloring enforces zero incompatibility violations. Since the two constructions optimize different fidelity criteria, these numbers do not define a universal ranking between the methods.

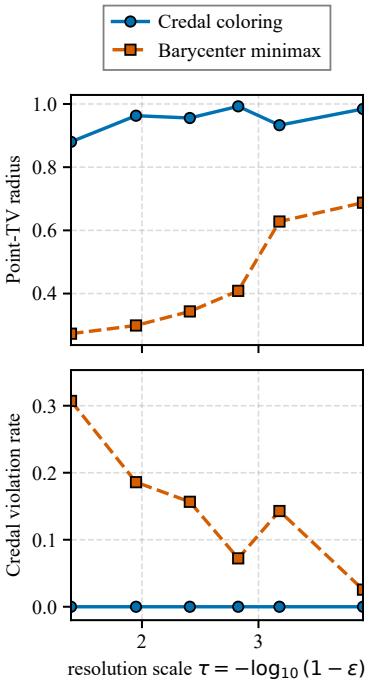

Fig. 5. Matched-cardinality comparison on the same $N = 60$ inputs with $k = \chi(\widehat{G}_\varepsilon)$. Top: barycentric medoid radius R_{TV} of the reported minimum-color credal partition and the optimal barycentric k-center partition. Bottom: within-cluster credal violation rate V_ε for the same two partitions.

XIII. Conclusion

We developed a rate–distortion framework in which each source symbol is associated with a credal set and compression replaces the credal sets within a cell by their convex hull. Under the support-only semantics considered here, whole-cell fidelity reduces exactly to pairwise compatibility. This turns the one-shot problem into graph coloring: the minimum codebook size is the chromatic number of the incompatibility graph, while the minimum output entropy is its chromatic entropy.

For product credal families, the asymptotic coding law depends on the scale at which replacement is observed. Under coordinatewise observation, the block graphs are disjunctive powers of the one-shot graph; the fixed-length exponent is $\log_2 \chi_f(G_\varepsilon)$ and the expected-prefix rate is $H_{G_\varepsilon}(\pi)$. For a fixed bounded window, the block graphs need not tensorize, but the expected-prefix rate still exists and is given by a regularized graph-entropy formula. Under unrestricted observation, every fixed tolerance $\varepsilon < 1$ has the same asymptotic endpoints: the fixed-length exponent is $\log_2 |\mathcal{C}|$, where $\mathcal{C}$ is the set of distinct credal sets, and the expected-prefix rate is $H(\bar{\pi})$.

Minimal bad profiles determine the smallest observation window for which bounded-window and unrestricted incompatibility graphs agree at every blocklength, and the concentration bound gives an explicit finite upper bound on this saturation window. The three-symbol staircase construction shows that the observer scales can remain genuinely distinct at the entropy-rate level, with coordinatewise, two-coordinate, and unrestricted expected-prefix rates equal to 0, $2/3$, and $\log_2 3$ bits per source symbol under the uniform law. The SVHN experiment illustrates, at a matched number of codewords, how credal-preserving compatibility and barycentric point-summary fidelity can lead to different partitions.

APPENDIX A
PRODUCT ADDITIVITY

The logarithm linearizes independent product aggregation.

Definition A.1 (Product functions)**.** Let Y_1 and Y_2 be finite nonempty sets. The independent product of credal sets is the strong extension defined in Section VII; it differs in general from the independent natural extension [30]. For positive $g_i : Y_i \to (0,\infty)$ and $f_i : Y_i \to \mathbb{R}$, define $(g_1 \otimes g_2)(y_1, y_2) = g_1(y_1)g_2(y_2)$ and $(f_1 \oplus f_2)(y_1, y_2) = f_1(y_1) + f_2(y_2)$.

Proposition A.2. *For every $K_1 \in \mathscr{K}(Y_1)$, $K_2 \in \mathscr{K}(Y_2)$, and positive functions $g_i : Y_i \to (0,\infty)$,*

$$S_{K_1\otimes K_2}(g_1 \otimes g_2) = S_{K_1}(g_1)S_{K_2}(g_2). \tag{137}$$

Consequently, for every $f_i : Y_i \to \mathbb{R}$,

$$\Psi_{K_1\otimes K_2}(f_1 \oplus f_2) = \Psi_{K_1}(f_1) + \Psi_{K_2}(f_2). \tag{138}$$

Proof. Let $A = \{p_1 \otimes p_2 : p_i \in K_i\}$. Since $K_1 \otimes K_2 = \mathrm{co}(A)$, the linear functional $\mu \mapsto \langle \mu, g_1 \otimes g_2\rangle$ may be maximized over A. Hence

$$S_{K_1\otimes K_2}(g_1 \otimes g_2) = \sup_{p_1\in K_1, p_2\in K_2} \langle p_1 \otimes p_2, g_1 \otimes g_2\rangle. \tag{139}$$

For product measures,

$$\langle p_1 \otimes p_2, g_1 \otimes g_2\rangle = \sum_{y_1\in Y_1}\sum_{y_2\in Y_2} p_1(y_1)p_2(y_2)g_1(y_1)g_2(y_2) = \langle p_1, g_1\rangle\langle p_2, g_2\rangle. \tag{140}$$

Therefore

$$S_{K_1\otimes K_2}(g_1 \otimes g_2) = \sup_{p_1\in K_1, p_2\in K_2} \langle p_1, g_1\rangle\langle p_2, g_2\rangle = \Big(\sup_{p_1\in K_1}\langle p_1, g_1\rangle\Big)\Big(\sup_{p_2\in K_2}\langle p_2, g_2\rangle\Big) = S_{K_1}(g_1)S_{K_2}(g_2). \tag{141}$$

Now let $f_i : Y_i \to \mathbb{R}$. Since $e^{f_1\oplus f_2} = e^{f_1} \otimes e^{f_2}$, Proposition IV.2 gives

$$\begin{aligned}\Psi_{K_1\otimes K_2}(f_1 \oplus f_2) &= \log S_{K_1\otimes K_2}(e^{f_1} \otimes e^{f_2}) = \log\big(S_{K_1}(e^{f_1})S_{K_2}(e^{f_2})\big)\\ &= \log S_{K_1}(e^{f_1}) + \log S_{K_2}(e^{f_2}) = \Psi_{K_1}(f_1) + \Psi_{K_2}(f_2).\end{aligned}$$

□

Theorem A.3 (Uniqueness of the logarithmic transform)**.** *Let $\varphi : (0,\infty) \to \mathbb{R}$ be strictly increasing and, for positive g, set $\Phi^{\varphi}_K(g) = \varphi(S_K(g))$. If, for every pair of finite alphabets, credal sets, and positive functions, $\Phi^{\varphi}_{K_1\otimes K_2}(g_1 \otimes g_2) = \Phi^{\varphi}_{K_1}(g_1) + \Phi^{\varphi}_{K_2}(g_2)$, then $\varphi(t) = c\log t, t > 0$ for some $c > 0$. Conversely, every positive multiple of the logarithm has this additivity property. The normalization $\varphi(e) = 1$ selects $\varphi(t) = \log t$.*

The universal additivity assumption is stronger than necessary: singleton alphabets already force the multiplicative Cauchy equation. Strict monotonicity excludes its pathological nonlinear solutions and forces the constant c to be positive.

Proof. For singleton alphabets, the support envelopes are arbitrary $a, b > 0$ and the product envelope is ab, so $\varphi(ab) = \varphi(a) + \varphi(b)$. In particular, $\varphi(1) = 0$. The function $u(x) := \varphi(e^x)$ is strictly increasing and satisfies Cauchy's equation $u(x + y) = u(x) + u(y)$ on $\mathbb{R}$. Every monotone additive function is linear, so $u(x) = cx$ for some $c \in \mathbb{R}$, and therefore $\varphi(t) = c\log t$. Strict monotonicity forces $c > 0$. The converse follows immediately from Proposition A.2, and the stated normalization gives $c = 1$. □

APPENDIX B
ENTROPIC DUALITY

The free-energy envelope also has a relative-entropy interpretation. The relative-entropy penalty below measures the least information cost of representing r by a law in K; its zero set therefore recovers K itself. It is unrelated to the two operational coding rates.

Proposition B.1. *For every $K \in \mathscr{K}(Y)$ and every $r \in \Delta(Y)$, the infimum defining $I_K(r)$ is attained, and*

$$I_K(r) = 0 \quad\Longleftrightarrow\quad r \in K. \tag{142}$$

Proof. Relative entropy is lower semicontinuous in (r, p) on the finite simplex [5], [42], so its restriction to the compact set K attains its infimum. Gibbs' inequality [5] gives $\mathrm{KL}(r\|p) \geq 0$, with equality exactly when $r = p$. □

Theorem B.2 (Variational representation)**.** *For every $K \in \mathscr{K}(Y)$ and every $f : Y \to \mathbb{R}$,*

$$\Psi_K(f) = \sup_{r\in\Delta(Y)} \big\{\langle r, f\rangle - I_K(r)\big\}. \tag{143}$$

Proof. The finite-alphabet Gibbs variational formula states that, for each $p \in \Delta(Y)$,

$$\log \sum_y p(y) e^{f(y)} = \sup_{r \in \Delta(Y)} \big\{ \langle r, f \rangle - \mathrm{KL}(r \| p) \big\}; \tag{144}$$

see, for example, [41], [42]. Taking the supremum over $p \in K$ and exchanging the order of two suprema gives

$$\Psi_K(f) = \sup_{r \in \Delta(Y)} \left\{ \langle r, f \rangle + \sup_{p \in K} \big(- \mathrm{KL}(r \| p) \big) \right\} = \sup_{r \in \Delta(Y)} \big\{ \langle r, f \rangle - I_K(r) \big\}. \tag{145}$$

□

Corollary B.3 (Order duality)**.** *For $K, L \in \mathscr{K}(Y)$, the following are equivalent:*

$$L \subseteq K, \qquad d_+(K, L) = 0, \qquad I_K(r) \le I_L(r) \quad \textit{for every } r \in \Delta(Y). \tag{146}$$

Proof. Inclusion $L \subseteq K$ directly gives $I_K \le I_L$, while Theorem IV.4 gives the equivalence of the first two statements. If $I_K \le I_L$, Theorem B.2 implies $\Psi_K(f) \ge \Psi_L(f)$ for every f. Thus $d_+(K, L) = 0$, and Theorem IV.4 yields $L \subseteq K$. □

APPENDIX C
DATA PROCESSING

A downstream channel can only discard information. A replacement distortion intended for inference should therefore not increase when both credal sets are passed through the same channel. Let T be a channel from finite Y to finite Y'. For $K \in \mathscr{K}(Y)$, set $TK := \{Tp : p \in K\} \in \mathscr{K}(Y')$ and $(Tp)(y') := \sum_{y \in Y} T(y' \mid y) p(y)$.

Proposition C.1 (Data processing)**.** *For the channel T above and all $K, L \in \mathscr{K}(Y)$, $d_+(TK, TL) \le d_+(K, L)$.*

Proof. The image TK is compact and convex. For $f \in \mathcal{F}(Y')$, set $g(y) = \sum_{y'} T(y' \mid y) e^{f(y')}$. Then $1 \le g(y) \le e$ for every y. Moreover

$$\Psi_{TK}(f) = \log S_K(g), \qquad \Psi_{TL}(f) = \log S_L(g). \tag{147}$$

Let $\ell = \log g$. Then $\ell - \min \ell \in \mathcal{F}(Y)$, and by translation covariance,

$$\Psi_{TL}(f) - \Psi_{TK}(f) = \Psi_L(\ell) - \Psi_K(\ell) = \Psi_L(\ell - \min \ell) - \Psi_K(\ell - \min \ell). \tag{148}$$

This is at most $d_+(K, L)$. Taking the supremum over $f \in \mathcal{F}(Y')$ proves the claim. □

Remark C.2. *The symmetric distance d contracts under channels. Consequently, output garbling can only remove edges from the incompatibility graph. Both the fixed-length chromatic costs and the prior-weighted chromatic-entropy costs are therefore monotone under output processing.*

APPENDIX D
LOCAL AND GRAPH-SATURATION PROOFS

A. Local tensorization

Proof. For (i), set $A = K_1 \otimes \cdots \otimes K_n$ and $B = L_1 \otimes \cdots \otimes L_n$. Every $F \in \mathcal{F}_n^{\mathrm{loc}}$ has the form $F = \Gamma_i f$. Since $e^{\Gamma_i f} = \mathbf{1} \otimes \cdots \otimes e^f \otimes \cdots \otimes \mathbf{1}$ and $S_{K_j}(\mathbf{1}) = 1$, Proposition A.2 gives $S_A(e^{\Gamma_i f}) = S_{K_i}(e^f)$. Thus

$$\Psi_A(\Gamma_i f) = \Psi_{K_i}(f), \qquad \Psi_B(\Gamma_i f) = \Psi_{L_i}(f). \tag{149}$$

It follows that $\Psi_B(F) - \Psi_A(F) \le \max_j d_+(K_j, L_j)$, and taking the supremum over F proves the upper bound. Conversely, fixing i and taking the supremum over all $F = \Gamma_i f$ gives $d_{+,n}^{\mathrm{loc}}(A, B) \ge d_+(K_i, L_i)$. Maximizing over i proves the directed formula. Applying it in both directions proves the symmetric formula.

For (ii), fix $x^n \in C$. Theorem IV.6 and the cylinder identity above give, for $F = \Gamma_i f$,

$$\Psi_{K_C^{\otimes}}(F) - \Psi_{K_{x^n}^{\otimes}}(F) = \max_{u^n \in C} \Psi_{K_{u_i}}(f) - \Psi_{K_{x_i}}(f). \tag{150}$$

The finiteness of C allows the supremum and maximum to be interchanged, so

$$\begin{aligned} d_{+,n}^{\mathrm{loc}}(K_{x^n}^{\otimes}, K_C^{\otimes}) &= \max_{1 \le i \le n} \max_{u^n \in C} \sup_{f \in \mathcal{F}} \big(\Psi_{K_{u_i}}(f) - \Psi_{K_{x_i}}(f) \big) \\ &= \max_{u^n \in C} \max_{1 \le i \le n} d_+(K_{x_i}, K_{u_i}). \end{aligned}$$

Maximizing over $x^n \in C$ gives the cluster formula.

For (iii), the cluster formula and both orientations of each pair show that $\mathsf{D}_n^{\mathrm{loc}}(C) \le \varepsilon$ iff $\max_i d(K_{x_i}, K_{u_i}) \le \varepsilon$ for all $x^n, u^n \in C$. By definition of the OR power, this is exactly independence of C in $G_\varepsilon^{\vee n}$. □

B. Local chromatic number

Proof. By Theorem VII.3(iii), locally ε-faithful encoders are precisely proper colorings of $G_\varepsilon^{\vee n}$. Minimizing codewords therefore minimizes colors. □

C. Local chromatic entropy

Proof. The same equivalence identifies minimum encoder-output entropy with the chromatic entropy of $(G_\varepsilon^{\vee n}, \pi^{\otimes n})$. □

D. Local entropy rate

Proof. Proposition VII.8 reduces the claim to the classical graph-entropy theorem for OR powers [18], [19]. □

E. Graph-saturation window

Proof. First suppose that $\mathcal{B}_\varepsilon^{\mathrm{glob}} = \varnothing$. If a pair of blocks were globally adjacent, its nonzero full ordered-pair profile would belong to $\mathcal{B}_\varepsilon^{\mathrm{glob}}$, a contradiction. Hence $G_{n,\varepsilon}^{\mathrm{glob}}$ is empty for every n. Since $G_{n,\varepsilon}^{(k)} \subseteq G_{n,\varepsilon}^{\mathrm{glob}}$ for every $n, k \geq 1$, all the k-local graphs are empty as well. Thus equality holds for every positive integer k, and the smallest such integer is $1 = k_{\min}(\varepsilon)$.

Now suppose that $\mathcal{B}_\varepsilon^{\mathrm{glob}} \neq \varnothing$. Assume that $k \geq k_{\min}(\varepsilon)$ and let x^n, u^n be globally adjacent. Their full ordered-pair profile $H := \mathsf{P}(x^n, u^n) \in \mathbb{N}_0^{X \times X} \setminus \{0\}$ belongs to $\mathcal{B}_\varepsilon^{\mathrm{glob}}$. Since this set is upward closed, H dominates some minimal bad profile $h \in \mathrm{Min}(\mathcal{B}_\varepsilon^{\mathrm{glob}})$. By definition, $\|h\|_1 \leq k$. Therefore one can choose at most k coordinates of the block pair whose profile is exactly h. This subwindow is already globally bad, hence the pair is adjacent in $G_{n,\varepsilon}^{(k)}$ by Lemma VIII.4. The reverse inclusion $G_{n,\varepsilon}^{(k)} \subseteq G_{n,\varepsilon}^{\mathrm{glob}}$ is immediate because $\mathcal{F}_n^{(k)} \subseteq \mathcal{F}_n^{\mathrm{glob}}$.

Conversely, suppose that $1 \leq k < k_{\min}(\varepsilon)$. Choose a minimal bad profile $h \in \mathrm{Min}(\mathcal{B}_\varepsilon^{\mathrm{glob}})$ with $\|h\|_1 > k$. Construct two blocks whose full ordered-pair profile is exactly h. They are globally adjacent. However every nonempty subwindow of size at most k has profile strictly dominated by h, and therefore cannot be bad by minimality of h. Hence this pair is not adjacent in the k-local graph. Thus equality with the global graph fails. □

APPENDIX E
SAFE COORDINATES

Finite global witnesses rely on two facts: a common coordinate preserves global distortion, and a directed-safe coordinate cannot increase it. Here the alphabets are generic and the blocklength subscript is suppressed. For credal sets A, B on a common finite alphabet E, write

$$d_+^{\mathrm{glob}}(A, B) := \sup_{f \in \mathcal{F}(E)} \big(\Psi_B(f) - \Psi_A(f)\big), \qquad d^{\mathrm{glob}}(A, B) := \max\{d_+^{\mathrm{glob}}(A, B), d_+^{\mathrm{glob}}(B, A)\}. \tag{151}$$

Lemma E.1 (Common-coordinate invariance)**.** *Let $A, B \in \mathscr{K}(U)$ and $C \in \mathscr{K}(V)$, where U, V are finite alphabets. Then*

$$d_+^{\mathrm{glob}}(A \otimes C, B \otimes C) = d_+^{\mathrm{glob}}(A, B). \tag{152}$$

Consequently, $d^{\mathrm{glob}}(A \otimes C, B \otimes C) = d^{\mathrm{glob}}(A, B)$.

Proof. The lower bound follows by testing only the U coordinate. For the upper bound, let $0 \leq F \leq 1$, set $G = e^F$, and for $r \in C$ define $g_r(u) = \sum_v r(v) G(u, v)$. Then $1 \leq g_r \leq e$. Let $\delta = d_+^{\mathrm{glob}}(A, B)$. Applying the definition of δ to $\log g_r$ after subtracting its minimum gives

$$S_B(g_r) \leq e^\delta S_A(g_r), \qquad r \in C. \tag{153}$$

Indeed, for $m_r = \min_u g_r(u)$ and $f_r = \log(g_r/m_r) \in \mathcal{F}(U)$, $\log S_B(g_r) - \log S_A(g_r) = \Psi_B(f_r) - \Psi_A(f_r) \leq \delta$. Taking suprema over $r \in C$ gives

$$S_{B \otimes C}(G) = \sup_{r \in C} S_B(g_r) \leq e^\delta \sup_{r \in C} S_A(g_r) = e^\delta S_{A \otimes C}(G). \tag{154}$$

Taking logarithms and the supremum over F gives the upper bound. Exchange A and B for the symmetric statement. □

Lemma E.2 (Directed-safe coordinates do not help)**.** *Let $A, B \in \mathscr{K}(U)$ and $C, D \in \mathscr{K}(V)$. If $D \subseteq C$, then*

$$d_+^{\mathrm{glob}}(A \otimes C, B \otimes D) \leq d_+^{\mathrm{glob}}(A, B). \tag{155}$$

Proof. Let $0 \leq F \leq 1$, set $G = e^F$, and for $r \in D$ let $g_r(u) = \sum_v r(v) G(u, v)$. If $\delta = d_+^{\mathrm{glob}}(A, B)$, then, as in the proof of Lemma E.1,

$$S_B(g_r) \leq e^\delta S_A(g_r), \qquad r \in D. \tag{156}$$

Hence

$$S_{B \otimes D}(G) = \sup_{r \in D} S_B(g_r) \leq e^\delta \sup_{r \in D} S_A(g_r) \leq e^\delta \sup_{s \in C} S_A(g_s) = e^\delta S_{A \otimes C}(G). \tag{157}$$

Taking logarithms and then the supremum over F proves the claim. □

Corollary E.3 (Removing directed-safe coordinates)**.** *Let $K_i, L_i \in \mathscr{K}(Y)$ for $1 \le i \le n$ and*

$$J := \{i \in [n] : L_i \not\subseteq K_i\}.$$

Then

$$d_+^{\mathrm{glob}}\Big(\bigotimes_{i=1}^n K_i, \bigotimes_{i=1}^n L_i\Big) \le d_+^{\mathrm{glob}}\Big(\bigotimes_{i\in J} K_i, \bigotimes_{i\in J} L_i\Big), \tag{158}$$

with the convention that the right-hand side is 0 when $J = \varnothing$.

Proof. Apply Lemma E.2 one directed-safe coordinate at a time. □

APPENDIX F
GLOBAL TESTS

Global block distortion allows arbitrary normalized observables on Y^n.

Since $\mathcal{F}_n^{\mathrm{loc}} \subseteq \mathcal{F}_n^{\mathrm{glob}}$, for every $A, B \in \mathscr{K}(Y^n)$ one has

$$d_{+,n}^{\mathrm{loc}}(A,B) \le d_{+,n}^{\mathrm{glob}}(A,B), \qquad d_n^{\mathrm{loc}}(A,B) \le d_n^{\mathrm{glob}}(A,B). \tag{159}$$

Moreover, because Y^n is again a finite alphabet, Theorem IV.4 applies verbatim at block level.

Proposition F.1 (Global cluster distortion)**.** *For every nonempty cluster $C \subseteq X^n$ and every $x^n \in C$,*

$$d_{+,n}^{\mathrm{glob}}(K_{x^n}^{\otimes}, K_C^{\otimes}) = \max_{u^n \in C} d_{+,n}^{\mathrm{glob}}(K_{x^n}^{\otimes}, K_{u^n}^{\otimes}). \tag{160}$$

Consequently, the global cluster distortion defined in Section VIII satisfies

$$\mathsf{D}_n^{\mathrm{glob}}(C) = \max_{x^n, u^n \in C} d_{+,n}^{\mathrm{glob}}(K_{x^n}^{\otimes}, K_{u^n}^{\otimes}). \tag{161}$$

Proof. Theorem IV.6 gives $\Psi_{K_C^{\otimes}}(F) = \max_{u^n \in C} \Psi_{K_{u^n}^{\otimes}}(F)$. Hence, for fixed $x^n \in C$,

$$\begin{aligned} d_{+,n}^{\mathrm{glob}}(K_{x^n}^{\otimes}, K_C^{\otimes}) &= \sup_{F \in \mathcal{F}_n^{\mathrm{glob}}} \Big(\Psi_{K_C^{\otimes}}(F) - \Psi_{K_{x^n}^{\otimes}}(F)\Big) \\ &= \sup_{F \in \mathcal{F}_n^{\mathrm{glob}}} \max_{u^n \in C} \Big(\Psi_{K_{u^n}^{\otimes}}(F) - \Psi_{K_{x^n}^{\otimes}}(F)\Big). \end{aligned}$$

Since C is finite, the supremum and maximum commute. Maximizing over $x^n \in C$ gives the second identity. □

Proposition F.2 (Finite-n global coloring laws)**.** *For every $n \ge 1$ and every $\varepsilon \in [0,1]$,*

$$M_{n,\mathrm{glob}}^*(\varepsilon) = \chi(G_{n,\varepsilon}^{\mathrm{glob}}), \qquad \mathsf{H}_{\pi,n}^{*,\mathrm{glob}}(\varepsilon) = H_\chi(G_{n,\varepsilon}^{\mathrm{glob}}, \pi^{\otimes n}). \tag{162}$$

Proof. By Proposition F.1, a deterministic block encoder is globally ε-faithful if and only if each fiber is an independent set of $G_{n,\varepsilon}^{\mathrm{glob}}$. Thus globally ε-faithful deterministic block encoders are exactly proper colorings of $G_{n,\varepsilon}^{\mathrm{glob}}$. Minimizing the number of codewords gives the chromatic number, while minimizing the entropy of the encoder output gives the corresponding chromatic entropy. □

Unlike $G_\varepsilon^{\vee n}$, the unrestricted graph $G_{n,\varepsilon}^{\mathrm{glob}}$ need not reduce to a one-shot OR power.

A. Global collapse

Proof. If $m = 1$, take $r_\varepsilon = 1$ and $B_\varepsilon(n) = 1$; the constant encoder makes all bounds equalities. Assume $m \ge 2$. For each of the finitely many (x,u) with $K_u \not\subseteq K_x$, choose $q_{x,u} \in K_u$ and $h_{x,u} : Y \to [0,1]$ such that

$$\langle q_{x,u}, h_{x,u}\rangle \ge \sup_{p \in K_x} \langle p, h_{x,u}\rangle + 2\eta_{x,u} \tag{163}$$

for some $\eta_{x,u} > 0$, and let $\eta_* = \min\{\eta_{x,u} : K_u \not\subseteq K_x\} > 0$.

Fix x^n, u^n and set $I(x^n,u^n) = \{i : K_{u_i} \not\subseteq K_{x_i}\}$ and $m_\to(x^n,u^n) = |I(x^n,u^n)|$. Write $J = I(x^n,u^n)$ and $r = |J|$, and assume $r > 0$. For $i \in J$, set $a_i = \sup_{p \in K_{x_i}} \langle p, h_{x_i,u_i}\rangle$ and $b_i = \langle q_{x_i,u_i}, h_{x_i,u_i}\rangle$, so that $b_i \ge a_i + 2\eta_*$. Define the event

$$E_{x^n,u^n} := \left\{ y^n \in Y^n : \frac{1}{r}\sum_{i\in J} h_{x_i,u_i}(y_i) \ge \frac{1}{r}\sum_{i\in J}(a_i + \eta_*) \right\}, \tag{164}$$

and let $F_{x^n,u^n} := \mathbf{1}_{E_{x^n,u^n}} \in \mathcal{F}_n^{\mathrm{glob}}$.

Choose $\nu \in K_{u^n}^{\otimes}$ using q_{x_i,u_i} on J and arbitrary admissible laws elsewhere. Under ν, the independent $Z_i = h_{x_i,u_i}(Y_i) \in [0,1]$ have means at least $a_i + 2\eta_*$. Hence Hoeffding's inequality [33] gives

$$\nu(E^c_{x^n,u^n}) \le e^{-2\eta_*^2 r}. \tag{165}$$

Hence

$$\Psi_{K_{u^n}^{\otimes}}(F_{x^n,u^n}) \ge \log\bigl(1 + (e-1)(1 - e^{-2\eta_*^2 r})\bigr). \tag{166}$$

On the other hand, for every generating product measure

$$\mu = p_1 \otimes \cdots \otimes p_n, \qquad p_i \in K_{x_i},$$

the same variables are independent, bounded in $[0,1]$, and have means at most a_i. A second application of Hoeffding's inequality gives

$$\mu(E_{x^n,u^n}) \le e^{-2\eta_*^2 r}. \tag{167}$$

Since $K_{x^n}^{\otimes}$ is the convex hull of these generating product measures, the same bound holds for every $\mu \in K_{x^n}^{\otimes}$, so

$$\Psi_{K_{x^n}^{\otimes}}(F_{x^n,u^n}) \le \log\bigl(1 + (e-1)e^{-2\eta_*^2 r}\bigr). \tag{168}$$

Therefore, with

$$\phi_*(r) := \log\bigl(1 + (e-1)(1 - e^{-2\eta_*^2 r})\bigr) - \log\bigl(1 + (e-1)e^{-2\eta_*^2 r}\bigr), \qquad r \in \mathbb{N}, \tag{169}$$

The function ϕ_* is nondecreasing and tends to one. Consequently,

$$d_{+,n}^{\mathrm{glob}}(K_{x^n}^{\otimes}, K_{u^n}^{\otimes}) \ge \phi_*(r). \tag{170}$$

Let $\rho(x^n,u^n) = |\{i : K_{x_i} \ne K_{u_i}\}|$, the Hamming distance between the class sequences. At each such coordinate, at least one ordered noninclusion holds. Hence

$$\max\{m_{\to}(x^n,u^n), m_{\to}(u^n,x^n)\} \ge \frac{\rho(x^n,u^n)}{2}. \tag{171}$$

It follows that

$$d_n^{\mathrm{glob}}(K_{x^n}^{\otimes}, K_{u^n}^{\otimes}) \ge \phi_*\Bigl(\Bigl\lceil \frac{\rho(x^n,u^n)}{2} \Bigr\rceil\Bigr). \tag{172}$$

Choose $r_\varepsilon \ge 1$ with $\phi_*(\lceil r_\varepsilon/2 \rceil) > \varepsilon$. Then whenever $\rho(x^n,u^n) \ge r_\varepsilon$, the pair x^n, u^n is adjacent in $G_{n,\varepsilon}^{\mathrm{glob}}$. Therefore, if $I \subseteq X^n$ is an independent set of $G_{n,\varepsilon}^{\mathrm{glob}}$, the class sequences in $\kappa^n(I)$ have pairwise Hamming distance strictly less than r_ε. Fix $x^n \in I$. Then every element of $\kappa^n(I)$ lies in the Hamming ball of radius $r_\varepsilon - 1$ around $\kappa^n(x^n)$ in the m-ary cube $\mathcal{C}^n$. Hence

$$|\kappa^n(I)| \le B_\varepsilon(n) := \sum_{j=0}^{\min\{r_\varepsilon - 1, n\}} \binom{n}{j} (m-1)^j. \tag{173}$$

This proves the independent-set bound.

For the codebook bound, the encoder that outputs the class sequence $\kappa^n(x^n)$ is globally 0-faithful, hence globally ε-faithful, and uses exactly m^n codewords. This gives $M^*_{n,\mathrm{glob}}(\varepsilon) \le m^n$. Conversely, every color class in a proper coloring of $G_{n,\varepsilon}^{\mathrm{glob}}$ contains at most $B_\varepsilon(n)$ distinct class sequences. Since there are m^n class sequences in total,

$$M^*_{n,\mathrm{glob}}(\varepsilon) \ge \frac{m^n}{B_\varepsilon(n)}. \tag{174}$$

Polynomial growth of $B_\varepsilon(n)$ gives the codebook-rate limit.

For the entropy bound, let $Z_n = e_n(X^n)$ be any globally ε-faithful deterministic block encoder, and write $C_n := \kappa^n(X^n)$. Conditional on $Z_n = z$, the class sequence C_n can take at most $B_\varepsilon(n)$ values, because each fiber of e_n is an independent set of $G_{n,\varepsilon}^{\mathrm{glob}}$. Hence

$$H(C_n \mid Z_n) \le \log_2 B_\varepsilon(n). \tag{175}$$

Since $C_n \sim \bar{\pi}^{\otimes n}$,

$$H(Z_n) \ge H(C_n) - H(C_n \mid Z_n) \ge nH(\bar{\pi}) - \log_2 B_\varepsilon(n). \tag{176}$$

Taking the infimum gives the theorem's entropy lower bound. Encoding $\kappa^n(X^n)$ itself gives the upper bound $nH(\bar{\pi})$; division by n and polynomial growth of $B_\varepsilon(n)$ prove the limit. □

Remark F.3 (Interpretation)**.** *Unlike the local OR-power law, global testing yields no new graph-entropy functional. At every fixed tolerance below one, it distinguishes blocks whose credal profiles differ in linearly many coordinates. Only the sequence of one-shot equality classes remains asymptotically relevant. The rigidity comes from concentration across every directed noninclusion, not from a product law for the global incompatibility graphs.*

APPENDIX G
A NON-ATTAINMENT EXAMPLE

The following singleton construction gives strict regularization and non-attainment. Its finite-event comparisons use the rational verifier from Lemma XI.1.

Theorem G.1 (Finite-block non-attainment)**.** *There exist finite alphabets, two singleton credal sets, a threshold ε, and a source law π such that, for $k = 2$,*

$$R_2(\varepsilon, \pi) = \sup_{n\geq 1} \frac{1}{n} H_{G^{(2)}_{n,\varepsilon}}(\pi^{\otimes n}) > H_{G^{(2)}_{1,\varepsilon}}(\pi). \tag{177}$$

Moreover, the supremum is not attained at any finite blocklength.

Proof. Take $k = 2$, $X = \{0, 1\}$, $Y = \{1, 2, 3\}$, $K_0 = \{p_0\}$, and $K_1 = \{p_1\}$, where

$$p_0 = (0.47, 0.26, 0.27), \qquad p_1 = (0.10, 0.17, 0.73). \tag{178}$$

Let $\varepsilon = 11/20$. For laws P, Q on finite E, the singleton directed distortion has the finite-event formula

$$d_+(\{P\}, \{Q\}) = \max_{A\subseteq E} \log \frac{1 + (e-1)Q(A)}{1 + (e-1)P(A)}. \tag{179}$$

The rational finite-event calculation described in Lemma XI.1 gives the outward-rounded enclosures

$$\begin{aligned}
d(\{p_0\}, \{p_1\}) &\in \big[0.43343074818367933317934326 6, \\
&\qquad 0.43343074818367933317934326 7\big] < \frac{11}{20}, \\
d(\{p_0^{\otimes 2}\}, \{p_1^{\otimes 2}\}) &\in \big[0.53880959502921841858376474 6, \\
&\qquad 0.53880959502921841858376474 7\big] < \frac{11}{20}, \\
d(\{p_0 \otimes p_1\}, \{p_1 \otimes p_0\}) &\in \big[0.56233334815263768908652081 7, \\
&\qquad 0.56233334815263768908652081 8\big] > \frac{11}{20}.
\end{aligned}$$

Before decimal presentation, each interval has width below 10^{-111}; the verifier checks all events of Y and Y^2 in both directions. Hence $G^{(2)}_{1,\varepsilon}$ is empty, while the mixed profile $e_{01} + e_{10}$ is bad.

It remains to identify the whole n-block graph. Equip $\{0, 1\}^n$ with the coordinatewise order

$$x^n \leq u^n \quad \Longleftrightarrow \quad x_i \leq u_i \text{ for every } i \in [n]. \tag{180}$$

If x^n and u^n are incomparable, some coordinates have types $(0, 1)$ and $(1, 0)$. The window $\{i, j\}$ therefore realizes the bad profile $e_{01} + e_{10}$, so the two words are adjacent in $G^{(2)}_{n,\varepsilon}$.

Conversely, suppose that x^n and u^n are comparable. Then all off-diagonal coordinates have the same orientation. A two-coordinate window containing one diagonal and one off-diagonal coordinate is safe by Lemma E.1, since the diagonal coordinate contributes the same credal factor on both sides. If both coordinates are off-diagonal, the window profile is either $2e_{01}$ or $2e_{10}$, both of which are safe at threshold ε. Hence no two-coordinate window is bad. Therefore $G^{(2)}_{n,\varepsilon}$ is exactly the incomparability graph of the Boolean lattice.

Let $X \sim \text{Bernoulli}(q)$ with $q \in (0, 1)$, and let π be its law. Write $h(q) := -q \log_2 q - (1-q) \log_2(1-q)$. Every independent set in the incomparability graph is a chain. Every chain in the Boolean lattice has size at most $n + 1$. Thus, for any proper coloring Z,

$$H(X^n \mid Z) \leq \log_2(n + 1), \tag{181}$$

and hence

$$H(Z) = H(X^n) - H(X^n \mid Z) \geq nh(q) - \log_2(n + 1). \tag{182}$$

Taking the infimum over all proper colorings gives

$$H_\chi(G^{(2)}_{n,\varepsilon}, \pi^{\otimes n}) \geq nh(q) - \log_2(n + 1). \tag{183}$$

The trivial coloring by the whole sequence gives the reverse bound

$$H_\chi(G^{(2)}_{n,\varepsilon}, \pi^{\otimes n}) \leq nh(q). \tag{184}$$

Therefore

$$\lim_{n\to\infty} \frac{1}{n} H_\chi(G^{(2)}_{n,\varepsilon}, \pi^{\otimes n}) = h(q). \tag{185}$$

TABLE III
INTERVAL-CERTIFIED ONE-SHOT AND WEIGHT-TWO SYMMETRIC PROFILE DISTANCES FOR THE INTERMEDIATE INSTANCE

Profile	Certified interval	Relation at ε_0
01	$[0.46154942821784841363081129 1,\ 0.46154942821784841363081129 2]$	compatible
02	$[0.37722708101226479564932353 8,\ 0.37722708101226479564932353 9]$	compatible
12	$[0.56672723315324930999937420 4,\ 0.56672723315324930999937420 5]$	compatible
$01+12$	$[0.69527313031536221505976225 2,\ 0.69527313031536221505976225 3]$	incompatible (h)
$10+21$	$[0.69527313031536221505976225 2,\ 0.69527313031536221505976225 3]$	incompatible $(h^\top)$
$01+01$	$[0.67018608862500123444417562 7,\ 0.67018608862500123444417562 8]$	compatible
$10+10$	$[0.67018608862500123444417562 7,\ 0.67018608862500123444417562 8]$	compatible
$01+21$	$[0.65429471232620938950021135 1,\ 0.65429471232620938950021135 2]$	compatible
$10+12$	$[0.65429471232620938950021135 1,\ 0.65429471232620938950021135 2]$	compatible
$12+12$	$[0.65041134694504736636069384 1,\ 0.65041134694504736636069384 2]$	compatible
$21+21$	$[0.65041134694504736636069384 1,\ 0.65041134694504736636069384 2]$	compatible
$01+10$	$[0.64009313695242465383908271 5,\ 0.64009313695242465383908271 6]$	compatible
$02+12$	$[0.63396719265956773433314582 8,\ 0.63396719265956773433314582 9]$	compatible
$20+21$	$[0.63396719265956773433314582 8,\ 0.63396719265956773433314582 9]$	compatible
$12+21$	$[0.63033953589152574762227666 1,\ 0.63033953589152574762227666 2]$	compatible
$02+21$	$[0.60937405054272588194792231 6,\ 0.60937405054272588194792231 7]$	compatible
$12+20$	$[0.60937405054272588194792231 6,\ 0.60937405054272588194792231 7]$	compatible
$01+02$	$[0.60145684392407334446160593 8,\ 0.60145684392407334446160593 9]$	compatible
$10+20$	$[0.60145684392407334446160593 8,\ 0.60145684392407334446160593 9]$	compatible
$01+20$	$[0.59371530488068459417928957 8,\ 0.59371530488068459417928957 9]$	compatible
$02+10$	$[0.59371530488068459417928957 8,\ 0.59371530488068459417928957 9]$	compatible
$02+20$	$[0.51866415821120658597517084 4,\ 0.51866415821120658597517084 5]$	compatible
$02+02$	$[0.51454619227398434789673364 0,\ 0.51454619227398434789673364 1]$	compatible
$20+20$	$[0.51454619227398434789673364 0,\ 0.51454619227398434789673364 1]$	compatible

By Theorem IX.2, $R_2(\varepsilon,\pi) = h(q)$. Since $K_0 \neq K_1$, the one-shot credal equality classes are the two singleton symbols, so in this example $\bar{\pi} = \pi$ and the same value also equals the global equality-class entropy $H(\bar{\pi})$. The point of the example is therefore not the absence of a closed form, but the failure of one-shot and finite-block attainment.

Since $G^{(2)}_{1,\varepsilon}$ is empty, $H_{G^{(2)}_{1,\varepsilon}}(\pi) = 0 < h(q) = R_2(\varepsilon,\pi)$.

For finite n, the graph is not complete: 0^n and $10\cdots0$ are comparable and have positive probability. Merging them gives a proper coloring with entropy strictly below that of the identity coloring:

$$H_\chi(G^{(2)}_{n,\varepsilon}, \pi^{\otimes n}) < nh(q). \tag{186}$$

Consequently,

$$H_{G^{(2)}_{n,\varepsilon}}(\pi^{\otimes n}) < nh(q) \tag{187}$$

for every finite n, while the normalized supremum equals $h(q)$. Hence the supremum is not achieved at any finite blocklength.

Thus non-attainment is not caused by an unknown limiting value: the limit is the elementary binary entropy $h(q)$. It arises because every finite Boolean lattice still contains comparable positive-probability words that can be merged. □

APPENDIX H
STAIRCASE ENTROPY-RATE PROOFS

A. Finite-Event Certificates

This calculation records the numerical separation that identifies the staircase graph. In Table III, the string ab abbreviates the unit profile e_{ab}, and sums of strings abbreviate sums of unit profiles. Set $h = e_{01} + e_{12}$ and $h^\top = e_{10} + e_{21}$.

Proof of Lemma XI.1. Each comparison reduces to a finite calculation. For probability laws P, Q on a finite set E, exponentiating the normalized test function and using scale invariance reduces the optimization to $g \in [1,e]^E$. With all other coordinates fixed, the ratio

$$\frac{\sum_y Q(y)g(y)}{\sum_y P(y)g(y)}$$

is a fractional-linear function of one coordinate, so its derivative has constant sign. Iterating over the coordinates shows that a maximizer is a vertex of the box. The vertices are $g = 1 + (e-1)\mathbf{1}_A$, $A \subseteq E$, and hence

$$d_+(\{P\},\{Q\}) = \max_{A\subseteq E} \log \frac{1+(e-1)Q(A)}{1+(e-1)P(A)}. \tag{188}$$

For each profile, the product laws and event masses in (188) are rational. The verifier enumerates all events of Y and Y^2 in both directions using rational arithmetic. Its transcendental inputs are enclosed as follows. If $S_N = \sum_{j=0}^N 1/j!$, the exponential series and a geometric majorant give

$$S_N < e < S_N + \frac{1}{(N+1)!}\frac{N+2}{N+1}. \tag{189}$$

For $x \geq 1$ and $z = (x-1)/(x+1)$,

$$\log x = 2 \sum_{j=0}^{M-1} \frac{z^{2j+1}}{2j+1} + R_M, \qquad 0 < R_M < \frac{2z^{2M+1}}{(2M+1)(1-z^2)}. \tag{190}$$

With $N = 75$ and $M = 115$, monotone propagation through every event ratio encloses each table entry within width 10^{-111}. Comparing interval endpoints with $17/25$ proves that only the two displayed oriented profiles are bad. □

B. Separator upper bound

Proof. Let $\delta > 0$ and draw $M_n = \lceil n 2^{(2/3+2\delta)n} \rceil$ independent random subsets of $[n]$, including each coordinate with probability $1/2$. A word x lies in $\mathcal{I}_{A_j}^{12}$ with probability $2^{-(|O_x|+|T_x|)}$. Let $\mathcal{A}_n = \{x : |O_x| + |T_x| \leq (2/3+\delta)n\}$. For every $x \in \mathcal{A}_n$, the probability that no A_j covers x is at most

$$\left(1 - 2^{-(2/3+\delta)n}\right)^{M_n} \leq \exp(-n 2^{\delta n}). \tag{191}$$

A union bound over at most 3^n words gives probability tending to zero that some word in $\mathcal{A}_n$ is uncovered. Fix a realization covering $\mathcal{A}_n$.

Use colors $1, \ldots, M_n$ for the least-indexed separator that covers a word in $\mathcal{A}_n$, and give all words outside $\mathcal{A}_n$ a fallback injective color. The separator color classes are independent, and singleton fallback classes are independent. If $X^n \sim U^{\otimes n}$ and C_n is the resulting color, while B_n indicates the fallback event, then

$$H(C_n) \leq H(B_n) + \log_2 M_n + \mathbb{P}(B_n = 1) n \log_2 3. \tag{192}$$

Since $|O_{X^n}| + |T_{X^n}| \sim \mathrm{Bin}(n, 2/3)$, Hoeffding's inequality [33] gives $\mathbb{P}(B_n = 1) \leq \exp(-2\delta^2 n)$. Therefore

$$\limsup_{n\to\infty} \frac{1}{n} H(C_n) \leq \frac{2}{3} + 2\delta. \tag{193}$$

Letting $\delta \downarrow 0$ gives $R_\star(U) \leq 2/3$. □

C. Two-fibre freezing

Proof. Let $x \in I_S$ and $u \in I_R$. Since $O_x = S$ and $O_u = R$, with $A = S \setminus R$ and $B = R \setminus S$, one has

$$Z_x \cap O_u = B \setminus T_x, \qquad T_x \cap O_u = B \cap T_x,$$
$$O_x \cap T_u = A \cap T_u, \qquad O_x \cap Z_u = A \setminus T_u.$$

The pair (x, u) is adjacent if either

$$(B \setminus T_x) \neq \varnothing \quad \text{and} \quad (A \cap T_u) \neq \varnothing, \tag{194}$$

or

$$(A \setminus T_u) \neq \varnothing \quad \text{and} \quad (B \cap T_x) \neq \varnothing. \tag{195}$$

Since I is independent, neither event occurs. If $T_x \cap B$ were a nonempty proper subset of B, avoiding both conditions would require $A \cap T_u = \varnothing$ and $A \subseteq T_u$, impossible because $A \neq \varnothing$. Thus $T_x \cap B \in \{\varnothing, B\}$.

If $T_x \cap B = \varnothing$, then the adjacency condition involving $B \setminus T_x$ forces $A \cap T_u = \varnothing$ for every $u \in I_R$. If $B \subseteq T_x$, then the second adjacency condition forces $A \subseteq T_u$ for every $u \in I_R$. Thus the type is determined by x. If two elements $x, x' \in I_S$ had different types, then every $u \in I_R$ would satisfy both $A \cap T_u = \varnothing$ and $A \subseteq T_u$, again impossible. Hence the same alternative holds uniformly over both fibres. □

D. Exact staircase extremum

Proof. For the lower bound, the separator star $\{x \in \mathcal{T}_m : O_x \subseteq A,\ T_x \subseteq A^c\}$ is independent and has size $\binom{|A|}{m}\binom{3m-|A|}{m}$. Maximizing over $|A|$ gives S_m.

For the upper bound, let $I \subseteq \mathcal{T}_m$ be independent, write $I_S = \{x \in I : O_x = S\}$, and choose an occupied S maximizing $|I_S|$. For every other occupied R, put $B_R = R \setminus S$. By Lemma XI.4, the pivot fibre either avoids or contains each B_R. Define

$$F := \bigcup_{R:\, T_x \cap B_R = \varnothing \ \forall x \in I_S} B_R, \qquad P := \bigcup_{R:\, B_R \subseteq T_x \ \forall x \in I_S} B_R. \tag{196}$$

Every T_x in the pivot fibre contains P and avoids F, so $P \cap F = \varnothing$. Put $p = |P|$ and $f = |F|$. Both are at most m, and

$$|I_S| \leq \binom{2m - p - f}{m - p}. \tag{197}$$

Every occupied include-type fibre satisfies $R \subseteq S \cup P$, and every occupied avoid-type fibre satisfies $R \subseteq S \cup F$. Since $P \cap F = \varnothing$, these two families of m-sets intersect only at S. Hence the number of occupied fibres is at most

$$\binom{m+p}{m} + \binom{m+f}{m} - 1. \tag{198}$$

The maximality of I_S gives

$$\begin{aligned} |I| &\le B_m(p,f), \\ B_m(p,f) &:= \left(\binom{m+p}{m} + \binom{m+f}{m} - 1 \right) \binom{2m-p-f}{m-p}. \end{aligned} \tag{199}$$

The expression is symmetric in p and f, since $\binom{2m-p-f}{m-p} = \binom{2m-p-f}{m-f}$. It is therefore enough to assume $p \le f$.

For $m \ge 4$, put

$$A_t := \binom{m+t}{m}, \qquad F_m(t) := A_t \binom{2m-t}{m}, \qquad 0 \le t \le m. \tag{200}$$

The successive ratio is

$$\frac{F_m(t+1)}{F_m(t)} = \frac{(m+t+1)(m-t)}{(t+1)(2m-t)}. \tag{201}$$

The numerator minus the denominator of the last fraction is $m(m-2t-1)$. Hence the integer maximum of F_m is attained at the integer or integers nearest to $m/2$, and

$$\max_{0 \le t \le m} F_m(t) = \mathsf{S}_m. \tag{202}$$

If $p = 0$, then $B_m(0,f) = F_m(f) \le \mathsf{S}_m$. Assume $p \ge 1$. For $2 \le f \le m-2$, the discrete convexity of A_t and $A_0 = 1$ give

$$A_p - 1 \le \frac{p}{f}(A_f - 1). \tag{203}$$

Moreover,

$$\frac{\binom{2m-p-f}{m-p}}{\binom{2m-f}{m}} = \prod_{i=0}^{p-1} \frac{m-i}{2m-f-i} \le \left(\frac{m}{2m-f} \right)^p. \tag{204}$$

It follows from Bernoulli's inequality that

$$\begin{aligned} B_m(p,f) &\le F_m(f) \left(1 + \frac{p}{f}\right) \left(\frac{m}{2m-f} \right)^p \\ &\le F_m(f) \left(\frac{m(f+1)}{f(2m-f)} \right)^p \le F_m(f) \le \mathsf{S}_m. \end{aligned} \tag{205}$$

The penultimate inequality uses $f(m-f) \ge 2(m-2) \ge m$.

For the boundary values, if $f = 1$, then $p = 1$ and

$$B_m(1,1) = (2m+1)\binom{2m-2}{m-1} \le F_m(2) \le \mathsf{S}_m, \tag{206}$$

where the first inequality is equivalent to $2m(2m+1) \le (m-1)(m+1)(m+2)$. If $f = m$, then

$$B_m(p,m) < 2A_m \le F_m(2) \le \mathsf{S}_m, \tag{207}$$

where the middle inequality follows from $(m-1)(m+1)(m+2) \ge 8(2m-1)$. Finally, if $f = m-1$ and $m \ge 6$, then

$$B_m(p,m-1) \le 2mA_{m-1} \le F_m(2) \le \mathsf{S}_m. \tag{208}$$

The middle inequality is equivalent to $(m-1)(m+1)(m+2) \ge 4m(2m-1)$, which holds for $m \ge 6$. The two omitted boundary checks are

m	f	p	$B_m(p,f)$
4	3	1, 2, 3	156, 147, 138
5	4	1, 2, 3, 4	655, 584, 543, 502

$\mathsf{S}_4 = 225, \quad \mathsf{S}_5 = 1176.$ (209)

This proves $B_m(p,f) \le \mathsf{S}_m$ for every $m \ge 4$.

For $m \le 3$, direct evaluation of the finite bound $B_m(p,f)$ already gives the claim except, under $p \le f$, for

$$(m,p,f) = (1,1,1),\ (2,1,1),\ (2,2,2),\ (3,1,1). \tag{210}$$

For $p = f = 1$, write $P = \{p_0\}$ and $F = \{f_0\}$. Apart from the pivot S, the possible occupied O-sets occur in m pairs

$$R_s = (S \setminus \{s\}) \cup \{p_0\}, \qquad Q_s = (S \setminus \{s\}) \cup \{f_0\}, \qquad s \in S. \tag{211}$$

The pivot fibre has size at most $\binom{2m-2}{m-1}$. Apply the two-fibre freezing lemma to R_s and Q_s. In one alternative their respective sizes are bounded by $\binom{2m-2}{m-1}$ and $\binom{2m-2}{m}$; in the other alternative the two bounds are reversed. Therefore

$$|I| \le \binom{2m-2}{m-1} + m\left[\binom{2m-2}{m-1} + \binom{2m-2}{m}\right]. \tag{212}$$

For $m = 1, 2, 3$ the right-hand side is respectively $2, 8, 36$, whereas S_m is respectively $2, 9, 40$.

Consider finally $(m, p, f) = (2, 2, 2)$. Write $S = \{s_1, s_2\}$, $P = \{p_1, p_2\}$, and $F = \{f_1, f_2\}$. The pivot fibre has size at most one, so maximality of the pivot implies that every occupied fibre has size one. The ten possible nonpivot fibres can be paired as

$$\begin{aligned} &\{s_1, p_1\} \leftrightarrow \{s_2, f_1\}, && \{s_1, p_2\} \leftrightarrow \{s_2, f_2\}, \\ &\{s_2, p_1\} \leftrightarrow \{s_1, f_1\}, && \{s_2, p_2\} \leftrightarrow \{s_1, f_2\}, \\ &P \leftrightarrow F. \end{aligned} \tag{213}$$

Every word in one fibre of a pair is adjacent to every word in the other. For example, the include relation forces a word with $O_x = \{s_1, p_1\}$ to satisfy $s_2 \in T_x$, while the avoid relation forces a word with $O_y = \{s_2, f_1\}$ to satisfy $s_1 \notin T_y$. Thus $s_2 \in T_x \cap O_y$ and $s_1 \in O_x \cap Z_y$, which is a staircase generator. The other three mixed pairs follow by symmetry. For the last pair, the constraints force $(O_x, T_x, Z_x) = (P, S, F)$ and $(O_y, T_y, Z_y) = (F, P, S)$, which realizes the other generator. At most one fibre in each pair can therefore be occupied, and

$$|I| \le 1 + 5 = 6 \le \mathsf{S}_2. \tag{214}$$

Together with the separator-star lower bound, this proves equality. □

E. Classification of extremizers

Proof of Theorem XI.6. For $m = 1$, the six permutations of $(0, 1, 2)$ show that each maximum independent pair is a separator star. For $m \ge 2$, use the pivot fibre I_S, sets P, F, and parameters p, f from the extremal proof. The coordinatewise symbol reversal $0 \leftrightarrow 2$ is an automorphism of Γ^{st} and exchanges the include and avoid alternatives of Lemma XI.4. It therefore exchanges p and f, so it is enough to consider $p \le f$.

Suppose $|I| = \mathsf{S}_m$. The equality analysis in the binomial bound forces

$$p = 0, \qquad f \in \{\lfloor m/2 \rfloor, \lceil m/2 \rceil\}. \tag{215}$$

Indeed, if $p = 0$, then $B_m(0, f) = F_m(f)$, and the successive-ratio calculation shows that equality with S_m occurs exactly at the integer or integers nearest to $m/2$. If $p \ge 1$ and $m \ge 4$, every corresponding inequality in the upper-bound proof is strict: in the interior range this follows from strict discrete convexity when $p < f$, while the term -1 makes the estimate strict when $p = f$; the boundary estimates are strict as well. For $m = 2, 3$, the direct checks above give the same conclusion; the exceptional parameter values have upper bounds $8 < \mathsf{S}_2$, $6 < \mathsf{S}_2$, and $36 < \mathsf{S}_3$, respectively.

Set $A_* = S \cup F$, $M = \binom{2m-f}{m}$, and $N = \binom{m+f}{m}$. Since $p = 0$, the pivot argument gives at most N occupied O-fibres, each of size at most $|I_S| \le M$. Moreover,

$$NM = F_m(f) = \mathsf{S}_m. \tag{216}$$

Equality of the total size therefore forces exactly N occupied fibres, each of size M. Every nonpivot fibre is of avoid type and has its O-set contained in A_*. Since A_* has exactly N subsets of size m, every m-set $R \subseteq A_*$ occurs as an occupied O-set, and the pivot fibre is $I_S = \{x \in \mathcal{T}_m : O_x = S,\ T_x \subseteq A_*^c\}$.

Fix an m-set $R \subseteq A_*$ and $z \in A_* \setminus R$. Choose $s \in R$ and put $Q = (R \setminus \{s\}) \cup \{z\}$. Both fibres are occupied. The freezing lemma makes $\ell(R, z) = \mathbf{1}_{\{z \in T_x\}}$ constant over $x \in I_R$. If $f = 1$, any two m-subsets of A_* differ by one exchange, and freezing shows directly that all incidence labels $\ell(R, z)$ coincide. If $f \ge 2$, fix distinct $z, w \in A_* \setminus R$ and $s, t \in R$, and set $Q = (R \setminus \{s, t\}) \cup \{z, w\}$. Freezing between R and Q forces $\ell(R, z) = \ell(R, w)$. Thus the label is constant at each R. Freezing between adjacent m-sets then identifies their constants, and the Johnson graph on the m-subsets of A_* is connected. Consequently there is one global label $\ell \in \{0, 1\}$.

If $\ell = 0$, then every word in the fibre I_R has $T_x \subseteq A_*^c$. There are exactly M such words, so fullness of the fibre gives

$$I = \mathcal{I}^{12}_{A_*} \cap \mathcal{T}_m. \tag{217}$$

If $\ell = 1$, then $A_* \setminus R \subseteq T_x$ for every $x \in I_R$. The remaining $m - f$ elements of T_x lie in A_*^c, giving

$$\binom{2m-f}{m-f} = \binom{2m-f}{m} = M \tag{218}$$

possibilities. Hence every fibre is full and

$$I = \mathcal{I}^{01}_{A_*^c} \cap \mathcal{T}_m. \tag{219}$$

By (215), both A_* and A_*^c have cardinality in $\{\lfloor 3m/2 \rfloor, \lceil 3m/2 \rceil\}$. Undoing the possible symbol reversal preserves the two asserted classes of separator stars. The converse was established by the separator-star construction.

The counts in the stability remark are direct. If $m = 2q$, $|A| = 3q+1$, and $|B| = 3q$, then the largest same-orientation intersection occurs when $B \subset A$ and has size

$$\binom{3q}{2q}\binom{3q-1}{2q}. \tag{220}$$

Its ratio to $|\mathcal{I}_A^{12} \cap \mathcal{T}_{2q}|$ is $(q+1)/(3q+1)$. An opposite-orientation intersection is nonempty only when $|A \cap B| = q+1$, in which case its size is $\binom{2q-1}{q-1}$, exponentially smaller than S_{2q}. □

F. Type-class bridge

Proof. For a type $Q = (n_0, n_1, n_2)$ of length n, let $T_Q = \{x \in \{0,1,2\}^n : |Z_x| = n_0,\ |O_x| = n_1,\ |T_x| = n_2\}$ and $M = \max_i n_i$. The padding map $\iota(x) = x\,0^{M-n_0}1^{M-n_1}2^{M-n_2}$ embeds the type class T_Q into $\mathcal{T}_M$. For two words in the same type class, the appended coordinates have diagonal pair type (a,a), and the staircase generators use only off-diagonal types. Hence the padding map preserves adjacency and non-adjacency, so

$$\alpha(\Gamma^{\mathrm{st}}[T_Q]) \le \alpha(\Gamma^{\mathrm{st}}[\mathcal{T}_M]) = \mathsf{S}_M \le 2^{3Mh(1/3)}. \tag{221}$$

Let c be a proper coloring, $C = c(X^n)$ for $X^n \sim U^{\otimes n}$, and Q_n its empirical type. Then $H(C) = n\log_2 3 - H(X^n \mid C)$ and

$$H(X^n \mid C) \le H(Q_n) + H(X^n \mid C, Q_n). \tag{222}$$

For each color and type, the fiber is independent inside the type class, so

$$H(X^n \mid C, Q_n) \le \mathbb{E}\log_2 |c^{-1}(C) \cap T_{Q_n}|. \tag{223}$$

Also $H(Q_n) \le 2\log_2(n+1)$. Fix $\delta > 0$ and let $\mathcal{E}_n$ be the event $\|Q_n/n - U\|_\infty \le \delta$. Hoeffding's inequality [33] gives $\mathbb{P}(\mathcal{E}_n^c) \le 6e^{-2n\delta^2}$. On $\mathcal{E}_n$, the padding length satisfies $M \le (1/3+\delta)n$, hence

$$\log_2 |c^{-1}(C) \cap T_{Q_n}| \le (1+3\delta)nh(1/3) + 1. \tag{224}$$

On $\mathcal{E}_n^c$, the same logarithm is at most $n\log_2 3$. Therefore

$$\liminf_{n\to\infty} \frac{1}{n} H(C) \ge \log_2 3 - (1+3\delta)h(1/3). \tag{225}$$

Letting $\delta \downarrow 0$ gives

$$R_\star(U) \ge \log_2 3 - h(1/3) = \frac{2}{3}. \tag{226}$$

□

G. Product-refinement obstruction

Proof of Theorem XI.11. Let $C_m = \{0,1\}^m$ and $D_m = \{1,2\}^m$. No pair inside C_m can realize either staircase generator because it has no ordered-pair type 12 or 21; similarly, no pair inside D_m has type 01 or 10. Corollary XI.2 therefore shows that both sets are independent in $G^{(2)}_{m,\varepsilon_0}$.

Define $T(0) = 2$, $T(1) = 1$ coordinatewise and let $\mathcal{S}_m = \{(x, T(x)) : x \in \{0,1\}^m\} \subseteq C_m \times D_m$. Identify binary words with subsets of $[m]$ through their 1-coordinates. If two words x, u are incomparable, there is a coordinate of type 01 and another of type 10. In the second block these contribute types 21 and 12, respectively. The concatenated profile therefore dominates both $e_{01} + e_{12}$ and its transpose $e_{10} + e_{21}$, so the corresponding vertices of $\mathcal{S}_m$ are adjacent. If x and u are comparable, all their differing coordinates have the same orientation. Their concatenated off-diagonal profile is then a multiple of either $e_{01} + e_{21}$ or $e_{10} + e_{12}$, neither of which dominates a bad generator. Hence the graph induced on $\mathcal{S}_m$ is exactly the incomparability graph of the Boolean lattice.

An independent set in that induced graph is a chain. Consequently, any partition of $C_m \times D_m$ into independent sets restricts to a chain partition of the Boolean lattice. By Dilworth's theorem and Sperner's theorem [43], [44], such a partition has at least its width,

$$\binom{m}{\lfloor m/2 \rfloor}, \tag{227}$$

nonempty parts. This proves the claim. □

REFERENCES

[1] N. Bonacorsi, M. Caprio, F. Meneghello, and F. Restuccia, "Credal rate–distortion theory," accepted for presentation at the *IEEE Information Theory Workshop (ITW)*, Tempe, AZ, USA, Nov. 2026.

[2] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, Jul. and Oct. 1948.

[3] ——, "Coding theorems for a discrete source with a fidelity criterion," in *IRE National Convention Record*, vol. 7, pt. 4, pp. 142–163, 1959.

[4] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1971.

[5] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ, USA: Wiley-Interscience, 2006.

[6] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge, U.K.: Cambridge University Press, 2011.

[7] J. Forrest and R. Lougee-Heimer, "CBC user guide," in *Emerging Theory, Methods, and Applications*, INFORMS, 2005, pp. 257–277.

[8] E. Hüllermeier and W. Waegeman, "Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods," *Machine Learning*, vol. 110, no. 3, pp. 457–506, 2021.

[9] B. Frénay and M. Verleysen, "Classification in the presence of label noise: A survey," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 5, pp. 845–869, May 2014.

[10] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," in *Proceedings of the 34th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, vol. 70, 2017, pp. 1321–1330.

[11] P. Walley, *Statistical Reasoning with Imprecise Probabilities*. London, U.K.: Chapman & Hall, 1991.

[12] M. C. M. Troffaes and G. de Cooman, *Lower Previsions*. Chichester, U.K.: Wiley, 2014.

[13] A. Mahmood and A. B. Wagner, "Minimax rate–distortion," *IEEE Transactions on Information Theory*, vol. 69, no. 12, pp. 7712–7737, Dec. 2023.

[14] ——, "Lossy compression with universal distortion," *IEEE Transactions on Information Theory*, vol. 69, no. 6, pp. 3552–3573, Jun. 2023.

[15] V. Malik, T. Kargin, V. Kostina, and B. Hassibi, "A distributionally robust approach to Shannon limits using the Wasserstein distance," in *Proceedings of the IEEE International Symposium on Information Theory*, Athens, Greece, Jul. 2024, pp. 861–866.

[16] J. Abellán and M. Gómez-Olmedo, "Measures of divergence on credal sets," *Fuzzy Sets and Systems*, vol. 157, no. 11, pp. 1514–1531, Jun. 2006.

[17] S. L. Chau, M. Caprio, and K. Muandet, "Integral imprecise probability metrics," in *Advances in Neural Information Processing Systems*, vol. 38, 2025.

[18] J. Körner, "Coding of an information source having ambiguous alphabet and the entropy of graphs," in *Transactions of the Sixth Prague Conference on Information Theory, Statistical Decision Functions, Random Processes*, Prague, Czechoslovakia: Academia, 1973, pp. 411–425.

[19] N. Alon and A. Orlitsky, "Source coding and graph entropies," *IEEE Transactions on Information Theory*, vol. 42, no. 5, pp. 1329–1339, Sep. 1996.

[20] A. Orlitsky and J. R. Roche, "Coding for computing," *IEEE Transactions on Information Theory*, vol. 47, no. 3, pp. 903–917, Mar. 2001.

[21] V. Doshi, D. Shah, M. Médard, and M. Effros, "Functional compression through graph coloring," *IEEE Transactions on Information Theory*, vol. 56, no. 8, pp. 3901–3917, Aug. 2010.

[22] J. Cardinal, S. Fiorini, and G. Van Assche, "On minimum entropy graph colorings," in *Proceedings of the IEEE International Symposium on Information Theory*, 2004, p. 43.

[23] Y. Netzer, T. Wang, A. Coates, A. Bissacco, B. Wu, and A. Y. Ng, "Reading digits in natural images with unsupervised feature learning," in *Proceedings of the NeurIPS Workshop on Deep Learning and Unsupervised Feature Learning*, 2011. [Online]. Available: https://ai.stanford.edu/~twangcat/papers/nips2011_housenumbers.pdf

[24] N. Tishby, F. C. Pereira, and W. Bialek, "The information bottleneck method," in *Proceedings of the 37th Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, USA, 1999, pp. 368–377.

[25] N. Slonim and N. Tishby, "Agglomerative information bottleneck," in *Advances in Neural Information Processing Systems 12*, S. A. Solla, T. K. Leen, and K.-R. Müller, Eds. Cambridge, MA, USA: MIT Press, 2000, pp. 617–623.

[26] H. S. Witsenhausen, "The zero-error side information problem and chromatic numbers," *IEEE Transactions on Information Theory*, vol. 22, no. 5, pp. 592–593, Sep. 1976.

[27] H. Föllmer and T. Knispel, "Entropic risk measures: Coherence vs. convexity, model ambiguity, and robust large deviations," *Stochastics and Dynamics*, vol. 11, no. 2–3, pp. 333–351, 2011.

[28] V. Kostina and S. Verdú, "Fixed-length lossy compression in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 58, no. 6, pp. 3309–3338, Jun. 2012.

[29] R. T. Rockafellar, *Convex Analysis*, ser. Princeton Mathematical Series. Princeton, NJ, USA: Princeton University Press, 1970, no. 28.

[30] G. de Cooman, E. Miranda, and M. Zaffalon, "Independent natural extension," *Artificial Intelligence*, vol. 175, no. 12–13, pp. 1911–1950, 2011.

[31] C. T. Li and A. El Gamal, "Strong functional representation lemma and applications to coding theorems," *IEEE Transactions on Information Theory*, vol. 64, no. 11, pp. 6967–6978, Nov. 2018.

[32] L. E. Dickson, "Finiteness of the odd perfect and primitive abundant numbers with n distinct prime factors," *American Journal of Mathematics*, vol. 35, no. 4, pp. 413–422, Oct. 1913.

[33] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *Journal of the American Statistical Association*, vol. 58, no. 301, pp. 13–30, Mar. 1963.

[34] M. Sion, "On general minimax theorems," *Pacific Journal of Mathematics*, vol. 8, no. 1, pp. 171–176, 1958.

[35] E. R. Scheinerman and D. H. Ullman, *Fractional Graph Theory: A Rational Approach to the Theory of Graphs*. New York, NY, USA: Wiley, 1997.

[36] P. Erdős, C. Ko, and R. Rado, "Intersection theorems for systems of finite sets," *Quarterly Journal of Mathematics*, vol. 12, no. 1, pp. 313–320, 1961.

[37] C. Godsil and K. Meagher, *Erdős–Ko–Rado Theorems: Algebraic Approaches*. Cambridge, U.K.: Cambridge University Press, 2016.

[38] K. Meagher and L. Moura, "Erdős–Ko–Rado theorems for uniform set-partition systems," *Electronic Journal of Combinatorics*, vol. 12, no. 1, Art. no. R40, 2005.

[39] A. Kupavskii, "Erdős–Ko–Rado type results for partitions via spread approximations," *European Journal of Combinatorics*, vol. 132, Art. no. 104288, 2026.

[40] A. Q. M. S. Sayyed, M. Caprio, N. D. Bastian, and F. Restuccia, "Decision-driven credal information fusion," in *Proceedings of the 29th International Conference on Information Fusion*, 2026.

[41] M. D. Donsker and S. R. S. Varadhan, "Asymptotic evaluation of certain Markov process expectations for large time. IV," *Communications on Pure and Applied Mathematics*, vol. 36, no. 2, pp. 183–212, Mar. 1983.

[42] I. Csiszár, "Information-type measures of difference of probability distributions and indirect observations," *Studia Scientiarum Mathematicarum Hungarica*, vol. 2, pp. 299–318, 1967.

[43] R. P. Dilworth, "A decomposition theorem for partially ordered sets," *Annals of Mathematics*, vol. 51, no. 1, pp. 161–166, Jan. 1950.

[44] E. Sperner, "Ein Satz über Untermengen einer endlichen Menge," *Mathematische Zeitschrift*, vol. 27, pp. 544–548, 1928.

[45] V. Doshi, D. Shah, and M. Médard, "Source coding with distortion through graph coloring," in *Proceedings of the IEEE International Symposium on Information Theory*, Nice, France, 2007, pp. 1501–1505.

[46] D. Brélaz, "New methods to color the vertices of a graph," *Communications of the ACM*, vol. 22, no. 4, pp. 251–256, Apr. 1979.

[47] S. Basu, D. Seo, and L. R. Varshney, "Hypergraph-based source codes for function computation under maximal distortion," *IEEE Journal on Selected Areas in Information Theory*, vol. 3, no. 4, pp. 824–838, Dec. 2022.

[48] J. Cardinal, S. Fiorini, and G. Joret, "Minimum entropy coloring," *Journal of Combinatorial Optimization*, vol. 16, no. 4, pp. 361–377, Nov. 2008.

Nicolò Bonacorsi received the B.Sc. degree in Statistical Sciences from the University of Bologna, Bologna, Italy, in 2024, and the B.Sc. degree in Mathematical Sciences for Artificial Intelligence from Sapienza University of Rome, Rome, Italy, in 2025. He is currently pursuing the M.S. degree in Applied Mathematics at Columbia University, New York, NY, USA. His research interests include stochastic control, optimal stopping, stochastic analysis, information theory, imprecise probability, and uncertainty quantification.

Michele Caprio received the Ph.D. degree in Statistics from Duke University, Durham, NC, USA, in 2022. He was a Postdoctoral Researcher with the Department of Computer and Information Science, University of Pennsylvania. He is currently an Assistant Professor of Statistical Machine Learning with the Department of Computer Science, University of Warwick, Coventry, U.K., and an Honorary Lecturer in Machine Learning at the University of Manchester. His research interests include probabilistic machine learning, imprecise probability, applied probability, and uncertainty quantification.

Francesca Meneghello received the B.Sc. degree in Information Engineering in 2016, the M.Sc. degree in Telecommunications Engineering in 2018, and the Ph.D. degree in Information Engineering in 2022, all from the University of Padova, Padua, Italy. She is currently a Principal Research Scientist with the Institute for Intelligent Networked Systems, Northeastern University, Boston, MA, USA. Her research interests include wireless communications, wireless sensing, machine learning, and signal processing for next-generation data-driven wireless systems.

Francesco Restuccia (M'16–SM'21) received the B.S. and M.S. degrees in Computer Engineering from the University of Pisa, Pisa, Italy, in 2009 and 2011, respectively, and the Ph.D. degree in Computer Science from the Missouri University of Science and Technology, Rolla, MO, USA, in 2016. He is currently an Associate Professor with the Department of Electrical and Computer Engineering, Northeastern University, Boston, MA, USA. His research interests include artificial intelligence, computer networks, and intelligent cyber-physical systems. He is a Senior Member of IEEE and ACM.