Beyond Prediction

Reachability, Admissibility, and the Geometry of Future Preservation

Flyxion

Independent Researcher

June 2026

Part of the RSVP / CLIO / Admissibility series.
Companion papers: *Frozen Processes*, *Constraint Before Content*, *Persistent Anomalies and the Geometry of Ontology Revision*.

Abstract

Contemporary artificial intelligence is overwhelmingly framed as prediction: a language model predicts the next token, a planner predicts future rewards, a retrieval system predicts relevance. This framing, while computationally convenient, obscures what capable systems actually do when they succeed.

This monograph develops an alternative account centred on three related ideas: *admissibility* (which futures remain coherent with current constraints), *reachability* (which of those futures are navigable from the present state), and *future preservation* (the objective of maintaining access to admissible trajectories rather than identifying any particular next state).

We motivate the account through a detailed reading of a recent industrial result in machine learning — the Bebop paper on rejection-sampling Multi-Token Prediction (MTP) during reinforcement learning training [3] — which demonstrates, largely without theoretical fanfare, that acceptance in speculative decoding is governed by distributional *overlap* (total variation distance) rather than by predictive accuracy. We argue that TV overlap is a discretised, flat-simplex approximation to a deeper geometric object: the intersection of admissible future sets.

The analysis proceeds across four levels of description, from classical token prediction (Level 0) through distributional overlap (Level 1) to reachable trajectory overlap (Level 2) and full admissibility overlap (Level 3). At each level the dominant question shifts from *What is the next state?* toward *Which futures remain jointly accessible?*

We connect this hierarchy to the RSVP (Relativistic Scalar–Vector Plenum), CLIO (Compressed Latent Information Ontology), and Admissibility frameworks developed in companion papers. We show that entropy, conventionally treated as the independent variable controlling speculative-decoding acceptance, is better understood as a proxy for admissible volume: a scalar measure of local branching whose limitations are revealed as soon as geometry enters the picture.

We close by sketching an admissibility-based learning objective $\mathcal{L}_A = 1 - \mathrm{Vol}(\mathcal{A}_p \cap \mathcal{A}_q) \,/\, \mathrm{Vol}(\mathcal{A}_p)$ that would directly optimise future preservation, and by identifying the Reachability Principle as the unifying claim: *for adaptive systems, durable performance depends more strongly on preservation of admissible future sets than on preservation of present states*.

Contents

Chapter 1

Introduction: The Prediction Assumption

1.1 The Standard Story

Modern AI systems are universally described in predictive terms. A large language model is trained to predict the next token given a context. A reinforcement learning agent predicts cumulative reward. A search engine predicts document relevance. A diffusion model predicts noise residuals. The word *prediction* has become so pervasive that it is rarely examined as a theoretical commitment rather than a neutral description.

This monograph argues that the identification of intelligence with prediction is a contingent historical choice, not a conceptual necessity. Prediction is a special case of a more general operation, and the special case is neither the most powerful nor the most stable one available.

The more general operation is *future preservation*: the maintenance of access to admissible continuations under changing conditions. We develop this claim through a convergence of three sources:

1. A reading of recent industrial machine learning results (§5).
2. The RSVP and CLIO theoretical frameworks (§9, 8).
3. A formal account of admissibility and reachability geometry (§6–§13).

1.2 The Central Thesis

Let A_t denote the *admissible future set* of a system at time t: the collection of trajectories that remain coherent with its current constraints. A system's performance is robust when A_t is preserved across disturbances, not when the system correctly names an element of A_t.

Principle 1 (Reachability Principle). For adaptive systems, durable performance depends more strongly on preservation of admissible future sets than on preservation of present states.

Prediction targets a specific element of A_t. Future preservation targets the set itself. The distinction matters whenever the environment is uncertain, when the system undergoes internal updates, or when multiple continuations are simultaneously viable.

1.3 A Motivating Observation

Principle 1 might seem abstract, but it has a very concrete recent instantiation. In speculative decoding, a draft model proposes tokens that a target model accepts or rejects. The rate of acceptance determines the speed of generation. The conventional assumption is that acceptance depends on how accurately the draft predicts the target's top token.

A 2026 paper by Li et al. [3] shows that this assumption is wrong in a precise and informative way. Acceptance under rejection sampling is given by

$$\alpha_{\mathrm{RS}} = \sum_{v} \min\big(p(v),\, q(v)\big) = 1 - d_{\mathrm{TV}}(p, q),$$

where p is the target distribution and q is the draft distribution. This quantity measures distributional *overlap*, not predictive accuracy. The draft does not need to name the same next token. It needs to cover the same probability mass.

We read this result as an empirical instance of Principle 1 operating within the vocabulary simplex. The rest of the monograph generalises the observation.

1.4 A Concrete Entry Point: The Frozen Archive

Before the formal development begins, it is worth grounding the abstract thesis in a concrete engineering failure that most readers will recognise.

Consider every file format in daily use: MP3, JPEG, FLAC, ZIP. All are built on Shannon's source-coding framework, which was designed for a specific task — transmitting data through a physical channel with minimal latency. A codec is, in Shannon's terms, a *thin pipe*: data enters one end, is compressed, crosses the wire, and is reconstructed at the other. For this purpose, three structural commitments are necessary virtues: a *fixed representational basis* (the vocabulary is chosen at design time and cannot evolve), *write-once encoding* (once a block is compressed it is never revisited), and *modality isolation* (an audio codec knows nothing of co-located video or geometry).

These commitments purchase speed, predictability, and bounded memory. But an archive is not a pipe. A pipe is a transient space; data crosses it and is forgotten. An archive persists for years, accretes evidence, and is consulted across decades. For an archive, the three thin-pipe rules become pathologies: they freeze the system's ontology at the arbitrary moment of first ingestion. The modern digital archive is, in this precise sense, a graveyard of past states of knowledge — every file encoded using whatever the system understood on the day the data arrived, with no mechanism for improvement as understanding grows.

The rematching archive of *Waves of Collapse* [1] inverts all three commitments. Its representational basis grows without bound via a hierarchical template library. Its encodings are retroactive: whenever the library gains a new template, every prior encoding is re-evaluated and rewritten if the new template explains it more cheaply. And its library is shared across modalities, treating photographs, audio recordings, LiDAR sweeps, and text as projections of

a single shared latent world.

This architectural inversion has a consequence that is philosophically sharp before it is technically useful. A system with a growing, retroactively applied, cross-modal template library is not merely a better compression algorithm. It is an *epistemic agent*: its archive is a living theory of the data it has seen, its compression ratio is a direct measure of how well that theory aligns with the generative structure of the world, and the punctuated jumps in that ratio — the waves of collapse — are the system's scientific revolutions, moments at which a new template retroactively reorganises the interpretation of everything previously stored.

The formal development of this monograph is largely an account of why *prediction is not the fundamental operation underlying such a system*, and what replaces it. The thin pipe needs a good predictor of the next symbol. The rematching archive needs something different: a theory of which futures remain jointly navigable by the archive and the world it is trying to model. That is a reachability problem, not a prediction problem.

1.5 Structure of the Monograph

across several domains. Chapter 3 develops the contrast between search and navigation. Chapter 4 documents an independent industrial convergence on the Reachability Principle via the AI+HW 2035 roadmap. Chapter 5 gives a detailed reading of the Li et al. result. Chapter 6 introduces the four-level hierarchy from token prediction to admissibility overlap. Chapter 7 reinterprets entropy as a proxy for admissible volume. Chapter 8 connects the account to CLIO projection geometry. Chapter 9 translates the framework into RSVP field-theoretic terms. Chapter 10 connects future preservation to the theory of repair. Chapter 11 examines the rematching archive of *Waves of Collapse* as a running operational system where all components of the Reachability Principle are simultaneously active and directly measurable. Chapter 12 develops the formal mathematics of explanation as ontological enlargement, deriving scientific discovery, collapse events, and knowledge growth as consequences of deficit reduction. Chapter 13 sketches the admissibility-based learning objective. Chapter 14 states the Reachability Principle formally and draws consequences.

Chapter 2

The Failure of Object-Centric Thinking

2.1 A Recurring Pattern

Across domains, effective systems are routinely described in object terms — as storing, retrieving, and identifying discrete items — while their actual operation is better described in process terms: as maintaining viable continuation, repairing degraded access, and preserving navigable structure.

The gap between description and operation is not merely verbal. Object-centric descriptions generate object-centric objectives. Object-centric objectives generate brittle systems that fail as soon as the objects in question shift, merge, split, or disappear.

2.2 Examples Across Domains

2.2.1 Memory

Classical models of memory treat it as a retrieval system: a stored object m is successfully remembered when the system produces m. But episodic memory is less like retrieval from a fixed store and more like reconstruction under constraint. The relevant criterion is not whether the output matches a stored item but whether the output is admissible: coherent with the constraints encoded across the network.

2.2.2 Identity and Repair

The Ship of Theseus problem arises because identity is modelled as object-identity (same parts) rather than process-identity (same admissible future set). A repaired system is the same system when its future accessibility A is preserved, even if its physical substrate x is entirely replaced. We return to this in Chapter 10.

2.2.3 Programs and Software

A running program is not its source code. It is an execution process with a set of admissible future states. Software maintenance preserves not source identity but computational reachability: the ability to reach the same observable outputs under the same conditions.

Object-centric version control tracks file differences; process-centric understanding tracks behavioural equivalence.

2.2.4 Institutions

An institution persists not because its personnel, bylaws, or physical location are unchanged, but because its characteristic patterns of decision and response remain continuous. Institutional failure is not replacement of objects but collapse of admissible trajectories.

2.3 The General Structure

In each case the pattern is:

1. Object-centric description identifies performance with state identity.
2. Process-centric analysis reveals that performance tracks continuation viability.
3. Robustness comes from preserving A, not from preserving x.

Define formally:

Definition 2.1 (Future Accessibility)**.** Let $\mathcal{T}$ be a space of trajectories and $\mathcal{C}$ a set of constraint relations. The *admissible future set* at state x under constraints $\mathcal{C}$ is

$$A(x) = \{\, \gamma \in \mathcal{T} \mid \gamma(0) = x,\ \gamma \text{ satisfies } \mathcal{C} \,\}.$$

Definition 2.2 (Future Preservation)**.** A transformation $\phi : x \mapsto x'$ is a *future-preserving repair* if $A(x') \approx A(x)$ under some appropriate metric on trajectory sets, even when $x' \neq x$.

The claim of this chapter is that successful adaptive systems implement future-preserving transformations even when their designers describe them in object-preserving terms.

Chapter 3

Search Is Navigation

3.1 Two Models of Intelligence

Classical AI is organised around search: given an objective function f and a domain $\mathcal{X}$, find

$$x^* = \arg\max_{x \in \mathcal{X}} f(x).$$

The paradigm frames intelligence as the identification of a privileged object, the optimal state, answer, move, or token.

An alternative paradigm frames intelligence as navigation: given a current state $x(t)$ and a local constraint field $\mathcal{C}(x, t)$, move to a state $x(t+1)$ that preserves access to admissible continuations:

$$x(t+1) = x(t) + v(x, t), \qquad \gamma(t) \subset A(x(t)).$$

The objective is not finding the unique optimal item but staying inside the navigable corridor.

3.2 Why Navigation Generalises Search

Search is a special case of navigation in which the admissible future set collapses to a single trajectory: the one leading to x^*. When the environment is deterministic, stationary, and fully observed, and when f has a unique maximum, search and navigation coincide.

None of those conditions hold in general. Under uncertainty, the admissible future set has non-trivial volume. Under non-stationarity, optimal states shift, and the relevant question is not whether the current state is optimal but whether optimal states remain reachable. Under partial observability, the relevant set is the intersection of admissible futures across all states consistent with current observations.

3.3 The Vocabulary Simplex as Navigation Space

Consider language generation. The vocabulary simplex $\Delta^{|V|}$ is the space of distributions over next tokens. At each generation step the target model p defines a distribution over this simplex.

Token selection is conventionally framed as search: find the most probable next token (greedy decoding) or sample from the distribution.

Speculative decoding introduces a draft model q that navigates this simplex in parallel. Under target-only sampling, the draft must find the same mode as p: a search problem. Under rejection sampling, the draft only needs to remain in the same *region* of the simplex: a navigation problem.

The Bebop result [3] is, among other things, evidence that language generation is better modelled as navigation than as search.

Chapter 4

Industrial Convergence Toward Reachability

4.1 An Independent Confirmation

The argument developed so far rests on evidence drawn from speculative decoding, distributional geometry, and the theoretical frameworks of RSVP and CLIO. It is therefore striking that an independent research community — spanning semiconductor manufacturers, cloud providers, academic computer architecture groups, and AI laboratories — has converged on a structurally similar conclusion through a different route: the economics and physics of large-scale AI deployment.

The AI+HW 2035 roadmap [2] is nominally an engineering document. Its central target is a thousand-fold improvement in "intelligence per joule" by 2035, achieved through co-design of models, chips, memory, compilers, networks, cooling, and institutions as a single coupled system. Beneath the engineering language, however, the document is making a philosophical claim: the dominant challenge of the next decade is not improving predictive accuracy but preserving access to useful capabilities under increasingly severe physical constraints. This is the Reachability Principle expressed in the language of systems engineering.

4.2 From FLOPs to Admissibility

The historical metric of progress in artificial intelligence has been computational throughput: parameter count, training compute, floating-point operations, or benchmark scores. Such metrics implicitly treat intelligence as a property of a present state, a function of model size or peak performance in isolation.

The roadmap proposes a different metric: intelligence per joule. Although framed as an engineering target, this is more precisely an *admissibility objective*.

Proposition 4.1 (Efficiency as Future Preservation)**.** *Suppose two systems achieve identical task performance. If system A requires one hundred times less energy, memory bandwidth, and communication overhead than system B, then the admissible future set of A strictly contains that of B:*

$$\mathcal{A}(A) \supset \mathcal{A}(B).$$

The lower-cost system is deployable in more locations, on more devices, for longer durations, and under more environmental conditions.

The roadmap identifies data movement, memory bandwidth, interconnect capacity, thermal density, and power delivery as the dominant bottlenecks of future AI systems. These are not prediction failures. They are accessibility failures. The desired computations remain theoretically possible but become *physically inadmissible* under energy, cooling, and cost constraints.

The "1000× efficiency" target is therefore an admissibility-restoration programme: an attempt to reopen computational futures that brute-force scaling is closing. Efficiency is not merely an optimisation target; it is a future-preservation mechanism.

4.3 AI+HW as an RSVP Field

The AI+HW roadmap translates naturally into RSVP terms.

Let $\Phi(x)$ denote the local accessibility of computational state x. Regions requiring excessive power, memory traffic, or communication occupy low-Φ regions; regions realisable efficiently occupy high-Φ regions. The roadmap's proposed technologies are then *field-reshaping operations*:

- **Compute-in-memory** reduces the distance between storage and execution, raising Φ along trajectories previously dominated by memory-movement cost.
- **Photonic interconnects** reduce communication cost, flattening gradients that previously separated distributed computation from local computation.
- **Three-dimensional integration** increases local connectivity and enlarges the accessible basin surrounding a computational state.
- **Adaptive runtimes** modify the transport field $\vec{v}(x,t)$ to steer computation toward regions of lower energy expenditure and higher effective throughput.

The resulting picture is not one of isolated hardware improvements. It is a geometry of accessibility: the future of AI appears governed less by the predictive capability of individual models than by the shape of the constraint field through which those models must navigate.

4.4 The CLIO Failure of Current Metrics

From a CLIO perspective, the roadmap diagnoses a bad projection. Current AI discourse projects intelligence onto parameter count, benchmark score, or peak FLOPs. These projections $\pi : \mathcal{X} \to \mathcal{M}$ discard precisely the constraints that determine whether a system is deployable.

The roadmap argues implicitly for a better projection:

$$\pi_{\text{FLOP}} : \mathcal{X} \to (\text{accuracy, parameter count}), \tag{4.1}$$

$$\pi_{\text{adm}} : \mathcal{X} \to (\text{accuracy, energy, latency, deployability, robustness}). \tag{4.2}$$

Projection (4.1) has low representational entropy for currently dominant models, but high entropy with respect to deployment constraints: many models indistinguishable by FLOP count have wildly different admissible future sets. Projection (4.2) better preserves the information that governs which computational futures remain reachable.

4.5 Convergence Without Common Language

What is most significant about the AI+HW roadmap is the structural convergence it represents.

A large consortium of researchers — working from the physics of semiconductor devices, the economics of data centres, the logistics of hardware supply chains, and the requirements of physical AI deployment — arrives at the same conclusion that the analysis of speculative decoding in Chapter 5 arrives at from the theory of distributions: *the unit of analysis is no longer the model in isolation but the model–memory–compiler–hardware–runtime–energy–application loop.*

This loop is an admissibility structure. Every link constrains which computational futures remain reachable. Breaking any link — running out of memory bandwidth, exceeding thermal limits, losing connectivity, exhausting energy budgets — collapses the admissible future set.

The roadmap frames this as an engineering challenge. The framework developed here suggests it is a deeper theoretical claim: intelligent behaviour is a constrained trajectory through a physical substrate, and the dominant challenges of the next decade concern the geometry of that constraint field rather than the predictive accuracy of the models it must support.

Chapter 5

The Bebop Result: Overlap Governs Acceptance

5.1 Background: Speculative Decoding and MTP

Speculative decoding is a technique for accelerating large language model inference by using a small, fast draft model to propose candidate tokens that are verified in parallel by the full target model. The speedup depends on the *acceptance rate*: the fraction of draft tokens that pass verification.

Multi-Token Prediction (MTP) is a related technique in which a single forward pass of the backbone generates γ draft tokens using lightweight auxiliary heads trained during pre-training or fine-tuning. The challenge addressed by [3] is that MTP acceptance degrades during reinforcement learning training. As the RL objective updates the backbone, a gap opens between the distribution the MTP heads were trained on and the current policy.

5.2 The Standard Explanation and Its Limitation

The obvious explanation is *draft-model mismatch*: the draft heads lag behind the backbone. If this were the dominant effect, online retraining of the draft heads during RL should recover most of the lost acceptance.

Li et al. show empirically and theoretically that this explanation is incomplete. The dominant effect is not mismatch but *policy entropy increase*. As RL training proceeds, the policy entropy $H(p)$ rises — the model becomes less confident — and this rise drives acceptance down independently of any staleness in the draft heads.

5.3 The Acceptance Rate Derivations

5.3.1 Target-Only Sampling

Under target-only sampling (the draft proposes $\hat{y} \sim q$, accepted iff $\hat{y} = \arg\max_v p(v)$), acceptance is approximately:

$$\alpha_{\text{top}} \approx p\Big(\arg\max_v q(v)\Big),$$

which is directly bounded by the maximum probability of the target distribution. When p is nearly uniform (high entropy), $\max_v p(v)$ is small and α_{top} collapses.

5.3.2 Rejection Sampling

Under rejection sampling, the draft proposes $\hat{y} \sim q$ and it is accepted with probability $\min(1, p(\hat{y})/q(\hat{y}))$. The expected acceptance rate is:

$$\alpha_{\text{RS}} = \mathbb{E}_{\hat{y}\sim q}\left[\min\left(1,\, \frac{p(\hat{y})}{q(\hat{y})}\right)\right] = \sum_{v} \min\big(p(v),\, q(v)\big) = 1 - d_{\text{TV}}(p, q). \tag{5.1}$$

This is the paper's central analytical result. Acceptance is governed by total variation overlap, not by whether the draft identifies the same top token.

5.3.3 The TV Loss

Conventional draft training uses cross-entropy (CE) or KL divergence:

$$\mathcal{L}_{\text{CE}} = -\sum_{v} p(v) \log q(v), \qquad \mathcal{L}_{\text{KL}} = \sum_{v} p(v) \log \frac{p(v)}{q(v)}.$$

Li et al. propose instead a total variation loss:

$$\mathcal{L}_{\text{TV}} = d_{\text{TV}}(p, q) = 1 - \sum_{v} \min\big(p(v), q(v)\big).$$

The gradient of $\mathcal{L}_{\text{TV}}$ concentrates learning on the probability mass that the draft already covers, rather than distributing gradient signal across the long tail of the vocabulary. The result is a *probability-proportional mismatch condition*:

$$|q^*(v) - p(v)| \lesssim \delta\, p(v) \implies d_{\text{TV}}(p, q^*) \leq \tfrac{\delta}{2},$$

making acceptance approximately independent of entropy.

5.3.4 Multi-Step Acceptance

For γ draft tokens, expected accepted length is:

$$\mathbb{E}[L] = \sum_{j=1}^{\gamma} \prod_{i=1}^{j} \alpha_i, \tag{5.2}$$

where α_i is the per-step acceptance. Because later tokens are only reached if earlier ones are accepted, the multiplicative structure penalises per-step losses acutely. The paper's end-to-end TV loss optimises (5.2) directly.

5.4 The Key Empirical Result

Li et al. decompose the drop in acceptance during RL into an entropy-driven component $\Delta\alpha_{\text{entropy}}$ and a mismatch-driven component $\Delta\alpha_{\text{mismatch}}$. They find:

$$\Delta\alpha_{\text{entropy}} \gg \Delta\alpha_{\text{mismatch}},$$

with the mismatch contribution approximately zero. This is empirically surprising: the backbone moves substantially during RL, but the draft heads do not become stale in any practically significant sense.

5.5 What the Result Shows

Equation (5.1) states that acceptance is a linear function of the overlap between two distributions on the vocabulary simplex. This is a statement about *shared coverage*, not about correct prediction.

The TV loss result states that directly optimising overlap outperforms indirectly optimising it via CE or KL. This is a statement about the primacy of *geometric overlap* as an objective, over *predictive accuracy* as an objective.

Both results are instances of the general principle that access to a shared region matters more than identification of a shared point. The next chapter frames this observation in the four-level hierarchy.

Chapter 6

A Hierarchy of Future Orientation

6.1 Four Levels of Description

The Bebop result motivates a hierarchy of increasingly general objectives for future-oriented systems. We identify four levels, defined by what the system attempts to preserve.

Level	Object	Question	Metric
0	Next token	What is y_t?	$p(y_t \mid y_{<t})$
1	Token distribution	What mass is shared?	$1 - d_{\mathrm{TV}}(p, q)$
2	Reachable trajectories	What futures overlap?	$\mathrm{Vol}(\mathcal{R}_p(T) \cap \mathcal{R}_q(T))$
3	Admissible futures	What remains coherent?	$\mathrm{Vol}(\mathcal{A}_p \cap \mathcal{A}_q)$

6.2 Level 0: Token Prediction

Classical language modelling. The system names the next token. The relevant performance measure is $p(y_t \mid y_{<t})$. Failure mode: the correct token is not the same as the useful one; or there is no unique correct token; or the distribution over tokens is flat and prediction collapses to guessing.

6.3 Level 1: Distributional Overlap

Bebop. The system preserves distributional mass over the next token. The relevant performance measure is $1 - d_{\mathrm{TV}}(p, q)$. Failure mode: distributions over the simplex can have identical overlap and wildly different semantic behaviour; flat-simplex geometry ignores semantic structure.

6.4 Level 2: Reachable Trajectory Overlap

Generalisation to trajectories. Define:

$$\mathcal{R}_p(T) = \left\{ \gamma : [0, T] \to \mathcal{X} \;\middle|\; P(\gamma) > 0 \right\} \tag{6.1}$$

as the set of trajectories reachable with positive probability under policy p within horizon T. The overlap metric is:

$$R(p, q; T) = \mathrm{Vol}\big(\mathcal{R}_p(T) \cap \mathcal{R}_q(T)\big),$$

measuring how much of the future accessible to the target model is also accessible to the draft model. TV distance on the next token is a one-step, flat approximation to $R(p, q; 1)$.

6.5 Level 3: Admissibility Overlap

Full admissibility. Define $\mathcal{A}_p$ as the set of globally coherent continuations under the constraints operative for policy p (semantic coherence, factual consistency, task completion, physical plausibility, and so on). The admissibility overlap metric is:

$$\mathcal{O}(p, q) = \frac{\mathrm{Vol}(\mathcal{A}_p \cap \mathcal{A}_q)}{\mathrm{Vol}(\mathcal{A}_p)}, \tag{6.2}$$

measuring the fraction of the target model's admissible futures that the draft model also regards as admissible.

Level 1 is a discrete, flat approximation to Level 3. The journey from Level 0 to Level 3 is a progressive geometrisation of the future orientation problem. Chapter 11 examines a system — the rematching archive of *Waves of Collapse* [1] — in which Level 3 admissibility overlap operates as the explicit design criterion, with the compression ratio serving as a direct instrument reading of how much of the world's reachability structure the archive has incorporated.

6.5.1 TV Overlap as a Projection of Trajectory Overlap

The transition from Level 1 to Level 2 may appear to rest on analogy rather than derivation. We establish the formal relationship here.

Let $\mathcal{T}$ denote trajectory space and let $\pi_1 : \mathcal{T} \to V$ be the projection mapping each trajectory to its first emitted token. Policy p induces a probability measure μ_p on $\mathcal{T}$; the token distribution is the pushforward $p(v) = \mu_p(\pi_1^{-1}(v))$. Likewise for policy q, with dominating measure ν on $\mathcal{T}$.

Define the *trajectory overlap*:

$$\mathcal{O}_{\mathcal{T}}(p, q) = \int_{\mathcal{T}} \min\left(\frac{\mathrm{d}\mu_p}{\mathrm{d}\nu}, \frac{\mathrm{d}\mu_q}{\mathrm{d}\nu}\right) \mathrm{d}\nu.$$

Proposition 6.1 (TV as first-coordinate projection)**.** *The one-step overlap $\mathcal{O}_1(p, q) = \int_V \min(p(v), q(v))\, \mathrm{d}v$ equals the trajectory overlap projected onto the first-step σ-algebra:*

$$\mathcal{O}_1(p, q) = 1 - d_{\mathrm{TV}}(p, q).$$

Total variation overlap is therefore not a distinct quantity but the lowest-dimensional marginal of a

more general reachability geometry.

Proof. By the layer-cake identity for min, integrating $\min(p(v), q(v))$ over V recovers exactly $1 - d_{\mathrm{TV}}(p, q)$. Since $p(v) = \mu_p(\pi_1^{-1}(v))$, the token-level integral is the pushforward of the trajectory-level integral restricted to the first-step σ-algebra generated by π_1. □

Proposition 6.1 closes the dimensionality gap: the hierarchy does not merely use TV as a metaphor for trajectory overlap. TV overlap *is* trajectory overlap, restricted to horizon 1 and projected onto a flat vocabulary simplex. Every level of the hierarchy is the same geometric object at a different resolution.

6.6 Each Level Subsumes the Previous

Proposition 6.2. *When the state space is a discrete vocabulary and the horizon is one step, Level 2 and Level 3 reduce to Level 1. When in addition the draft model concentrates mass on a single token, Level 1 reduces to Level 0.*

The hierarchical structure shows that classical language modelling is not wrong but is a limiting special case. It is exact in a degenerate setting (deterministic, one-step, discrete) and approximate elsewhere. The Bebop paper identifies empirically that the next level up — Level 1 — is already substantially better matched to the actual generation task.

6.7 Level 4: Ontological Reachability

Levels 0–3 share a common assumption: the ontology — the language in which futures are described — is fixed. Admissible future sets may expand or contract, but the description vocabulary does not change.

The rematching archive of Chapter 11 reveals that this assumption is not fundamental. A collapse event does not merely expose new trajectories within an existing description space; it enlarges the description space itself. A new template kind is admitted and futures that were previously indescribable become reachable.

This motivates a fifth level of the hierarchy.

Definition 6.3 (Ontology reachability)**.** Let T be a template library. The *ontological reachability set* of T is:

$$\mathcal{R}_\Omega(T) = \{\, T' \mid T' \text{ is reachable from } T \text{ through admissible template admissions} \,\}.$$

Definition 6.4 (Level 4 overlap: Ontological reachability overlap)**.**

$$\Omega(T_1, T_2) = \frac{\mathrm{Vol}\big(\mathcal{R}_\Omega(T_1) \cap \mathcal{R}_\Omega(T_2)\big)}{\mathrm{Vol}\big(\mathcal{R}_\Omega(T_1)\big)}.$$

Two systems may possess identical admissible futures under their current ontologies while differing substantially in their capacity for ontological growth. Levels 0–3 preserve futures *inside* an ontology. Level 4 preserves the *ability to discover new ontologies*.

The full hierarchy is therefore:

Level	**Object**	**Question**	**Metric**
0	Next token	What is y_t?	$p(y_t \mid y_{<t})$
1	Token distribution	What mass is shared?	$1 - d_{\mathrm{TV}}(p, q)$
2	Reachable trajectories	What futures overlap?	$\mathrm{Vol}(\mathcal{R}_p \cap \mathcal{R}_q)$
3	Admissible futures	What coherent futures remain?	$\mathcal{O}(p, q)$
4	Ontological reachability	What description spaces are reachable?	$\Omega(T_1, T_2)$

The distinction between Level 3 and Level 4 is the distinction that runs through the explanation chapter: preservation is conservative (Level 3), but explanation is generative — it creates futures that did not previously exist within the reachable description space (Level 4).

Chapter 7

Entropy as Proxy for Admissible Volume

7.1 The Standard Entropy Story

Li et al. report a strong empirical relationship:

$$H(p) \uparrow \quad \Rightarrow \quad \alpha_{\text{RS}} \downarrow .$$

They interpret this as *entropy bounding acceptance*, and the paper's title refers to a method for *breaking entropy bounds*. The interpretation is locally correct but theoretically incomplete.

7.2 The Geometric Reinterpretation

Entropy is a scalar summary of distributional spread. It conflates two geometrically distinct phenomena:

1. **Isotropic diffusion**: probability mass spreading uniformly in all directions, increasing the effective dimensionality of the admissible region uniformly.

2. **Anisotropic diffusion**: probability mass spreading in structured directions corresponding to semantically coherent continuations while remaining concentrated orthogonal to that structure.

Two distributions with the same entropy can have very different overlap with a fixed draft distribution. Consider:

- p_1: uniform over a coherent cluster of 100 semantically related tokens.

- p_2: uniform over 100 tokens drawn at random from the vocabulary.

Both have $H(p_1) = H(p_2) = \log 100$. But a draft model trained on semantic coherence will overlap strongly with p_1 and weakly with p_2.

Entropy cannot distinguish these cases. Admissible volume can.

7.3 Admissible Volume as the Fundamental Variable

Definition 7.1 (Admissible Volume)**.** Let $\mathcal{A}(p) \subset \mathcal{T}$ be the admissible future set of policy p. The *admissible volume* is

$$V_A(p) = \mathrm{Vol}\big(\mathcal{A}(p)\big).$$

7.3.1 Admissibility Is Not Support

Admissible volume should not be confused with effective support. Define $\mathrm{supp}(p) = \{x : p(x) > 0\}$. Support identifies states that are probabilistically reachable. Admissibility identifies states that remain coherent under the operative constraints. Consequently:

$$\mathcal{A}(p) \subseteq \mathrm{supp}(p).$$

Two policies may possess identical support while exhibiting dramatically different admissibility structure: $\mathrm{supp}(p) = \mathrm{supp}(q)$ but $\mathcal{A}(p) \neq \mathcal{A}(q)$.

This distinction is particularly important in language generation. A policy may assign non-zero probability to every token in the vocabulary — full support — while only a tiny subset of trajectories are semantically coherent. Support measures *possibility*; admissibility measures *coherent possibility*. The latter is strictly stronger, and the gap between them is the space in which the interesting problems of reachability theory live.

Definition 7.2 (Admissibility Curvature)**.** Let $g_{ij} = \mathbb{E}\big[\partial_i \log P(\gamma)\, \partial_j \log P(\gamma)\big]$ be the Fisher information metric on policy space induced by the trajectory distribution. The *admissibility curvature* at p is:

$$\kappa_A(p) = \mathrm{Tr}\big(g^{-1} \nabla^2 \log V_A(p)\big).$$

This definition is coordinate-independent: g^{-1} supplies the Riemannian structure that the informal norm $\|\nabla^2 \log V_A\|$ lacks. Regions with large positive κ_A correspond to futures whose accessible volume changes rapidly under small policy perturbations — high sensitivity basins. Regions with small κ_A correspond to stable admissibility plateaux. The quantity therefore measures local sensitivity of future accessibility rather than distributional uncertainty alone.

The claim is that acceptance in speculative decoding is governed more directly by V_A and κ_A than by $H(p)$. Entropy is a proxy for V_A that is easy to compute but geometrically blind.

Proposition 7.3. *For a policy p over a discrete vocabulary of size $|V|$, the maximum-entropy distribution (uniform) achieves the largest possible admissible volume on the simplex. Entropy is therefore monotone with admissible simplex volume in the isotropic case.*

The isotropic case is the one Bebop implicitly assumes: probability mass is treated as undifferentiated quantity spread over a flat vocabulary space. The moment semantic or trajectory structure is admitted into the geometry, the entropy-volume correspondence breaks.

7.3.2 From Entropy to Deficit

The progression developed throughout this monograph can be summarised as:

$$H \longrightarrow V_A \longrightarrow \delta_T.$$

Entropy H measures *uncertainty*: how spread out the probability mass is.

Admissible volume V_A measures *possibility*: how much future remains coherently reachable under current constraints.

Ontological deficit δ_T measures *missing structure*: how much of the world's reachable description space the current ontology lacks the concepts to express.

The distinctions matter. Two systems may have identical entropy and identical admissible volume while differing dramatically in their ontological deficit — one has a rich template library that captures the generator's structure, the other does not, even though their probability distributions and constraint sets look the same from the outside. Entropy asks how many futures are available. Admissible volume asks how much future remains reachable. Ontological deficit asks which futures remain unreachable because the current ontology cannot yet name them.

7.4 Breaking the Entropy Bound vs. Bypassing It

Li et al. describe their TV loss as *breaking* the entropy bound: training the draft to match distributional shape under high-entropy conditions. The reinterpretation offered here is more radical.

The entropy bound is not broken; it is bypassed. By optimising TV overlap directly, the draft learns to approximate the *shape* of the admissible region rather than the *location* of its mode. This is not breaking a constraint but operating in a space where the constraint does not apply: the space of admissible volume rather than the space of predictive accuracy.

Chapter 8

CLIO: Projection and Admissibility

8.1 The CLIO Framework

CLIO (Compressed Latent Information Ontology) begins from the observation that representational systems operate on projections rather than full states. A projection $\pi : \mathcal{X} \to \mathcal{M}$ collapses a high-dimensional state space onto a lower-dimensional macroscopic representation. Many microscopic states x map to the same macroscopic representation m.

The *representational entropy* of m under π is:

$$S_\pi(m) = \log \mathrm{Vol}\big(\pi^{-1}(m)\big),$$

measuring the number of microscopic states consistent with macroscopic observation m.

8.2 The Draft Head as a Projection Operator

In the MTP architecture, the backbone produces a rich hidden state $h \in \mathcal{H}$, and each draft head applies a lightweight linear map $\pi_k : \mathcal{H} \to \Delta^{|V|}$ to produce a distribution over the vocabulary.

The draft head is therefore a projection. Many different backbone states h may produce the same draft distribution $q = \pi_k(h)$. The relevant question is:

How much of the target model's admissible future region
survives projection through the draft head?

This is precisely the CLIO question applied to the draft-head architecture.

8.3 CE vs. TV Through the CLIO Lens

CE and KL training optimise fidelity of the projection to the target distribution at training-time states. They measure how accurately $\pi_k(h)$ reproduces p when the backbone has state h.

TV training optimises the acceptance boundary induced by the rollout verifier. It measures how much of the target model's probability mass survives projection into the draft head's output simplex.

In CLIO terms:

- CE/KL optimise π_k as a *faithful representation*.
- TV optimises π_k as a *future-preserving projection*.

Faithful representation and future-preserving projection are not the same objective. A projection can be perfectly faithful in representational terms (low KL divergence) while projecting the wrong region of the future (low TV overlap, low admissibility overlap). The Bebop experiments show empirically that TV-projected futures outperform CE-projected futures during RL training. This is CLIO's projection-loss problem instantiated in a production machine learning system.

8.4 Representational Entropy and Policy Entropy

The paper's central finding — that entropy drives acceptance degradation — can be reread through CLIO as:

- As $H(p)$ increases, the admissible region $\mathcal{A}(p)$ expands.
- The draft head's projection π_k can only compress this region, not expand it.
- Under CE/KL training, the projection is calibrated to the pre-RL entropy level.
- When entropy rises, the projection becomes too narrow: it captures only a fraction of the now-larger admissible region.
- TV training adaptively widens the projection to cover the current admissible volume.

The entropy bound is therefore a projection capacity problem, not merely an uncertainty problem. And projection capacity is a CLIO concept.

Chapter 9

RSVP: A Field-Theoretic Translation

9.1 The RSVP Framework

RSVP (Relativistic Scalar–Vector Plenum) models physical and cognitive systems through coupled scalar and vector fields:

- A *scalar field* $\Phi(x,t)$ encoding local future accessibility, entropy density, or constraint intensity.
- A *vector field* $\vec{v}(x,t)$ encoding transport through possibility space: the flow of probability mass, attention, or causal influence.
- A *constraint field* $\mathcal{C}(x,t)$ encoding the admissibility structure that bounds the flow.

9.2 Speculative Decoding as Field Transport

In RSVP terms, the target model defines a scalar field Φ over the possibility space $\mathcal{X}$. High $\Phi(x)$ indicates that x is in a region of rich future accessibility; low $\Phi(x)$ indicates a region of constrained or terminal states.

The draft model is an approximation attempting to follow the flow $\vec{v}(x,t) = -\nabla\Phi(x,t)$ without evaluating the full scalar field. It is a compressed transport surrogate.

The rejection-sampling acceptance criterion,

$$\alpha_{\text{RS}} = 1 - d_{\text{TV}}(p,q),$$

measures whether the surrogate remains inside the same level set of Φ as the full model. Acceptance fails when the surrogate drifts to a different level set — not necessarily because its parameters are wrong, but because the scalar field has been reshaped by RL updates.

9.3 RL as Metric Deformation

Reinforcement learning does not move the policy arbitrarily through distribution space. It moves the policy within the same admissible basin: the region of $\mathcal{X}$ where task-relevant constraints are satisfied. The language of "field reshaping" requires a precise clarification,

because the empirical result $\Delta\alpha_{\text{mismatch}} \approx 0$ depends on a specific claim about what RL does and does not change.

Let $\mathcal{A}_t$ and $\mathcal{A}_{t+\Delta t}$ denote the admissible manifolds before and after an RL update. The claim is:

1. **Topology preserved**: $\mathcal{A}_t \simeq \mathcal{A}_{t+\Delta t}$ (homeomorphic) — connected components, holes, and basin boundaries are unchanged.

2. **Geometry deformed**: $g_t \neq g_{t+\Delta t}$ — the Riemannian metric on the manifold, which governs accessibility gradients, basin depths, and transport costs, is altered.

RL acts primarily as a *metric deformation* rather than a topological transition. It reweights trajectories within the admissible basin without merging or splitting connected components.

Proposition 9.1 (Admissibility Stability Under RL)**.** *If RL updates preserve the topology of $\mathcal{A}(p)$ — moving probability mass within the admissible basin without merging or splitting connected components — then TV overlap is largely preserved and draft-model mismatch is negligible.*

This explains the empirical result. The backbone moves substantially in weight space, but the admissible manifold's topology is stable. The draft heads track the manifold's shape, not its exact metric, so they remain approximately valid after topological-preservation updates. Only a topological transition — a genuine paradigm shift in which the admissible basin splits, merges, or acquires new connected components — would require retraining the draft heads.

9.4 CE Is Pointwise; TV Is Geometric

In RSVP language:

- CE training is *pointwise*: it minimises prediction error at each location x in the scalar field independently.

- TV training is *geometric*: it minimises failure of transport across the field by preserving the coverage of high-Φ regions.

The distinction is not about the magnitude of error. It is about whether the objective is aligned with the field's topology. A system that minimises pointwise error can still fail globally if its errors consistently occur in the high-Φ (high-accessibility) regions. TV training avoids this by weighting errors proportionally to probability mass, which is a proxy for local field strength.

Chapter 10

Repair and Future Preservation

10.1 Repair as Admissibility-Preservation

In the companion monograph *Frozen Processes*, repair is defined not as state restoration but as the preservation of future accessibility. An entity x is repaired into x' when:

$$A(x') \approx A(x), \quad \text{even when} \quad x' \neq x.$$

This definition is ontologically minimal: it does not require that the original state be reconstructed, only that the relevant future trajectories remain accessible.

10.2 The Bebop Result as an Instance of Repair Theory

The Bebop paper's empirical finding — that RL does not significantly degrade MTP acceptance because $\Delta\alpha_{\text{mismatch}} \approx 0$ — is an instance of admissibility-preservation under repair.

The backbone undergoes significant weight updates during RL. In parameter space, this is substantial change. But in admissibility space, the change is mild. The RL updates constitute a *repair*: they modify the implementation of the policy while preserving the structure of its admissible future set.

This is why online retraining of the draft heads is largely unnecessary. The draft heads were trained to approximate $\mathcal{A}(p)$. After RL, $\mathcal{A}(p)$ is approximately the same. The implementation p changed; the admissibility structure did not.

10.3 The Ship of Theseus in Distribution Space

The Ship of Theseus problem asks whether an object with all parts replaced is the same object. The Bebop result suggests a distributional analogue: a policy with substantially updated weights can be *the same policy* in the sense that its admissible future set is preserved.

This is not a metaphor. It is a quantitative claim: $d_{\text{TV}}(p_{\text{before}}, p_{\text{after}})$ may be large, but $\text{Vol}(\mathcal{A}(p_{\text{before}}) \cap \mathcal{A}(p_{\text{after}}))$ may remain close to $\text{Vol}(\mathcal{A}(p_{\text{before}}))$. The policy changed; the policy's reachable futures did not.

10.4 Implications for Continual Learning

The repair interpretation suggests a criterion for safe continual learning:

Definition 10.1 (Admissibility-Preserving Update)**.** A parameter update $\theta \mapsto \theta'$ is *admissibility-preserving* if

$$\mathcal{O}(p_\theta, p_{\theta'}) = \frac{\mathrm{Vol}(\mathcal{A}(p_\theta) \cap \mathcal{A}(p_{\theta'}))}{\mathrm{Vol}(\mathcal{A}(p_\theta))} \geq 1 - \epsilon.$$

Catastrophic forgetting, in this framework, is not loss of stored information but collapse of admissible future overlap: after the update, large portions of previously admissible futures are no longer reachable. An admissibility-preserving update criterion would regulate this directly, without requiring replay or regularisation toward old parameters.

Chapter 11

The Rematching Archive: Admissibility Made Operational

11.1 From Principle to Running System

The preceding chapters have developed a theoretical account of admissibility and reachability primarily through analysis of existing machine learning results — speculative decoding acceleration, RL policy stability, AI hardware constraints. The framework gains additional traction when the same principles are instantiated not as a reading of someone else's system but as the explicit design of a new one.

The monograph *Waves of Collapse: Compression as the Geometry of Reachable Description* [1] develops precisely this instantiation. Its central object is a *rematching archive*: a compression system that maintains a growing hierarchy of templates and, whenever the hierarchy gains a new entry, retroactively re-evaluates every prior encoding to determine whether the new template explains it more cheaply. The system is introduced there as an engineering idea, but the engineering forces a philosophical commitment identical to the one argued here: the primitive notion in the archive's operation is not what a signal *is* but what encodings of it are *reachable* from the current library.

11.2 The Admissibility Log and the Compression Staircase

The archive maintains an *Admissibility Log*: an append-only ledger recording which descriptions became reachable, and when. Compression does not improve smoothly over time. It improves in waves — long plateaus during which the archive encodes competently under its current template ontology, punctuated by *collapse events*: discontinuous drops in total archive size triggered by the admission of a single highly general template.

The resulting compression trajectory has the shape of a devil's staircase:

$$1.0\times \;\to\; 1.3\times \;\to\; 1.8\times \;\to\; 2.7\times \;\to\; 4.1\times \;\to\; \cdots$$

with jumps clustered around discoveries rather than spread uniformly. This is not an implementation artifact. It is the empirical signature of reachability structure: within a given

ontological stratum, all descriptions reachable by cheap moves are explored and exploited; progress requires crossing a barrier into a new stratum by admitting a template whose short-term cost (the library grows) is paid back by a cascade of rewrites.

The staircase is therefore the Reachability Principle made visible as an instrument reading. Each plateau is a stratum of description space; each vertical drop is a transition; the Admissibility Log is the event-sourced ledger of stratum history.

11.3 The Ontological Deficit as a Measured Quantity

The rematching archive introduces a precise measurable analogue to the concept of admissible volume developed in Chapter 7. Define the *reachable optimum* from archive state σ as:

$$\Lambda^*(\sigma) = \min\{\Lambda(\sigma') : \sigma' \in \text{Reach}(\sigma)\},$$

the least total description length achievable by admissible moves from σ. The *ontological deficit* is then:

$$\delta_T(\sigma) = \Lambda^*(\sigma) - K(\mathcal{D}),$$

where $K(\mathcal{D})$ is the (ideal, uncomputable) Kolmogorov complexity of the corpus.

The deficit δ_T is the frozen invariant of the current stratum: the portion of description length that cannot be reduced by any move available within the current template hierarchy. It measures, precisely, what the archive's present ontology cannot say. Its decrease over time — stepwise, at collapse events — is the archive's intellectual history.

This is the same concept as admissible volume $V_A(p)$ of Chapter 7, but made computable and directly observable: the deficit is logged as a numerical quantity at every collapse event, with a timestamp, a magnitude m, and an identifiable cause (the admitted template). The abstract claim that "entropy is a proxy for admissible volume" here has an operational counterpart: δ_T is the volume of futures trapped inside a stratum boundary, waiting for an admission that opens the channel.

11.4 Collapse Events as Phase Transitions

The archive formalises a claim made informally in Chapter 2: that repair and revision are first-order transitions, not continuous drift.

Waves of Collapse states this as a theorem (schematic): a collapse event of magnitude m triggered by a single template admission has the structure of a first-order phase transition — the order parameter (realized-code fraction $S/(S+\Phi)$) jumps discontinuously by m, the two strata coexist during the cascade as rewritten and not-yet-rewritten regions propagate through the reference graph, and hysteresis obtains: the reverse transition does not occur at the same ledger threshold, because the rewritten archive now depends on the admitted template. Strata are therefore not merely conceptual distinctions. They are operational barriers with measurable heights, and the staircase is their empirical trace.

This connects the archive directly to the repair framework of Chapter 10. An archive repair — the admission of a template that retroactively restructures thousands of prior encodings — preserves the admissible future set of every stored observation (each can still be decoded exactly) while transforming the implementation. The observation's identity is stable across the rewrite because it was always indexed by its boundary conditions in the Spherepop sense, never by its bit pattern. What changes is not the object but the path by which it is reached. This is future preservation, not state preservation, implemented at scale.

11.5 Multimodal Reachability and the World-Model Crossover

The archive's most striking operational result — the world-model crossover — is a direct consequence of the level hierarchy developed in Chapter 6.

In a mature multimodal archive, observations from different sensors (photographs, LiDAR sweeps, audio recordings, thermal maps) are not stored as separate modality-specific files. They are stored as projections of a shared latent world model plus residuals. A 2018 photograph and a 2026 LiDAR scan of the same building compress each other, because both are evidence about the same persistent structure.

The crossover theorem (*Waves of Collapse*, Theorem 4.1) identifies a finite horizon n^* past which the archive's latent model is the dominant information-bearing object: past n^*, the per-observation marginal storage cost converges not to the sensor's data rate but to $H(O \mid M)$, the conditional entropy of the observation given the model — the world's novelty rate.

In the terms of this monograph, n^* is the point at which the archive's reachable future set transitions from being bounded by the modality kernel to being bounded by world novelty. Storage cost tracks admissible volume, not signal volume. A 16K camera recording an empty, unchanging room costs almost nothing past n^* because the room's geometry and lighting are already in the library; the archive spends bits only on the genuinely unmodeled.

This is Level 3 admissibility overlap operating at system scale. The archive's templates and the world's generative structure share admissible futures; the compression ratio is a direct measurement of how much of the world's reachability structure the archive has incorporated.

11.6 The Archive as an RSVP Apparatus

Waves of Collapse makes an explicit claim that deserves to be registered here: the rematching archive is the first system in the RSVP programme where all three field components are directly observable at native resolution.

In the RSVP identification:

- Φ is unexplained structure density over description space — residual mass weighted by its systematicity. High Φ marks compressible-but-uncompressed regions: the anomaly field.

- $\vec{v}$ is the rematching flow: directed transport of description mass through the reachability graph as rewrites propagate. Collapse events are shock fronts in $\vec{v}$, catastrophic local convergence as trapped mass funnels through a newly opened channel.
- S is realised code: entropy already extracted into efficient form, the archive's crystallised understanding.

The relaxation equation

$$\frac{\partial \Phi}{\partial t} + \nabla \cdot (\Phi \, \vec{v}) = -\Gamma(\Phi, T) + \eta(x, t)$$

is not merely a formal analogy; every term is loggable from the archive's event ledger. η is the ingestion rate (novelty source); $\Gamma(\Phi, T)$ is the rate at which anomalous density is converted to crystallised code, which vanishes on a plateau not because $\Phi = 0$ but because the constraint surface offers no admissible channel. The ontological deficit δ_T is $\Phi > 0$ trapped by a stratum boundary: a field quantity with a direct instrument reading.

11.6.1 An Order Parameter for Understanding

The archive naturally defines a scalar order parameter measuring the fraction of structure that has already been incorporated into the current ontology. Let:

$$\chi = \frac{S}{S + \Phi},$$

where S is crystallised description and Φ is unexplained structure density. The quantity satisfies $0 \leq \chi \leq 1$, with limiting interpretations:

- $\chi \approx 0$: anomaly-dominated regime — the archive has ingested much and understood little.
- $\chi \approx 1$: crystallised regime — most structure has been converted to efficient representation.

Collapse events are discontinuous jumps in χ: a template admission removes a stratum boundary, trapped Φ funnels through the newly opened channel, and S spikes. The archive thus admits a measurable order parameter directly analogous to those used in statistical physics to characterise phase transitions, and its time series is an observable record of the system's intellectual history.

11.6.2 The Deficit Without an Oracle

The appearance of Kolmogorov complexity in the deficit definition $\delta_T = \Lambda(T, \mathcal{D}) - K(\mathcal{D})$ does not imply operational access to $K(\mathcal{D})$. The ideal complexity serves as an unreachable lower bound, not a computed target.

In practice, one may instead define the *operational deficit*:

$$\hat{\delta}_T = \Lambda(T, \mathcal{D}) - \Lambda(T^*, \mathcal{D}),$$

where T^* is the best currently available ontology. The operational deficit is relative rather than absolute: as ontologies improve, estimates of deficit improve correspondingly. Kolmogorov complexity functions as a limiting attractor of the sequence $\hat{\delta}_{T_0} \geq \hat{\delta}_{T_1} \geq \cdots$ rather than a quantity directly evaluated. The framework therefore does not require an oracle; it requires only a succession of improving approximations, which is exactly what the archive's admission process produces.

11.7 File Size as a Relational Property

One engineering consequence of the rematching architecture deserves separate treatment because it overturns an assumption so deeply embedded in file-system design that it is rarely articulated as an assumption at all.

In every file system currently deployed, the size of a file is an *intrinsic* property of that file. A photograph reports 2.4 megabytes; the operating system reads that number from the file's metadata; the number is independent of everything else stored on the disk. This seems not just obvious but necessary.

In a rematching archive it is false. The correct definition of a file's storage cost is marginal:

$$\text{cost}(o) = \Lambda(\mathcal{D}) - \Lambda(\mathcal{D} \setminus \{o\}), \tag{11.1}$$

the amount by which total description length increases when o is present compared to when it is absent. This quantity depends entirely on what else is in the archive.

The consequence is sharp. The thousandth photograph of a well-modelled cat approaches zero marginal cost: the archive already holds the cat's geometry, lighting parameters, and characteristic poses; the new image is almost entirely predicted and contributes only a thin residual. The same photograph in an empty archive costs megabytes, because the archive must construct the cat model from scratch. File size is a property of the file's relationship to everything already known, not a property of the file itself.

This is not merely an engineering curiosity. It is a direct consequence of the Reachability Principle operating at the storage layer: the cost of an object is the increase in ontological deficit its presence introduces, which depends on the current state of the template hierarchy. When deficit is already low in a region of description space — because the archive has built deep, general templates covering that region — new observations in that region are almost free. Storage cost tracks reachability of description, not volume of data.

11.8 Knowledge as Navigation in the Archive

The archive's closing philosophical claim (Chapter 10 of *Waves of Collapse*) is the simplest and most direct statement of the thesis of this monograph:

> *Knowledge, throughout these frameworks, is navigation through reachable states; understanding is the growth of reachability. The archive renders this literal: a signal is known to the degree that it is reachable from a short description over the current hierarchy; learning is the construction of new admissible paths; and the punctuated collapse of the archive's size is the visible signature of new regions of description space becoming navigable. Compression is not a thing one does to knowledge. Compression is what knowledge looks like from the storage layer.*

This is the Reachability Principle stated as an engineering axiom rather than a philosophical thesis. The archive does not claim to store knowledge; it claims to store reachability. What persists across decades of rewrites is not any particular bit pattern but the capacity to regenerate an observation along an admissible path. Understanding accumulates by making more of the world navigable from a shorter sentence, and the compression staircase is the record of that accumulation.

Chapter 12

Compression, Reachability, and the Nature of Explanation

12.1 The Question Behind the Archive

Chapter 11 introduced the rematching archive as an operational instantiation of the Reachability Principle. A natural question follows: what exactly happens to explanation when the archive's description of a corpus shrinks? Compression is numerically measurable, but is it intellectually significant, or merely a useful engineering proxy?

This chapter argues that the question dissolves once the formal relationship between compression, reachable description space, and ontological deficit is made precise. Explanation does not *accompany* reachability expansion; it *is* reachability expansion, and the formalisation of this identity produces a sequence of theorems that unify archive dynamics, scientific discovery, CLIO projections, and the Reachability Principle under a single algebraic structure.

12.2 Reachable Description Space

Definition 12.1 (Reachable description space)**.** Let $\mathcal{D}$ be a corpus and T a template library. The *reachable description space* is:

$$\mathcal{R}_T(\mathcal{D}) = \left\{ \delta \;\middle|\; \delta \text{ describes } \mathcal{D} \text{ using only templates in } T \right\}.$$

The *shortest reachable description length* is:

$$\Lambda(T, \mathcal{D}) = \min_{\delta \in \mathcal{R}_T(\mathcal{D})} |\delta|.$$

The reachable description space is not a fixed object. It depends entirely on T, and T is a dynamic quantity — growing with each template admission, occasionally shrinking via retirement, and undergoing topology changes at collapse events. $\Lambda(T, \mathcal{D})$ is therefore not a property of the data. It is a property of the relationship between the data and the current state of knowledge.

Theorem 12.2 (Reachability Monotonicity)**.** *If $T \subseteq T'$ then $\mathcal{R}_T(\mathcal{D}) \subseteq \mathcal{R}_{T'}(\mathcal{D})$ and $\Lambda(T', \mathcal{D}) \leq \Lambda(T, \mathcal{D})$.*

Proof. Any description valid over T is valid over $T' \supseteq T$, so the minimisation in $\Lambda(T', \mathcal{D})$ is taken over a weakly larger set. Minimisation over a larger set cannot increase the minimum. □

Theorem 12.2 may appear trivial. Its significance is philosophical. It formally establishes that learning is *enlargement of reachable description space*, not approximation toward a fixed target. The Kolmogorov complexity $K(\mathcal{D})$ — the ideal shortest description — is a fixed lower bound, but it is uncomputable and unreachable from any finite library. The archive always operates at $\Lambda(T, \mathcal{D}) \geq K(\mathcal{D})$, and progress is reduction of the gap between them. This gap is the ontological deficit.

12.3 Ontological Deficit as the True Measure of Ignorance

The *ontological deficit* introduced in *Waves of Collapse* [1] is:

$$\delta_T(\mathcal{D}) = \Lambda(T, \mathcal{D}) - K(\mathcal{D}). \tag{12.1}$$

The deficit measures what the current ontology cannot say: the portion of description length that is not reducible by any admissible move within the current template family. It is ignorance converted into geometry rather than probability — not a distribution over unknown states but a distance in description space between where the archive stands and where it could stand.

Unlike compression gain, which measures local improvement, deficit reduction measures enlargement of the reachable universe:

Definition 12.3 (Explanatory gain)**.** The *explanatory gain* of admitting template τ is:

$$\Delta E(\tau) = \Lambda(T, \mathcal{D}) - \Lambda(T \cup \{\tau\}, \mathcal{D}) = \delta_T(\mathcal{D}) - \delta_{T \cup \{\tau\}}(\mathcal{D}).$$

Theorem 12.4 (Explanation as reachability expansion)**.** $\Delta E(\tau) > 0$ *if and only if there exists a description* $\delta \in \mathcal{R}_{T \cup \{\tau\}}(\mathcal{D}) \setminus \mathcal{R}_T(\mathcal{D})$ *with* $|\delta| < \Lambda(T, \mathcal{D})$.

Proof. By Theorem 12.2, $\Lambda(T \cup \{\tau\}, \mathcal{D}) \leq \Lambda(T, \mathcal{D})$. The gain is strictly positive exactly when the new minimum is achieved by a description not available in $\mathcal{R}_T(\mathcal{D})$. If no such description exists, the minimiser is unchanged and $\Delta E(\tau) = 0$. □

Theorem 12.4 states what explanation *is*: the creation of paths through description space that were previously inaccessible. A template is explanatory not because it names something true — truth is a constraint on admissibility, not the primary criterion — but because its admission opens routes to shorter descriptions that the current library could not construct.

12.4 The Collapse Staircase from Reachability Geometry

The wave structure of compression, observed empirically in *Waves of Collapse*, follows directly from the deficit geometry. Define the *collapse magnitude* of an admission:

$$C(\tau) = \Delta E(\tau) = \delta_T - \delta_{T \cup \{\tau\}}.$$

The archive's total description length trajectory is then:

$$\Lambda(t) = \Lambda_0 - \sum_i C_i \cdot \mathbf{1}[t \geq t_i], \tag{12.2}$$

where t_i is the timestamp of the i-th admission and C_i its collapse magnitude.

Equation (12.2) generates the devil's staircase structure directly: $\Lambda(t)$ is piecewise constant, decreasing only at discrete event times, with jump sizes C_i that are heavy-tailed because a highly general template reindexes a large fraction of the archive simultaneously. Compression improves discontinuously because reachable description space expands discontinuously. The staircase is not an implementation artifact. It is the generic behaviour of a monotone quantity that advances only at barrier crossings.

12.5 Persistent Anomalies as Gradients Toward New Ontology

Not all residuals are equal. In a mature rematching archive, the residual stream decomposes into three qualitatively distinct categories:

Noise Random, incompressible fluctuations that carry no systematic structure. No template of any kind can reduce this residual; it is the irreducible floor set by the source's intrinsic randomness. Noise is not informative about missing templates.

Novelty Genuinely new information not previously encountered. This residual is large for new objects or events but decreases rapidly as similar observations accumulate and templates are built. Novelty is what the archive *should* be spending its bits on.

Systematic misfit Structured, persistent residual that is compressible in principle but not expressible over the current library. This residual has regularity — it correlates with itself, responds to partial template matches, and clusters spatially in description space — but no template in T provides an admissible channel to encode it. Systematic misfit is the archive's research programme: its presence indicates that a profitable template admission exists but has not yet been found.

Only the third category signals an ontological gap. The TARTAN discipline [1] provides the spatial instrument: a recursive tiling of description space that makes systematic-misfit clusters visible as high-density regions in the anomaly field Φ, distinguishable from noise by their spatial coherence and from novelty by their persistence across repeated observations.

A *persistent anomaly* is a systematic-misfit region that no within-stratum move reduces:

Definition 12.5 (Persistent anomaly)**.** Residual density $R(x)$ at description-space coordinate x is a *persistent anomaly* if

$$\min_{\tau \in T} \Delta E(\tau; x) = 0,$$

where $\Delta E(\tau; x)$ is the local gain achievable by any template in T at x.

Definition 12.6 (Anomaly pressure)**.**

$$P_A = \int_{\Omega} R(x)\, \mathrm{d}x,$$

the total trapped anomaly mass within the current stratum Ω.

A persistent anomaly is not evidence of a theory's failure. It is a gradient pointing out of the current stratum: a region of description space where the deficit is locally high and where no template in T provides an admissible channel toward shorter descriptions.

Proposition 12.7 (Anomaly as ontological gradient)**.** *A persistent anomaly at x defines a direction in template space: the set*

$$\mathcal{N}(x) = \{\tau \notin T \mid \Delta E(\tau; x) > 0\}$$

is non-empty, and every element of $\mathcal{N}(x)$ is a candidate for stratum-crossing admission.

This is *Persistent Anomalies and the Geometry of Ontology Revision* stated as a consequence of the deficit formalism. The anomaly field $R(x)$ is not noise to be silenced but evidence for a missing template kind, and monitoring its spatial structure gives the archive an intrinsic research programme: the next collapse event is not arbitrary but is already indicated by the geometry of the current residual.

12.6 Three Levels of Discovery

The Weierstrass analysis of *Waves of Collapse* [1] reveals that template admissions are not all structurally equivalent. Three qualitatively distinct levels exist:

Entity discovery A new template τ of the existing kind is admitted. $\mathcal{R}_T$ expands but its generative grammar is unchanged. This is normal science: adding a new instance of a known kind.

Generator discovery A new template kind is admitted that can generate multiple instances through parameterisation. $\mathcal{R}_{T'}$ expands by a family, not a point. Mathematically, $T \to T'$ introduces a schema rather than a token.

Ontology revision The current template family T has the property that

$$\delta_T > \epsilon$$

for every finite extension within the family. No accumulation of entity or generator admissions reaches the necessary short description; the deficiency is architectural. The Weierstrass function is the canonical example: its Kolmogorov complexity is that of a one-line formula, but its ontological deficit under any finite fragment library is unbounded, because the generator is recursive and self-referential in a way that no non-recursive template family can express.

Proposition 12.8 (Forced stratum transition)**.** *Suppose $\delta_T(\mathcal{D}) > \epsilon$ for every library T' reachable from T within the current template class. Then reducing δ below ϵ requires admitting a template of a qualitatively new kind — one not representable as a composition of existing template types.*

This gives a rigorous content to the intuition that some scientific revolutions are different in kind from normal-science progress. The distinction is not sociological; it is geometrical. An ontology revision is a barrier crossing that requires a change in the *type* of available descriptions, not merely an extension of the current type.

12.7 The World-Model Crossover and the Primacy of Theory

The world-model crossover theorem (*Waves of Collapse*, Theorem 4.1) establishes a phase boundary in the growth of reachable description space. Under diverse measurement kernels, the joint information of accumulated observations about the latent structure M satisfies:

$$I(M;\,\{O_1,\ldots,O_n\}) \to H(M), \quad \text{while} \quad \max_i I(M;\,O_i) \le C < H(M).$$

There exists n^* past which the latent model M — which is the archive's template hierarchy — is the dominant information-bearing object.

In terms of the deficit, the crossover is the point at which the deficit reduction contributed by individual observations becomes negligible:

$$n > n^* \implies \frac{\mathrm{d}\,\delta_T}{\mathrm{d}n} \approx 0 \text{ and } \Lambda(T, \mathcal{D} \cup \{O_{n+1}\}) \approx \Lambda(T, \mathcal{D}) + H(O \mid M).$$

Past the crossover, marginal storage cost tracks the world's novelty rate, not the sensor's data rate.

The philosophical consequence is sharp. Before the crossover, observations are the primary information-bearing objects and the model is a summary. After the crossover, the model is the primary information-bearing object and observations are residuals — evidence used to update or extend the model. Theory becomes ontologically primary at a measurable, computable point.

This is CLIO's projection argument made temporal. CLIO asks: what information survives the projection $\pi : \mathcal{X} \to \mathcal{M}$? The crossover theorem answers: everything, asymptotically under diverse kernels. The latent manifold $\mathcal{M}$ eventually captures $H(M)$ bits of structure, at which point individual observations are fully explained as instances of the general theory plus novelty.

12.8 Knowledge Growth as Field Dynamics

The RSVP identification of Chapter 9 can now be extended to a conservation equation for knowledge. Let $\Phi(x,t)$ denote unexplained structure density (anomaly mass) and $\Gamma(\Phi,T)$ the rate at which it is converted to crystallised description S. From *Waves of Collapse*:

$$\frac{\partial \Phi}{\partial t} + \nabla \cdot (\Phi\, \vec{v}) = -\Gamma(\Phi,T) + \eta(x,t), \qquad \frac{\partial S}{\partial t} = \Gamma(\Phi,T).$$

Define *total embodied knowledge* as:

$$K(T,t) = \int_{\Omega} S(x,t)\, \mathrm{d}x. \tag{12.3}$$

Then:

$$\frac{\mathrm{d}K}{\mathrm{d}t} = \int_{\Omega} \Gamma(\Phi,T)\, \mathrm{d}x = -\frac{\mathrm{d}\delta_T}{\mathrm{d}t} \tag{12.4}$$

(since Γ converts anomaly mass to crystallised code while leaving total description mass $\Phi + S$ conserved up to ingestion).

Equation (12.4) is the central result of this chapter. Knowledge growth is the rate at which anomaly mass is converted into admissible structure. On a plateau, $\Gamma = 0$ even though $\Phi > 0$: the stratum boundary blocks conversion. At a collapse event, Γ spikes: a template admission removes the structural bottleneck and trapped mass funnels through as a shock front.

This is not a metaphor drawn from thermodynamics. It is the same equation, with the same conserved quantities, applied to the same dynamics — but now the conserved quantity is description length rather than energy, and the barriers are ontological rather than energetic.

12.9 A Unified Characterisation of Theory

The preceding sections converge on a single characterisation:

Principle 2 (Explanation as Ontological Enlargement). A theory τ is an admissible transformation $T \to T \cup \{\tau\}$ that:

1. reduces ontological deficit: $\delta_{T\cup\{\tau\}} < \delta_T$;
2. expands reachable description space: $\mathcal{R}_T(\mathcal{D}) \subsetneq \mathcal{R}_{T\cup\{\tau\}}(\mathcal{D})$;
3. increases future compression potential: for future observations $\mathcal{D}'$, $\Lambda(T \cup \{\tau\}, \mathcal{D}') < \Lambda(T, \mathcal{D}')$.

Condition (1) measures immediate explanatory gain. Condition (2) measures ontological enrichment: new paths in description space that were inaccessible are now traversable. Condition (3) measures transferability: the theory is not merely an efficient re-encoding of seen data but a genuine compression of the generator, applicable to unseen instances.

Remark 12.9. Condition (3) distinguishes genuine discovery from overfitting. A template that reduces $\Lambda(T, \mathcal{D})$ but not $\Lambda(T, \mathcal{D}')$ for fresh $\mathcal{D}'$ drawn from the same source has reduced deficit locally at the cost of increasing it elsewhere. MDL's two-part charge — the template must pay its own storage cost — is the practical approximation to this condition.

This characterisation unifies the disparate phenomena examined across the preceding chapters. Scientific theories, compression templates, RL policy updates, AI hardware co-design, repair operations, and archive collapse events are all admissible transformations that reduce ontological deficit, expand reachable description space, and increase future compression potential — with the only difference being the space over which the description is defined (signal space, weight space, physical possibility space, institutional policy space). The Reachability Principle is the claim that performance is governed by the second condition above, not by the state of T at any given moment.

12.9.1 Explanation Creates Futures

The Reachability Principle is often stated in a conservative form: successful systems *preserve* admissible futures. The deficit formalism reveals a stronger claim.

Explanation is not merely preservation. *Explanation is future creation.*

Suppose a template admission τ satisfies $\Delta E(\tau) > 0$. By Theorem 12.4, there exists a description $\delta \in \mathcal{R}_{T\cup\{\tau\}}(\mathcal{D}) \setminus \mathcal{R}_T(\mathcal{D})$ with $|\delta| < \Lambda(T, \mathcal{D})$.

The theory has therefore constructed a path through description space that *did not previously exist*. The newly reachable description is not merely a better way of saying something that was already sayable. It names a connection, an entity, or a law that the prior ontology had no resources to express. Scientific discovery is not the identification of facts about a fixed landscape. It is the construction of new admissible routes through a landscape that is itself enlarged by the act of discovery.

This shifts the monograph's final statement from:

Performance depends on preserving admissible futures.

to:

The highest form of performance is enlarging the space of futures that can be conceived, described, and reached.

The complete progression is therefore:

$$\text{Prediction} \rightarrow \text{Overlap} \rightarrow \text{Reachability} \rightarrow \text{Admissibility} \rightarrow \text{Ontological Expansion}.$$

Prediction identifies a single future. Overlap preserves a family of futures. Reachability preserves access to futures. Admissibility preserves coherent futures. Ontological expansion enlarges the space of futures that can be conceived, described, and reached.

Chapter 13

Toward an Admissibility-Based Learning Objective

13.1 The Progression of Objectives

We have traced a progression:

$$\mathcal{L}_{\text{CE}} \;\to\; \mathcal{L}_{\text{KL}} \;\to\; \mathcal{L}_{\text{TV}} \;\to\; \mathcal{L}_{A}.$$

Each step moves the learning objective closer to the quantity that actually governs system performance.

$\mathcal{L}_{\text{CE}}$ Minimise prediction error over the training distribution.

$\mathcal{L}_{\text{KL}}$ Minimise expected log-likelihood ratio; concentrates on modes.

$\mathcal{L}_{\text{TV}}$ Minimise total variation; concentrates on overlap of next-token distributions.

$\mathcal{L}_{A}$ Minimise admissibility divergence; concentrates on overlap of admissible future sets.

13.2 The Admissibility Objective

We propose:

$$\mathcal{L}_A(p, q) = 1 - \mathcal{O}(p, q) = 1 - \frac{\text{Vol}\big(\mathcal{A}(p) \cap \mathcal{A}(q)\big)}{\text{Vol}\big(\mathcal{A}(p)\big)}. \tag{13.1}$$

This objective measures the fraction of the target model's admissible futures that are *not* admissible under the draft model. Minimising $\mathcal{L}_A$ trains the draft to preserve access to the target model's futures, not merely to reproduce its next-token probabilities.

13.3 Relationship to TV

TV distance is $\mathcal{L}_A$ restricted to the one-step, flat-simplex case:

$$\mathcal{L}_{\text{TV}}(p, q) = \mathcal{L}_A(p, q)\big|_{\text{one step, flat simplex}}.$$

This confirms that TV is a degenerate case of the admissibility objective, not an approximation to it. The Bebop paper's move from CE to TV is therefore a move *toward* admissibility-based learning, even though it does not reach the full generalisation.

13.4 Challenges in Computing $\mathcal{L}_A$

The admissibility objective requires:

1. A definition of $\mathcal{A}(p)$: which trajectories are admissible.
2. A volume measure on trajectory space.
3. An efficient method for computing or approximating the overlap.

None of these is straightforward. For language models, admissibility is a complex, implicitly defined set (semantic coherence, factual accuracy, task relevance, safety). Trajectory space is exponentially large. Overlap computation is generally intractable.

Several approximation strategies are available:

- **Local approximation**: approximate $\mathcal{A}(p)$ by a neighbourhood of the current trajectory under p, and compute TV overlap on the marginal distribution at each step. This recovers $\mathcal{L}_{\text{TV}}$.
- **Learned admissibility boundary**: train an auxiliary model to classify trajectories as admissible or inadmissible under p, and minimise the draft model's inadmissibility probability.
- **RSVP field approximation**: model $\mathcal{A}(p)$ as a level set of the scalar field Φ and compute overlap via field integrals.

13.5 Admissibility-Based RL

The admissibility objective also suggests a reformulation of the RL training loop itself. Rather than rewarding the policy for producing high-reward trajectories, one could reward the policy for maintaining high admissibility overlap with a reference policy:

$$r_A(p, p_{\text{ref}}) = \mathcal{O}(p_{\text{ref}}, p) = \frac{\text{Vol}\big(\mathcal{A}(p_{\text{ref}}) \cap \mathcal{A}(p)\big)}{\text{Vol}\big(\mathcal{A}(p_{\text{ref}})\big)}.$$

This is a *future-preservation reward*: it measures how well the current policy preserves the reference policy's admissible futures, not how well it imitates its token-level behaviour. This differs from KL-penalised RL (which penalises distributional divergence) by targeting admissibility structure rather than probability structure.

Chapter 14

The Reachability Principle: Statement and Consequences

14.1 Formal Statement

We now state the Reachability Principle in its most general form.

Principle 3 (Reachability Principle, General Form). Let S be an adaptive system with state space $\mathcal{X}$, admissible future set $A : \mathcal{X} \to 2^{\mathcal{T}}$, and performance metric $\mathcal{P}$. Then for any perturbation $\phi : \mathcal{X} \to \mathcal{X}$,

$$\mathcal{P}(S \circ \phi) \approx \mathcal{P}(S) \text{ when } \mathcal{O}\big(A(x),\, A(\phi(x))\big) \approx 1,$$

and $\mathcal{P}(S \circ \phi)$ may diverge from $\mathcal{P}(S)$ even when $\phi(x) \approx x$ if $\mathcal{O}\big(A(x),\, A(\phi(x))\big) \ll 1$.

In words: system performance is stable under perturbation when admissible future sets overlap, and may be unstable even under small perturbations when admissible future sets diverge.

14.2 Special Cases

Corollary 14.1 (Speculative Decoding)**.** *Draft acceptance is high when $\mathcal{O}(\mathcal{A}(p), \mathcal{A}(q)) \approx 1$, regardless of token-level agreement between p and q. TV overlap is the one-step simplex approximation to this condition.*

Corollary 14.2 (RL Stability)**.** *If RL updates are admissibility-preserving (preserve the topology of $\mathcal{A}(p)$), then draft-model mismatch is negligible and online retraining of auxiliary heads is unnecessary.*

Corollary 14.3 (Catastrophic Forgetting)**.** *A continual learning update causes catastrophic forgetting if and only if it collapses $\mathcal{O}(A_{\text{before}}, A_{\text{after}})$, regardless of whether it preserves token-level accuracy on old tasks.*

Corollary 14.4 (Repair)**.** *An entity x is successfully repaired to x' if and only if $\mathcal{O}(A(x), A(x')) \approx 1$, even when $x' \neq x$.*

14.3 Identity as a Future Invariant

The examples examined throughout this monograph suggest a common structure that has been implicit but never stated as a theorem.

Definition 14.5 (Future identity)**.** The *future identity* of a system at state x is its admissible future set:

$$I(x) = A(x).$$

A transformation $\phi : x \to x'$ preserves identity whenever $\mathcal{O}(A(x), A(x')) \approx 1$.

Under this definition, apparently diverse phenomena become instances of the same invariant:

- **Repair** preserves identity by maintaining future accessibility across substrate changes: $x' \neq x$ but $A(x') \approx A(x)$.
- **Continual learning** preserves identity when parameter updates leave admissible future structure substantially unchanged.
- **Institutional continuity** reflects persistence of reachable patterns of action and response, not persistence of particular constituents.
- **Archive rewrites** preserve identity because observations remain reconstructible through admissible paths despite changes in encoding.
- **Scientific progress** occupies a special position: it preserves continuity with prior structure while enlarging $\mathcal{R}_\Omega(T)$ — the reachable ontology set. It is the only domain in the list where future identity genuinely grows rather than merely persisting.

In every case, identity is not indexed by present state but by the preservation and extension of admissible futures. The Reachability Principle can therefore be restated in its most compressed form:

Principle 4 (Identity as admissibility invariant). Identity is an invariant of future accessibility, not of present state.

14.4 The Inversion

We close by marking the inversion that runs through this monograph.

Standard framing	**Reachability framing**
Intelligence ≈ prediction	Intelligence ≈ future preservation
Next token is the objective	Access to continuations is the objective
Search: find the optimal state	Navigation: stay in the admissible corridor
Entropy bounds performance	Admissible volume governs tracking difficulty
CE/KL optimise faithfulness	TV/admissibility optimise coverage
State preservation ⇒ identity	Future preservation ⇒ identity
Repair ≈ state restoration	Repair ≈ admissibility restoration
Explanation ≈ true description	Explanation ≈ deficit reduction
Science finds facts	Science enlarges reachable description space
Discovery preserves options	Discovery creates new options

None of the left-hand entries is wrong. Each is a degenerate case of the right-hand entry, exact in limiting conditions that rarely obtain in practice.

The Bebop paper occupies an interesting historical position. It begins with the left column (token prediction, draft accuracy) and arrives, through empirical pressure, at the right column (overlap geometry, TV training). It does not articulate the theoretical shift explicitly. This monograph has attempted to provide that articulation.

14.5 Open Questions

Several directions remain open:

1. **Admissible volume estimation**: practical methods for approximating $\mathrm{Vol}(\mathcal{A}(p))$ for large language models.
2. **Admissibility curvature**: formalisation and measurement of κ_A as a predictor of tracking difficulty.
3. **Multi-step admissibility loss**: extension of $\mathcal{L}_{\mathrm{TV}}$ to the full trajectory horizon.
4. **Admissibility-preserving RL**: training algorithms that explicitly regulate $\mathcal{O}(p_{\mathrm{ref}}, p)$ during fine-tuning.
5. **RSVP field construction**: methods for computing the scalar field Φ from transformer hidden states.
6. **Collapse-size distribution**: the rematching archive [1] conjectures a power-law distribution over collapse-event magnitudes for natural data with hierarchical generative structure. Verifying or refuting this conjecture — and characterising the classes of data for which it holds — would ground the punctuated-compression account in a testable statistical claim.
7. **Conservation of transformability**: whether high compression depth achieved through deep template nesting minimises remaining admissible paths (ontological lock-in) or maximises them (by extracting genuine regularities) is posed as Open Problem 7 of *Waves of Collapse*. Resolving it would connect the admissibility framework directly to ecological network resilience theory, where the same tension between efficiency and adaptability is quantitatively well studied.

Each of these is a tractable research question that follows directly from the framework developed here.

Bibliography

[1] Flyxion. Waves of Collapse: Compression as the Geometry of Reachable Description. Monograph. Part of the RSVP / Admissibility series, 2026. https://standardgalactic.github.io/playfloor/coordinates/waves-of-collapse.pdf

[2] Deming Chen, Jason Cong, Azalia Mirhoseini, Christos Kozyrakis, Subhasish Mitra, Jinjun Xiong, Cliff Young, Anima Anandkumar, Michael Littman, Aron Kirschen, Sophia Shao, Serge Leef, Naresh Shanbhag, Dejan Milojicic, Michael Schulte, Gert Cauwenberghs, Jerry M. Chow, Tri Dao, Kailash Gopalakrishnan, Richard Ho, Hoshik Kim, Kunle Olukotun, David Z. Pan, Mark Ren, Dan Roth, Aarti Singh, Yizhou Sun, Yusu Wang, Yann LeCun, and Ruchir Puri. AI+HW 2035: Shaping the Next Decade. arXiv:2603.05225 [cs.AI], 2026. https://arxiv.org/abs/2603.05225

[3] Yucheng Li, Huiqiang Jiang, Yang Xu, Jianxin Yang, Yi Zhang, Yizhong Cao, Yuhao Shen, Fan Zhou, Rui Men, Jianwei Zhang, An Yang, Bowen Yu, Bo Zheng, Fei Huang, Junyang Lin, Dayiheng Liu, and Jingren Zhou. Breaking Entropy Bounds: Accelerating RL Training via MTP with Rejection Sampling. arXiv:2606.12370 [cs.LG], 2026. https://arxiv.org/abs/2606.12370

[4] Yaniv Leviathan, Matan Kalman, and Yossi Matias. Fast Inference from Transformers via Speculative Decoding. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*. PMLR, 2023.

[5] Charlie Chen, Sebastian Borgeaud, Geoffrey Irving, Jean-Baptiste Lespiau, Laurent Sifre, and John Jumper. Accelerating Large Language Model Decoding with Speculative Sampling. arXiv:2302.01318 [cs.LG], 2023.

[6] Fabian Gloeckle, Badr Youbi Idrissi, Baptiste Rozière, David Lopez-Paz, and Gabriel Synnaeve. Better & Faster Large Language Models via Multi-Token Prediction. arXiv:2404.19737 [cs.LG], 2024.

[7] DeepSeek-AI. DeepSeek-V3 Technical Report. arXiv:2412.19437 [cs.CL], 2024.

[8] Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, Y.K. Li, Y. Wu, and Daya Guo. DeepSeekMath: Pushing the Limits of Mathematical Reasoning in Open Language Models. arXiv:2402.03300 [cs.CL], 2024.

[9] Claude E. Shannon. A Mathematical Theory of Communication. *Bell System Technical Journal*, 27(3):379–423, 1948.

[10] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley-Interscience, 2nd edition, 2006.

[11] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal Policy Optimization Algorithms. arXiv:1707.06347 [cs.LG], 2017.

[12] Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. Training Language Models to Follow Instructions with Human Feedback. arXiv:2203.02155 [cs.CL], 2022.

[13] Vladimir N. Vapnik. An Overview of Statistical Learning Theory. *IEEE Transactions on Neural Networks*, 10(5):988–999, 1999.

[14] Cédric Villani. *Optimal Transport: Old and New*. Volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer, 2009.

[15] Alfred North Whitehead. *Process and Reality: An Essay in Cosmology*. Macmillan, 1929.

[16] W. Ross Ashby. *An Introduction to Cybernetics*. Chapman & Hall, 1956.

[17] Ilya Prigogine and Isabelle Stengers. *Order Out of Chaos: Man's New Dialogue with Nature*. Bantam Books, 1984.

[18] Karl J. Friston. The Free Energy Principle: A Unified Brain Theory? *Nature Reviews Neuroscience*, 11:127–138, 2010.