The Substance and Accident of Reasoning

Admissibility, Repair, and the Geometry of Verbosity

Flyxion

Independent Researcher

June 2026

Abstract

The central result of *Shorter but not Worse* [1] is often presented as an efficiency finding: models trained on a curriculum containing moderately easy mathematical problems produce substantially shorter solutions while preserving or improving reasoning performance. Yet the deeper significance of the result is not computational but geometric. The paper reveals a form of representational collapse in which an accidental property of successful reasoning becomes mistaken for an essential one. This essay analyses the failure through the lens of admissibility theory. A representation is admissible when it preserves distinctions required for future action. The hard-only training curriculum induces inadmissibility by systematically suppressing evidence that correct solutions can be short. The resulting policy identifies correctness with length—a projection collapse in which two properties that ought to be orthogonal become conflated. The easy-problem repair restores admissibility not by adding information but by expanding the fiber of successful trajectories to include compact ones. The information-theoretic argument in the paper explains why verbosity emerges under length-unpenalised training but misidentifies the goal: entropy reduction is not intelligence. The verbosity error is one of three instances of the same deeper mistake—the identification of a proxy metric with the target itself—alongside the latent-compression error (lower dimensionality as better representation) and the reasoning-length error (more steps as deeper thought). All three are projection collapses in which accidental properties of training distributions are mistaken for substantial properties of the task. The solution in each case is not additional optimisation pressure but representational repair: introducing training evidence that demonstrates the admissibility of the target without the proxy.

1. The Finding and Its Conventional Interpretation

Reinforcement learning pipelines for mathematical reasoning frequently discard easy problems on the grounds that they provide little optimisation signal. When all rollouts for a given problem are correct, the group relative advantage collapses to zero and the problem contributes no gradient. Training therefore concentrates on problems of intermediate and high difficulty whose successful solutions are typically long. Over time, reward becomes correlated with verbosity, and the model learns to associate correctness with length.

Bounhar et al. demonstrate that reintroducing moderately easy problems reverses this tendency. Models trained under the mixed curriculum produce solutions that are nearly half the length of baseline solutions, with no loss in accuracy on competition-level mathematics and improvements on several benchmarks. The intervention requires no explicit length penalty. The brevity is emergent.

The conventional interpretation frames this as a training efficiency result: easy problems act as implicit length regularisers by providing stable reward signals associated with short outputs. This framing is accurate as far as it goes. But it does not explain what has actually gone wrong in the baseline model or why a particular form of data augmentation specifically repairs it. To answer these questions requires a different theoretical vocabulary.

2. Representational Collapse

A representation is *admissible* when it preserves distinctions required for future action. It becomes inadmissible when states with different consequences are projected into the same representational region. The hard-only curriculum creates precisely such a projection.

Consider two solutions to a given mathematical problem:

$$s_1 = \text{short correct proof}, \qquad s_2 = \text{long correct proof}.$$

The downstream consequence of both trajectories is identical: the problem is solved and reward 1 is received. Yet the training distribution systematically exposes the model to successful trajectories of the second type while suppressing examples of the first. The learned representation therefore begins to identify correctness with length. Symbolically, the model's implicit mapping satisfies

$$F(\text{correct}) \approx F(\text{long})$$

rather than the admissible condition

$$F(\text{correct}) \perp F(\text{length}).$$

The distinction between solution quality and solution verbosity has collapsed.

This is not merely a statistical artifact. It is a loss of navigational structure. A planner operating on such a representation can no longer distinguish between properties that affect reachability and properties that merely accompany successful trajectories. The resulting policy cannot reliably separate what is necessary from what is incidental.

3. The Projection Operator and Its Distorted Fibers

The CLIO framework offers a precise geometric language for this failure. The hard-only curriculum induces a projection operator

$$\pi : (\text{quality, length}) \longrightarrow \text{quality}$$

whose empirical fibers become distorted. In an undistorted representation, the pre-image of good solutions would contain both short and long trajectories:

$$\pi^{-1}(\text{good}) = \{\text{short trajectories}\} \cup \{\text{long trajectories}\}.$$

Under the hard-only curriculum, the empirical support of this fiber contracts:

$$\pi^{-1}(\text{good}) \approx \{\text{long trajectories}\}.$$

The fiber's support has collapsed. What was a two-dimensional region in the quality-length product space has degenerated: the policy assigns negligible probability mass to the short-trajectory region of the fiber, placing it effectively outside the reach of the learned representation.

The easy-problem repair operates directly on this support collapse. By introducing successful trajectories that are both correct and compact, it repopulates the short-trajectory region of $\pi^{-1}(\text{good})$:

$$\pi^{-1}(\text{good}) \longleftarrow \pi^{-1}(\text{good}) \cup \{\text{short correct trajectories}\}.$$

The fiber's support expands and recovers its extent across the quality-length space. What appeared to be an intrinsic feature of successful reasoning—length—is revealed to be merely one coordinate among many. The projection no longer collapses the distinction between quality and verbosity.

Definition 3.1 (Admissible Reasoning Representation)**.** A policy's internal representation of solution strategies is *admissible* if the fibers of the correctness projection contain both short and long correct trajectories whenever both exist for a given problem class. Admissibility fails when the fiber contracts to a single strategy type under training pressure.

The baseline verbosity-trained model is inadmissible in this sense. The frugal model, after Stage 1 repair, is admissible: it can locate correct solutions across the full range of trajectory lengths and select among them based on problem structure.

4. Entropy Reduction Is Not Intelligence

The information-theoretic argument in the paper identifies a genuine mechanism underlying verbosity. By the chain rule of conditional entropy:

$$H(Y \mid X, Z_{t+1}) \;\leq\; H(Y \mid X, Z_t),$$

where Y is the final answer and Z_t is the length-t prefix produced by the autoregressive policy. Conditioning on a longer prefix can only decrease or preserve the conditional entropy of the final answer. This holds regardless of whether the additional tokens carry semantic content. In the absence of any length penalty, a policy rewarded for correct final answers has a weak statistical incentive to extend its output: verbosity becomes a shortcut for entropy reduction.

This observation is correct but incomplete. It identifies why verbosity emerges. It does not identify what is wrong with it.

Entropy reduction is not itself intelligence.

A language can reduce uncertainty by merging words—collapsing distinct meanings into a single form so that the distribution over interpretations narrows. A map can reduce uncertainty by erasing roads —reducing the number of possible routes so that the next turn becomes predictable. A bureaucracy can reduce uncertainty by prohibiting choices—constraining the action space until outcomes are fully determined.

In each case the system becomes more predictable. In each case it simultaneously becomes less useful.

The collapse of distinct meanings into a single form makes disambiguation harder. The erasure of roads reduces navigational capacity. The prohibition of choices destroys the agency the system was designed to support.

The relevant quantity is not entropy but admissibility.

The question is not whether uncertainty decreases after a longer prefix. The question is whether the distinctions necessary for navigation survive. A model that reduces uncertainty by making all successful solutions look long has reduced entropy at the cost of admissibility. It has made itself more predictable while simultaneously making itself less capable of producing the full range of correct solutions its task requires.

From this perspective, the model's verbosity resembles a language undergoing semantic merger. Distinct reasoning trajectories—short correct proofs and long correct proofs—are collapsed into a single category whose defining feature is length.

The representation becomes easier to optimise while simultaneously becoming less faithful to the structure of the task it is meant to solve.

5. Three Instances of the Same Error

The verbosity failure belongs to a class of machine learning failures that share a common structure. In each case, a proxy metric that correlates with the target in the training distribution becomes identified with the target itself. The identification holds as long as the training distribution sustains it. When the distribution changes, or when the model is deployed outside its training regime, the identification breaks and the failure becomes visible.

Three instances of this error are worth placing alongside each other.

Latent compression as intelligence. The latent fundamentalist mistake is to identify a low-dimensional representation with a better representation [4]. A model trained to compress input into a compact latent space learns that lower dimensionality correlates with generalisation in its training regime. The identification holds when the training distribution is well-covered and the latent structure captures genuine variation. It breaks when the latent projection discards variation that matters for downstream tasks—when it collapses distinctions required for navigation.

The formal structure is:

$$F(\text{general representation}) \approx F(\text{low-dimensional}),$$

when the admissible condition is

$$F(\text{general representation}) \perp F(\text{dimensionality}).$$

Entropy reduction as intelligence. The verbosity error is to identify lower conditional entropy with better reasoning. A model trained without length penalties learns that longer prefixes reduce uncertainty about the final answer. The identification holds when the training distribution consists primarily of hard problems where length genuinely correlates with correctness. It breaks when short correct solutions exist but have been excluded from the training distribution, leaving the model unable to produce them.

The formal structure is:

$$F(\text{correct reasoning}) \approx F(\text{low entropy}),$$

when the admissible condition is

$$F(\text{correct reasoning}) \perp F(\text{length}).$$

Long chains of thought as intelligence. The reasoning length error is the most culturally embedded of the three. Contemporary discourse about AI progress frequently equates extended chains of thought with deeper reasoning. Test-time compute scaling is discussed as if more computation were always better computation. A model that produces ten reasoning steps is presumed to be engaging in more careful reasoning than one that produces two, even when the two-step solution is correct and the ten-step solution is padded.

The formal structure is again the same:

$$F(\text{good reasoning}) \approx F(\text{many steps}),$$

when the admissible condition is

$$F(\text{good reasoning}) \perp F(\text{step count}).$$

The common structure. All three errors have the same form: a genuine correlation in the training distribution is mistaken for an identity. In the distribution where the model learned, the proxy and the target covary. Outside that distribution, they diverge.

The admissibility framework names this structure: it is a projection collapse. A property of trajectories that is accidental in the philosophical sense—contingent on the distribution, not constitutive of the target—has been treated as substantial. The repair in each case is the same: introduce training examples that demonstrate the admissibility of the target without the proxy. Short correct solutions for verbosity. High-dimensional correct representations for latent compression. Two-step correct proofs for reasoning length.

What makes these failures persist is that the proxies are not arbitrary. In many distributions, lower entropy does correlate with better reasoning. In many distributions, lower dimensionality does correlate with better generalisation. In many distributions, longer chains of thought do correlate with correct answers. The proxies are genuine signals. They become dangerous only when the training distribution is not representative of the full range of cases where the target behaviour is admissible.

The model is not failing to learn. It is succeeding in learning from the data it has. The data is the problem.

The common algebraic skeleton of all three cases can be stated as a single proposition.

Proposition 5.1 (Proxy Collapse and Repair)**.** *Let T be a target property and Q a proxy such that* $\mathrm{Corr}(T, Q) \approx 1$ *in the training distribution $\mathcal{D}$. A* proxy collapse *occurs when the learner replaces the distributional co-occurrence*

$$\Pr(T \mid Q) \approx 1$$

with the stronger identification $T \equiv Q$, treating Q as constitutive of T rather than merely correlated with it.

The collapse is repaired by introducing training examples in which $T \wedge \neg Q$ or $Q \wedge \neg T$ holds—examples that demonstrate the non-identity of T and Q directly.

Under this schema, the three cases reduce to:

Error	***Collapse***	***Repair signal***
Latent compression	*general $\equiv$ low-dim*	*high-dim correct representations*
Verbosity	*correct $\equiv$ long*	*short correct solutions*
Reasoning length	*good $\equiv$ many steps*	*two-step correct proofs*

In each case, the repair is not additional optimisation pressure but evidence of admissible trajectories in the region the training distribution had rendered sparse.

High correlation does not imply admissible substitution.

6. Morphological Repair

The easy-problem intervention is not a regulariser in the standard sense of adding a constraint to the objective function. It is a representational repair that restores admissible structure without modifying the reward function.

This distinction matters because it points to a different theory of what went wrong. A regulariser corrects an objective. A repair corrects a representation. The baseline model's failure is not that it is optimising the wrong objective—it is correctly maximising expected reward under the verifiable binary reward function. The failure is that its learned representation no longer faithfully distinguishes between strategies whose downstream consequences differ.

The repair mechanism in admissibility theory takes the following form. When a representation undergoes collapse by identifying two distinct states s_1 and s_2 with the same representational region, the minimal repair introduces a residual distinction that separates the two states without abandoning their shared structure:

$$\widetilde{F}(s_1) = (a, f), \qquad \widetilde{F}(s_2) = (b, f), \qquad a \neq b.$$

The shared component f preserves whatever the two states genuinely have in common (both are correct solutions). The residual components $a \neq b$ restore the distinction that collapse had erased (one is short, one is long; they are appropriate in different contexts).

Easy problems supply the $a \neq b$ signal. They are the training examples for which s_1 (short correct solution) receives positive reward and any attempt to produce s_2 (long solution) is either truncated or wasteful. The group relative advantage assigns positive weight to s_1 on easy problems in a way that the hard-only curriculum never could, because hard problems genuinely require s_2. The repair is morphological in the sense that it adds a minimal distinguishing marker—easy-problem context—to the learned representation, not by rewriting the underlying reward structure but by extending the evidence base from which the policy infers which strategy to select.

Theorem 6.1 (Easy Problems as Minimal Repair)**.** *Under hard-only training, the policy's learned representation satisfies $F(s_1) \approx F(s_2)$ for short and long correct solutions respectively. Reintroducing easy problems provides training examples on which s_1 is rewarded and s_2 is penalised by truncation or wasted budget. The GRPO advantage thereby assigns differential gradient to s_1 and s_2 on easy problems, progressively separating the two strategies in representation space. The repair is minimal: it does not alter the representation of hard-problem solutions but restores the short-solution fiber of the correctness projection.*

Proof. By Theorem 7.1, hard problems contribute zero differential gradient between s_1 and s_2. By the budget-truncated reward mechanism (formalised in Section 7), easy problems produce positive reward for s_1 and zero or negative effective reward for s_2 when s_2 exceeds the token budget or wastes allocation that could serve harder problems. The GRPO advantage on easy problems is therefore strictly positive for s_1 and non-positive for s_2. Gradient descent under this signal separates s_1 from s_2 in the policy's strategy representation. The separation is local to the easy-problem regime: on hard problems the policy continues to produce s_2 as before, since no new signal changes the hard-problem advantage. The repair is therefore minimal in the sense that it modifies the representation only in the short-solution fiber, not throughout $\mathcal{S}$. □

A penalty-based approach to the same problem would compress the long-solution region without expanding the short-solution region. It makes s_2 costly without demonstrating that s_1 is admissible. If the model has never produced short correct solutions under training, penalising long solutions provides no gradient toward finding them. The repair approach works by the opposite mechanism: it demonstrates that s_1 is correct before the model has internalised that it cannot be.

7. The Geometry of Advantage Collapse

The GRPO advantage function is the formal site where projection collapse manifests. We derive its behaviour under hard-only versus mixed distributions and show that the hard-only curriculum is structurally incapable of separating short from long correct trajectories.

7.1. Advantage blindness under hard-only training

For a group of G rollouts $\{y_i\}_{i=1}^G$ on query x, the GRPO advantage is

$$A_i = \frac{r(x, y_i) - \bar{r}}{\sigma_r},$$

where $\bar{r}$ and σ_r are the group mean and standard deviation of rewards. Under binary rewards $r \in \{0, 1\}$, if k rollouts are correct, $\bar{r} = k/G$ and $\sigma_r = \sqrt{k(G-k)}/G$.

For a hard problem with per-rollout success probability $p \in (0, 1)$:

$$\mathbb{E}[A_i \mid r_i = 1] = \sqrt{\frac{1-p}{p}}, \qquad \mathbb{E}[A_i \mid r_i = 0] = -\sqrt{\frac{p}{1-p}}.$$

This advantage is assigned identically to all correct rollouts regardless of length. For any two correct rollouts y_{short} and y_{long}:

$$A_{\text{short}} = A_{\text{long}} = \sqrt{\frac{1-p}{p}}.$$

Theorem 7.1 (Advantage Blindness to Length)**.** *Under the hard-only distribution, the GRPO advantage provides zero differential signal between correct trajectories of different lengths. For any y_1, y_2 with $r(x, y_1) = r(x, y_2) = 1$,*

$$A_1 - A_2 = 0$$

regardless of $|\ell(y_1) - \ell(y_2)|$.

Proof. $A_i = (r_i - \bar{r})/\sigma_r$ depends only on r_i and group statistics $(\bar{r}, \sigma_r)$, which are functions of $\{r_j\}_{j=1}^G$ alone. Since $r(x, y_1) = r(x, y_2) = 1$, both rollouts contribute identically to all group statistics, so $A_1 = A_2$. □

Remark. Theorem 7.1 is a structural result: the advantage estimator cannot distinguish any property of a rollout beyond its scalar reward. The gradient is therefore informationally incapable of teaching the policy to prefer short correct solutions over long ones when both exist. Repair must come from the training distribution, not the objective function.

7.2. Advantage asymmetry under the mixed distribution

Define the budget-truncated reward for budget B:

$$r^B(x, y) = \begin{cases} 1 & \text{if } r(x, y) = 1 \text{ and } \ell(y) \leq B, \\ 0 & \text{otherwise.} \end{cases}$$

Theorem 7.2 (Effective Advantage Asymmetry Under Budget Constraints)**.** *Under the budget-truncated reward r^B, the combined effect of the training distribution and the context limit produces an effective advantage asymmetry on easy problems: when the fraction $q > 0$ of rollouts exceeds budget B, the effective learning signal satisfies*

$$\mathbb{E}[A_{\text{short}}] > \mathbb{E}[A_{\text{long}}].$$

This is a property of the training distribution and budget constraint jointly, not of the GRPO objective in isolation.

Proof. Rollouts exceeding B receive $r^B = 0$ by the truncation rule. The group mean becomes $\bar{r} = k/G - q < k/G$ where k counts pre-truncation correct rollouts. A short correct rollout receives reward 1, advantage $(1 - \bar{r})/\sigma_{r^B} > 0$. A truncated long rollout receives reward 0, advantage $-\bar{r}/\sigma_{r^B} < 0$. Since $q > 0$ by hypothesis, the expected advantage differential is strictly positive: $\mathbb{E}[A_{\text{short}}] - \mathbb{E}[A_{\text{long}}] > 0$. The asymmetry arises from the conjunction of (i) the fixed budget B that converts long rollouts into truncation events, and (ii) the easy-problem distribution that ensures short correct rollouts exist within B. Neither factor alone suffices. □

Together, Theorems 7.1 and 7.2 characterise the repair mechanism: hard problems provide zero differential gradient between short and long correct solutions; easy problems provide positive differential gradient favouring short solutions. The mixed distribution has gradient structure the hard-only distribution structurally lacks.

8. Optimality of Repair over Penalty

8.1. The penalty approach and its limit

A length-penalised reward $r_\lambda(x, y) = r(x, y) - \lambda\ell(y)$ yields advantage differential

$$A^\lambda_{\text{short}} - A^\lambda_{\text{long}} = \frac{\lambda(\ell_{\text{long}} - \ell_{\text{short}})}{\sigma_{r_\lambda}} > 0.$$

This creates gradient toward shorter solutions. It appears to solve the problem. But it has an information-theoretic gap.

Definition 8.1 (Admissibility Evidence)**.** Training example (x, y) provides *admissibility evidence* for strategy s if $r(x, y) = 1$ and y instantiates s.

Theorem 8.2 (Information Gap of Length Penalties)**.** *Under hard-only training with any $\lambda > 0$, admissibility evidence for s_{short} is zero if the policy has not yet discovered short correct solutions. Under the mixed distribution with easy problems, admissibility evidence for s_{short} is positive regardless of the policy's prior behaviour.*

Proof. Admissibility evidence for s_{short} requires (x, y) with $r(x, y) = 1$ and $\ell(y) \ll B$. Under hard-only training, successful rollouts satisfy $\ell(y) \approx B$. If the policy has not yet produced short correct solutions, no rollout in the hard-only distribution satisfies both conditions. Adding λ to the reward does not change which rollouts the policy produces; it only reweights rewards. Therefore admissibility evidence is zero for any λ.

Under the mixed distribution, easy problems admit short correct rollouts with $r = 1$ and $\ell \ll B$ with positive probability even under a policy not yet specialised to brevity. □

Corollary 8.3 (Repair Dominates Penalty in Early Training)**.** *Before the policy discovers short correct solutions, repair provides strictly more information about s_{short} than any penalty-augmented hard-only distribution.*

Length penalties tell the policy that long solutions are costly. Easy problems show the policy that short solutions are correct. These are different pieces of information, and in the early training phase only the second is available.

Corollary 8.4 (Temporal Ordering)**.** *Repair is the appropriate early-phase intervention. Penalty augmentation may consolidate an established brevity preference in a later phase. Penalty before repair risks suppressing exploration of s_{short} by making long solutions costly without revealing that short ones exist.*

9. Fiber Support and Stability of the Repaired Representation

9.1. Formal fiber structure and the admissibility gap

Let the policy strategy space decompose as $\mathcal{S} = \mathcal{Q} \times \mathcal{L}$ with correctness projection $\pi(q, \ell) = q$. The admissible fiber at $q^* =$ correct is $\mathcal{F}^* = \{q^*\} \times \mathcal{L}$. The policy's empirical fiber $\hat{\mathcal{F}}$ contracts under hard-only training:

$$\hat{\mathcal{F}}_{\text{hard}} \approx \{q^*\} \times [\ell_{\min}^{\text{hard}}, B], \qquad \ell_{\min}^{\text{hard}} \gg 0,$$

and expands after repair:

$$\hat{\mathcal{F}}_{\text{repaired}} \approx \{q^*\} \times [\ell_{\min}^{\text{easy}}, B], \qquad \ell_{\min}^{\text{easy}} \ll \ell_{\min}^{\text{hard}}.$$

Definition 9.1 (Fiber Admissibility Gap)**.** $\Delta\mathcal{F} = \mu(\mathcal{F}^* \setminus \hat{\mathcal{F}})$, where μ weights lengths by their frequency in problem-appropriate solutions. Admissibility failure is $\Delta\mathcal{F} > 0$.

9.2. Stability of the repaired fiber

Theorem 9.2 (Stability Condition)**.** *Let ρ_{easy} be the fraction of easy problems in the ongoing training distribution and $\delta > 0$ the threshold gradient differential needed to sustain the short-solution region. The repaired fiber is stable if and only if*

$$\rho_{\text{easy}} \;\geq\; \rho^* := \frac{\delta}{\mathbb{E}_{x_e}\left[A_{\text{short}} - A_{\text{long}} \mid x \textit{ easy}\right]}.$$

Proof. By Theorem 7.1, hard problems contribute zero differential gradient to the short-solution fiber. By Theorem 7.2, each easy problem contributes expected differential $\mathbb{E}[A_{\text{short}} - A_{\text{long}}] > 0$. The net stabilising contribution is $\rho_{\text{easy}} \cdot \mathbb{E}[A_{\text{short}} - A_{\text{long}}]$. Stability requires this to exceed δ, giving the stated bound. □

Corollary 9.3 (A Testable Prediction)**.** *Setting $\rho_{\text{easy}} = 0$ in Stage 2 (eliminating easy problems entirely) violates the stability condition. The repair framework predicts that extended Stage 2 training on hard-only data will eventually regenerate verbosity, with re-collapse rate determined by δ and the average advantage differential on easy problems. This prediction is testable against extended training runs.*

9.3. Monotone correspondence with reachability loss

The fiber admissibility gap connects to the reachability loss Λ^θ of the companion framework [2]. The merged strategy $s_{\text{short}} \sqcup s_{\text{long}}$ is the undifferentiated policy for "correct solution", and reachability loss measures deployment contexts in which the merged strategy is inferior to the differentiated one:

$$\Lambda^\theta(s_{\text{short}}, s_{\text{long}}) = \mu^\theta(\mathcal{A}^\theta_{s_{\text{short}}} \cup \mathcal{A}^\theta_{s_{\text{long}}}) - \mu^\theta(\mathcal{A}^\theta_{s_{\text{short}} \sqcup s_{\text{long}}}).$$

Proposition 9.4 (Monotone Correspondence)**.** *The fiber admissibility gap and reachability loss are monotonically related: a larger gap cannot correspond to smaller loss. Formally, for two policies with fiber gaps $\Delta\mathcal{F}_1 < \Delta\mathcal{F}_2$,*

$$\Lambda^\theta_1 \;\leq\; \Lambda^\theta_2.$$

In particular, $\Delta\mathcal{F} = 0$ implies $\Lambda^\theta = 0$.

Proof. $\Delta\mathcal{F}_2 > \Delta\mathcal{F}_1$ means the second policy excludes a strictly larger region of $\mathcal{F}^*$ from its empirical fiber. Futures in $\mathcal{F}^* \setminus \hat{\mathcal{F}}_2$ but not in $\mathcal{F}^* \setminus \hat{\mathcal{F}}_1$ are reachable under the first policy's differentiated strategy but not under the second's. These futures contribute to Λ^θ_2 but not to Λ^θ_1. Therefore $\Lambda^\theta_2 \geq \Lambda^\theta_1$.

For the special case $\Delta\mathcal{F} = 0$: the empirical fiber equals the admissible fiber, so the differentiated and merged strategies access the same futures, giving $\Lambda^\theta = 0$. □

Remark. The monotone relation is sufficient for the essay's main claims and easier to defend than exact equivalence. The two quantities are defined over different formal objects—fiber gaps live in strategy-length product space, reachability losses live in admissible-futures measure space—and their relationship is conceptual as well as formal. Monotonicity captures the connection precisely without overclaiming identity.

Stage 1 achieves $\Delta\mathcal{F} \approx 0$ and $\Lambda^{\theta} \approx 0$ within the easy-problem domain. Stage 2 maintains this while extending the policy into harder domains, provided Theorem 9.2's condition holds.

10. Two-Stage Repair and Capability Expansion

The two-stage curriculum structure of the paper instantiates a general principle of admissibility-based learning: repair before expansion.

Stage 1 (emergent brevity) performs the morphological repair described above. Easy problems restore the short-solution fiber of the correctness projection. By the end of Stage 1, the model has learned that length and quality are orthogonal dimensions of a correct solution, not a single dimension.

Stage 2 (curriculum RLVR) performs capability expansion. Training moves progressively from moderately difficult to highly difficult problems. Crucially, Stage 2 does not reintroduce the collapse that Stage 1 repaired. The reason is structural: the model now has a representation in which correctness and length are separable. When it encounters a hard problem that genuinely requires a long solution, it extends its output because the problem requires it, not because length is constitutive of correctness. The policy can modulate solution length by problem structure rather than by a learned association with a spurious proxy.

This sequence—repair first, then expand—is the correct order of operations for any learning system that has undergone representational collapse. Expansion before repair reintroduces the collapse, because the harder domain's genuinely long solutions will again dominate the training signal and contract the short-solution fiber. The paper's empirical observation that Stage 2 brevity is preserved—minimum response length increases modestly but average length does not revert to baseline—confirms that the repair has been durable.

The two stages also correspond to two distinct forms of admissibility maintenance. Stage 1 restores local faithfulness within the current domain. Stage 2 extends admissibility into a new domain without sacrificing the local faithfulness that Stage 1 established. Together they implement what the domain divergence account of technical vocabulary emergence predicts: a representation that is stable in one domain can be extended into a harder domain without collapse, provided the extension begins from the edge of the already-stable region rather than from scratch.

11. Scientific Repair and the Broader Pattern

The repair structure observed in frugal reasoning is not specific to language models. It recurs across domains where a representation has undergone collapse under pressure from a biased sample of evidence.

A scientific anomaly repairs an ontology by forcing a distinction between concepts previously treated as identical. Newtonian mechanics treated mass and weight as proportional under terrestrial conditions, which they are; the anomaly of inertial versus gravitational mass forced the distinction open and revealed it to be theoretically fundamental. The anomaly did not add new information to the domain. It repopulated a previously empty fiber.

A biological mutation repairs an ecological niche by reopening a previously inaccessible trajectory. A lineage that has become specialised for one food source has undergone a representational collapse of sorts—its phenotype no longer distinguishes between environments where the specialised food is available and environments where it is not. A mutation that restores dietary flexibility expands the fiber of viable environments, restoring admissibility without changing the organism's fundamental architecture.

A mathematical counterexample repairs a theorem by revealing a hidden distinction. Cauchy's original proof of a theorem about continuous functions treated pointwise and uniform continuity as equivalent for the domain of interest; the counterexample forced the distinction open and the theorem was restated correctly. The counterexample did not contradict the original theorem under its intended interpretation. It demonstrated that the intended interpretation had collapsed a distinction the theorem required.

In every case, repair consists not of adding information but of restoring admissible structure by demonstrating that a previously empty fiber is non-empty. The easy mathematical problems are the counterexamples that demonstrate the non-emptiness of the short correct solution fiber. The repair is not a constraint. It is a demonstration.

12. The Broader Lesson for Machine Learning

The frugal reasoning result suggests a reframing of a class of machine learning failures that are conventionally attributed to insufficient data, insufficient model capacity, or insufficient optimisation pressure.

Many such failures may instead arise from projection collapses induced by unrepresentative training distributions. The model is not failing to learn from the data it has. It is succeeding in learning from the data it has, and the data happens to collapse a distinction the task requires.

In such cases, the correct intervention is not more data of the kind already present, not a stronger optimiser, and not a modified objective function. The correct intervention is data that repopulates the collapsed fiber—examples that demonstrate the admissibility of the strategy the model has learned to exclude.

This principle applies beyond reasoning length. A model trained on authoritative prose will conflate confidence with correctness. A model trained on long arguments will conflate thoroughness with validity. A model trained on formal language will conflate register with precision. In each case, the model has undergone a projection collapse: an accidental property of the training distribution has been identified with an essential property of the task.

The repair in each case follows the same structure as the frugal reasoning intervention. Introduce examples that demonstrate the admissibility of the strategy the model has collapsed: confident prose that is wrong, short arguments that are valid, informal language that is precise. The collapse is repaired not by modifying the objective but by expanding the fiber.

The lesson of *Shorter but not Worse* is therefore not that reasoning should be shorter. The lesson is that efficient reasoning emerges when a system learns which distinctions matter.

Length is not reasoning.

Length is merely one trajectory through the space of reasoning.

A model becomes capable not when it learns to think longer, but when it learns that thinking longer and thinking better are not the same thing—and that the difference between them is exactly the distinction that a well-curated training distribution must preserve.

13. Conclusion

Verbosity in reasoning models is a projection collapse. The hard-only training curriculum distorts the fiber of the correctness projection, contracting it from the full range of correct-solution lengths to a region concentrated on long trajectories. The model does not fail to optimise its objective; it succeeds too well, learning a representation in which the accidental property of training-distribution solutions—their length—has become identified with the essential property of correct reasoning.

The information-theoretic account of why verbosity emerges is correct: longer prefixes monotonically reduce conditional entropy of the final answer, and a length-unpenalised policy has a weak incentive to exploit this. But entropy reduction is not the goal. The goal is admissibility: producing representations that keep the right futures reachable. A model that reduces entropy by making all successful solutions look long has made itself more predictable while simultaneously reducing its navigational capacity.

The easy-problem repair restores admissibility by repopulating the short-solution fiber of the correctness projection. It is not a regulariser but a demonstration: it shows the model that short solutions are admissible, that they lie within the fiber of correct reasoning rather than outside it. The two-stage curriculum then extends this repaired representation into harder domains without reintroducing the collapse, because the policy has learned to modulate solution length by problem structure rather than by a spurious association with quality.

The pattern—collapse, repair, extension—is not specific to language models. It recurs wherever a biased sample of evidence causes a representation to treat an accidental property as constitutive. Scientific anomalies, biological mutations, and mathematical counterexamples all function as repair operators in this sense: they demonstrate the non-emptiness of a fiber that an accumulated evidence base had incorrectly treated as empty.

The measure of an intelligent system is not its ability to reduce uncertainty. It is its ability to preserve the distinctions that matter.

References

[1] Bounhar, A., Abdine, H., Dufraisse, E., Chamma, A., Mohamed, A., Bouch, D., Vazirgiannis, M., & Shang, G. (2026). Shorter but not worse: Frugal reasoning via easy samples as length regularizers in math RLVR. *arXiv:2511.01937v2*.

[2] Flyxion (2026). Why distinctions survive: A reachability theory of lexical preservation. *Independent Research Monograph*.

[3] Flyxion (2026). The persistence hierarchy: Adjective order as reachability gradient. *Independent Research Monograph*.

[4] Flyxion (2026). Against latent fundamentalism: Admissibility and the limits of dimensionality reduction. *Independent Research Monograph*.

[5] Shao, Z., Wang, P., Zhu, Q., Xu, R., Song, J., Bi, X., Zhang, H., Zhang, M., Li, Y. K., Wu, Y., & Guo, D. (2024). DeepSeekMath: Pushing the limits of mathematical reasoning in open language models. *arXiv:2402.03300*.