UNSEEN IS NOT NEW: PARTIAL IDENTIFICATION OF CAPABILITY EXPANSION IN REINFORCEMENT LEARNING WITH VERIFIABLE REWARDS

Anonymous authors
Paper under double-blind review

ABSTRACT

Reinforcement learning with verifiable rewards (RLVR) can expose reasoning behaviors that were unseen in a finite base-model sample. However, this observation does not establish capability expansion: a mode with zero probability and one with small positive probability are arbitrarily difficult to distinguish at any finite rollout budget. To address this ambiguity, we develop a probability-spectrum account of what checkpoint evaluations can identify. First, absolute continuity of response laws transports through semantic auditing, yielding support constraints and sharp divergence-to-visibility envelopes. Next, we show that expected discovery has a Hausdorff moment representation: its infinite curve identifies the positive mode-mass spectrum, whereas a finite prefix determines a sharp set of compatible future curves. Consequently, checkpoint rankings can reverse with the rollout budget, and semantic coarsening can conceal opposing probability flows. We connect these results to simultaneous probability-change regions and an access/execution analysis that separates natural conditioning from assigned interventions. Finally, exact certificates and matched neural checkpoints demonstrate identifiable amplification and a Transformer budget crossover. Under the verified full-softmax sampler, architectural support is equal even when observed sets differ. These results support capability claims about probability reallocation and operational visibility without treating finite nonobservation as evidence of structural novelty.

1 INTRODUCTION

Reinforcement learning with verifiable rewards (RLVR) can sharply increase the chance that a language model solves a mathematical or programming problem in one attempt (Shao et al., 2024; DeepSeek-AI, 2025). However, the interpretation of this gain remains unsettled. On one side of this debate, large-sample studies recover post-training answers or reasoning patterns from base checkpoints, suggesting probability reallocation (Yue et al., 2025; Wu et al., 2025). On the other side, studies report expansion under process-aware evaluation, prolonged training, guided exploration, or curriculum construction (Wen et al., 2025; Yao et al., 2025; Lee et al., 2026; Cai et al., 2026). Despite their different conclusions, these approaches often locate a capability boundary by asking whether a response appears in a finite sample.

More fundamentally, nonobservation does not identify a structural zero. A successful mode missing from n base rollouts may have probability zero or a positive probability below the experiment's resolution. These two worlds produce the all-zero sample with probabilities one and $(1-\varepsilon)^n$, respectively, and become arbitrarily hard to distinguish as ε approaches zero. More rollouts reduce the unresolved scale, but they do not turn finite absence into a probability equality. Fig. 1(a) therefore motivates our central question: *which support, probability, and budget-dependent capability claims can finite matched-checkpoint evaluations actually identify?*

To answer this question, we separate population support from finite-budget visibility. At the population level, absolute continuity of response laws passes through a fixed deterministic semantic auditor. In particular, full-softmax checkpoints sharing the same token and response universe retain the same semantic support. Nevertheless, their mode probabilities and observed sets can differ substantially. Thus structural assumptions and rollout evidence answer complementary questions.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

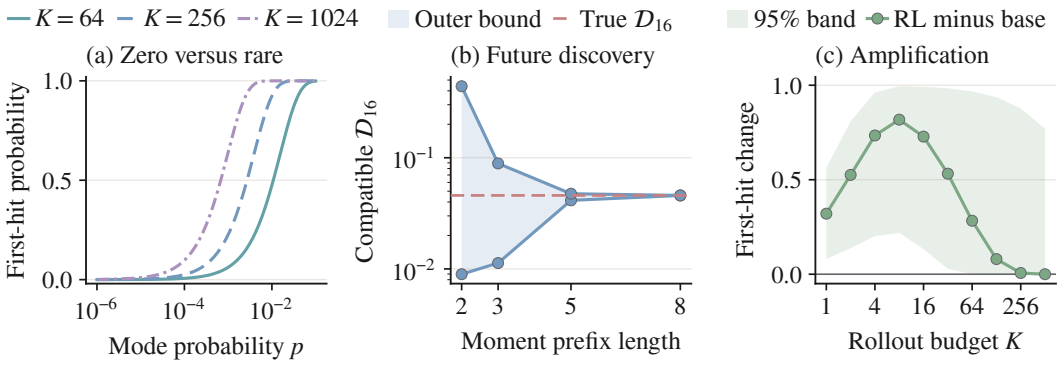

Figure 1: From finite observations to identifiable claims. (a) Zero and rare positive mass remain difficult to distinguish. (b) Additional discovery moments narrow the compatible future range; shading shows a Hausdorff outer bound. (c) Simultaneous first-hit intervals certify amplification of a mode already observed before RL.

We then replace a binary expansion label with a budget-indexed probability object. Each registered semantic mode has a first-hit curve: the probability that K rollouts expose it. Summing these curves gives expected discovered coverage, whose successive increments form a Hausdorff moment sequence (Hausdorff, 1921; Shohat & Tamarkin, 1943). The complete population curve identifies the positive mode-mass spectrum. A finite prefix instead yields a sharp set of compatible futures, with computable outer bounds illustrated in Fig. 1(b). This distinction replaces an arbitrary frequency cutoff with the information supplied by the observed curve.

Consequently, the rollout budget becomes part of the claim. Coverage is Schur-concave in the mode masses: at equal success probability, concentration reduces the number of distinct modes exposed. If RLVR changes both success mass and its concentration, a checkpoint can lead at one rollout and lose at a larger budget. Answer Pass@K for a fixed prompt depends only on total success and cannot distinguish within-prompt semantic spectra. Moreover, merging modes can conceal opposing probability changes. Neither a single budget nor a single semantic resolution therefore provides a universal capability ordering.

Guided by this account, we evaluate exact systems and matched neural checkpoints. An enumerable arithmetic environment fixes 2,623 held-out prompt-entrance atoms and verifies responses over rational arithmetic. Shared prompts, sampling laws, and rollout budgets support complete occupancy curves and simultaneous probability-change regions. These regions certify mode amplification without requiring novelty, as Fig. 1(c) previews. A secondary entrance intervention measures the gap between assigned access and natural conditioning, extending entrance-localization analysis (Zhou & Li, 2026). Finally, MATH-500 and GSM8K test the same distinction on ordinary mathematical reasoning (Hendrycks et al., 2021; Cobbe et al., 2021). The DeepScaleR lineage exhibits a budget-rank reversal that survives simultaneous curvewise inference, while other panels show different curve shapes. The empirical claim is therefore specific to the checkpoint, task panel, sampler, and deployment budget.

This paper makes three contributions:

- We establish finite-sample non-testability of zero mass and derive support transport and finite-prefix visibility sets.
- We characterize concentration, budget crossovers, and the loss of visibility and signed flow under semantic coarsening.
- We connect exact certificates to simultaneous matched-checkpoint inference and demonstrate a neural MATH-500 budget crossover.

2 Related Work

RLVR and capability boundaries. Outcome-reward training with PPO- and GRPO-family objectives underlies modern reasoning models (Shao et al., 2024; DeepSeek-AI, 2025). The resulting

capability debate contrasts large-budget evidence for reweighting or empirical support contraction (Yue et al., 2025; Wu et al., 2025) with evidence for new correct reasoning under process-aware metrics, later training phases, or modified exploration (Wen et al., 2025; Yao et al., 2025). Guided anchors and boundary-aware curricula explicitly target low-density or base-unsolved regions (Lee et al., 2026; Cai et al., 2026). Together, these studies analyze how training changes empirical reachability. We study the prior measurement question: which support, probability, and budget-indexed statements are identified by finite checkpoint rollouts, including when the base count is zero.

Pass@K, repeated sampling, and semantic breadth. Pass@K was popularized for code generation as the probability that at least one of K samples passes a verifier (Chen et al., 2021); self-consistency and best-of-N procedures similarly exploit repeated generation (Wang et al., 2023). Answer Pass@K is a function only of total success probability. Recent RLVR work supplements it with answer sets, reasoning patterns, entropy, process correctness, or enumerable solution families (Yue et al., 2025; Wen et al., 2025; Zhou & Li, 2026). Our coverage functional concerns distinct preregistered successful semantic modes, not merely one successful answer, and reports the complete curve over all budgets supported by the sample. A recent reliability report represents all-of-k success as moments of the across-task difficulty distribution and discusses its truncated moment problem and at-least-one task-level crossovers (TMLS AI Research, 2026). By comparison, our discovery moments encode the within-prompt semantic mass spectrum, support identification, signed checkpoint flow, and at-least-one budget crossovers. These are different measures over different probability objects.

Unseen species, moments, and partial identification. Estimating unseen categories has a long history in ecology, linguistics, and sampling theory (Fisher et al., 1943; Good & Toulmin, 1956; Efron & Thisted, 1976; Chao, 1984; Orlitsky et al., 2016). This literature studies observed frequency profiles and extrapolation to additional samples. Building on this literature, we use its occupancy logic but compare matched neural checkpoint laws over a task-conditioned semantic partition, derive signed inter-checkpoint flow, and connect the marginal discovery curve directly to a Hausdorff measure. Partial identification reports the full parameter set compatible with observables and assumptions rather than forcing point conclusions (Manski, 2003). Recent work applies that principle to latent labels inferred from panels of LLM reports (Chen et al., 2026); our latent object is instead a checkpoint's finite-rollout semantic probability spectrum, observed through executable generation.

3 Capability as a Probability Spectrum

We begin by defining the probability objects that remain meaningful at a finite budget. Appendix A collects the notation and sampling hierarchy before the first derivation. Let x be a prompt from a registered finite panel $\mathcal{X}$, let $j \in \{0, 1\}$ index matched checkpoints before and after RLVR, and hold the decoding law fixed. A deterministic semantic auditor maps a generated response r to either a successful mode $z \in \mathcal{Z}_x$ or a failure cell $\perp$. The prompt-specific successful-mode probability is

$$p_{j,x,z} = \mathbb{P}_j\{Z = z, Y = 1 \mid X = x\}. \tag{1}$$

Sampling X uniformly from the panel flattens (x, z) into an atom $a \in \mathcal{A}$ with joint mass $\widetilde{p}_{j,a} = p_{j,x,z}/|\mathcal{X}|$. A balanced design with the same number of rollouts per prompt estimates this mixture without changing its estimand. The positive semantic support is

$$S_j = \{(x, z) : p_{j,x,z} > 0\}. \tag{2}$$

Absolute capability expansion asserts $S_1 \setminus S_0 \neq \varnothing$.

Visibility at a deployment budget. Given this support definition, fix a prompt, suppress (j, x) and write the conditional successful-mode vector as $\boldsymbol{p} = (p_z)_z$. A mode's first-hit probability in K rollouts is

$$H_K(p) = 1 - (1 - p)^K. \tag{3}$$

The expected number of distinct successful modes visible for that prompt is

$$\mathcal{C}_K(\boldsymbol{p}) = \sum_{z \in \mathcal{Z}_x} H_K(p_z). \tag{4}$$

The panel curve averages conditional prompt curves:

$$\overline{\mathcal{C}}_{j,K} = \frac{1}{|\mathcal{X}|} \sum_{x} \mathcal{C}_K(\boldsymbol{p}_{j,x}). \tag{5}$$

Unlike $|S_j|$, these are finite-budget operational objects.

Probability flow. Writing $d_a = \widetilde{p}_{1,a} - \widetilde{p}_{0,a}$, define positive and negative semantic flows

$$F^+(\boldsymbol{d}) = \sum_a [d_a]_+, \qquad F^-(\boldsymbol{d}) = \sum_a [-d_a]_+. \tag{6}$$

They retain simultaneous entry and exit that a net success-rate difference can cancel. We analyze these objects at fine entrance, operator, and success granularities using a hierarchy fixed before held-out generation.

Access and execution. For a registered entrance b, let $q_j(b) = \mathbb{P}_j(B = b)$ and $\mu_j^{\mathrm{obs}}(b) = \mathbb{P}_j(Y = 1 \mid B = b)$. Then

$$p_j(b) = q_j(b)\mu_j^{\mathrm{obs}}(b). \tag{7}$$

A forced-entrance prompt estimates a different operational object, $\mu_j^{\mathrm{int}}(b) = \mathbb{P}_j(Y = 1 \mid I_b = 1)$, where I_b denotes assignment to the registered prompt intervention. Because a language model can disobey the requested entrance, this is an intention-to-treat quantity rather than an automatic identification of $\mathbb{P}(Y = 1 \mid \mathrm{do}(B = b))$. Section 4 keeps the objects separate through an estimable bridge gap and reports compliance.

4 What Finite Rollouts Identify

4.1 Absolute expansion is not finitely testable

With these objects in place, the first result isolates the inferential problem before any neural detail.

Theorem 4.1 (Zero versus rare)**.** *For n independent indicators, the total variation distance between the laws under $p = 0$ and $p = \varepsilon$ is*

$$\mathrm{TV}\big(\mathrm{Ber}(0)^{\otimes n}, \mathrm{Ber}(\varepsilon)^{\otimes n}\big) = 1 - (1 - \varepsilon)^n. \tag{8}$$

Consequently, any level-α test of zero mass has power at most $\alpha + 1 - (1 - \varepsilon)^n$ at mass ε. No finite-rollout test is uniformly more powerful than its size over the alternative $p > 0$.

The same bound applies to a matched-checkpoint expansion claim by holding the post-RLVR law fixed and changing only a base-mode mass from zero to ε. A Le Cam two-point argument also makes the minimax error for the binary support indicator arbitrarily close to $1/2$. Appendix B proves these statements and extends them to adaptive stopping; Appendix M gives assumptions that can restore support identification. This is uniform finite-sample non-testability, not failure of population point identification. The partial-identification result below instead concerns future functionals under a truncated moment observable.

Population assumptions supply complementary information. In particular, absolute continuity of response laws constrains semantic support even when rollout counts alone cannot resolve it.

Proposition 4.2 (Support transport)**.** *Let P_0 and P_1 be checkpoint response laws under the same prompt and decoding map. If $P_1 \ll P_0$, then every deterministic semantic pushforward also satisfies $p_{1,z} > 0 \Rightarrow p_{0,z} > 0$. If both likelihood ratios are positive almost surely, their semantic supports are equal.*

Finite reverse KL implies $P_1 \ll P_0$. Moreover, finite-logit full-softmax checkpoints assign positive probability to every admissible finite token string under the same vocabulary and stopping rules. Hence a valid response witness establishes positive semantic mass at both checkpoints, while its probability and finite-budget visibility can still change. Appendix C proves the proposition and derives sharp KL-to-visibility envelopes. Token truncation or other checkpoint-dependent filters require a separate support analysis.

4.2 THE VISIBILITY SPECTRUM IS A MOMENT SEQUENCE

Having separated zero mass from rare mass, we next characterize the observable discovery curve. Define the marginal probability of discovering a new successful mode when the budget grows from K to $K+1$:

$$\mathcal{D}_K(\boldsymbol{p}) = \mathcal{C}_{K+1}(\boldsymbol{p}) - \mathcal{C}_K(\boldsymbol{p}) = \sum_a p_a(1-p_a)^K. \tag{9}$$

Theorem 4.3 (Visibility-spectrum representation)**.** *Let $\nu_{\boldsymbol{p}} = \sum_a p_a \delta_{1-p_a}$ on $[0,1]$. Then*

$$\begin{aligned} \mathcal{D}_K &= \int_0^1 t^K \, d\nu_{\boldsymbol{p}}(t), \\ (-1)^r \Delta^r \mathcal{D}_K &= \sum_a p_a^{r+1}(1-p_a)^K \geq 0. \end{aligned} \tag{10}$$

The infinite sequence $(\mathcal{D}_K)_{K\geq 0}$ identifies $\nu_{\boldsymbol{p}}$ uniquely. For a finite semantic universe, it therefore identifies the multiset of positive mode masses up to permutation. If the registered universe has known size M, the exact population values $\mathcal{C}_1, \ldots, \mathcal{C}_M$ already identify all M masses, including zeros, through power-sum inversion.

Thus the full visibility curve has a precise status: it is not a collection of unrelated Pass@K values, but a completely monotone Hausdorff moment sequence. The positive mass at an atom $t < 1$ reveals the corresponding mode probability $1-t$ and its multiplicity. Zero-mass modes leave no trace, which is exactly the boundary in Theorem 4.1. Appendix D proves uniqueness and gives diagnostics implied by complete monotonicity.

Theorem 4.4 (Finite-prefix partial identification)**.** *For a registered universe of size M, let $\mathfrak{P}_{M,L}$ contain all $\boldsymbol{p} \in [0,1]^M$ with $\sum_a p_a \leq 1$ and $\ell_k \leq \mathcal{D}_k(\boldsymbol{p}) \leq u_k$ for $0 \leq k < L$. For $T \geq L$, the sharp future-discovery endpoints are*

$$\begin{aligned} \underline{D}_T &= \inf_{\boldsymbol{p}\in\mathfrak{P}_{M,L}} \mathcal{D}_T(\boldsymbol{p}), \\ \overline{D}_T &= \sup_{\boldsymbol{p}\in\mathfrak{P}_{M,L}} \mathcal{D}_T(\boldsymbol{p}). \end{aligned} \tag{11}$$

Both extrema are attained. Relaxing the finite-spectrum structure to all nonnegative measures matching the same moments gives valid outer endpoints from a linear program over measures, with finite atomic extremizers.

Accordingly, Equation 11 is the exact finite-spectrum identified set. The Hausdorff relaxation supplies computable outer bounds that shrink as more moments are measured. Appendix F gives both programs, their relation, the dual polynomial problem, and a numerical certificate.

4.3 NO SINGLE BUDGET ORDERS CAPABILITY SPECTRA

Theorem 4.5 (Concentration order and budget crossover)**.** *For every $K \geq 2$, $\mathcal{C}_K$ is symmetric, coordinatewise increasing, and Schur-concave in the successful-mode masses. Thus, at equal total success mass, a more concentrated spectrum has weakly lower expected coverage. If total mass also differs, there exist spectra $\boldsymbol{p}$ and $\boldsymbol{q}$ such that $\mathcal{C}_1(\boldsymbol{p}) > \mathcal{C}_1(\boldsymbol{q})$ but $\mathcal{C}_K(\boldsymbol{p}) < \mathcal{C}_K(\boldsymbol{q})$ for a larger K.*

For example, (0.60) leads $(0.20, 0.20, 0.10)$ at $K=1$ and loses from $K=2$ onward. Standard answer Pass@K equals $1-(1-s)^K$ for total success mass s, so it cannot see how s is distributed across semantic modes. Appendix G proves the theorem, while Appendix X gives additional collision and reversal witnesses.

4.4 WHAT PROBABILITY FLOW SURVIVES OBSERVATION

Proposition 4.6 (Coarsening contracts signed flow)**.** *Let a deterministic map merge fine atoms into coarse cells. If F^+_{fine} and F^-_{fine} are defined by Equation 6, then*

$$F^+_{\text{coarse}} \leq F^+_{\text{fine}}, \qquad F^-_{\text{coarse}} \leq F^-_{\text{fine}}. \tag{12}$$

Equality holds exactly when no coarse cell contains fine changes of opposing sign, apart from zeros.

In turn, the result follows from subadditivity of the positive part. It explains why an unchanged success rate can coexist with substantial redistribution among reasoning modes. Appendix J treats stochastic coarsenings and Appendix K instantiates the registered hierarchy.

Proposition 4.7 (Visibility under coarsening)**.** *For every deterministic semantic coarsening g and budget K, $\mathcal{C}_K(g_{\#}\boldsymbol{p}) \le \mathcal{C}_K(\boldsymbol{p})$. Coarsening contracts each checkpoint curve yet can reverse their matched ranking when collision losses change.*

Appendix J gives the collision-loss identity, proof, and an exact sign-reversal witness.

4.5 Simultaneous inference without a magnitude cutoff

To connect these population objects to data, consider a mode indicator observed on independent, potentially nonidentical prompts, let $\widehat{p}$ estimate the average panel probability $\bar{p}$. The Poisson-binomial Chernoff bound has exponent $n\,\mathrm{kl}(\widehat{p}\|\bar{p})$. Inverting both tails and allocating error across all registered cells and checkpoints gives a joint region $\mathcal{R}_{1-\alpha}$. We label a coordinate amplified only if

$$\inf_{\mathcal{R}_{1-\alpha}} (\widetilde{p}_{1,a} - \widetilde{p}_{0,a}) > 0, \tag{13}$$

and contracted by the analogous supremum below zero. Every other coordinate is unresolved. For a fine atom, restoring its known prompt weight converts the joint interval to task-conditional probabilities before the first-hit transform. Appendix H proves validity and records the allocation.

For $K \le n$, counts N_a from n rollouts also give the unbiased occupancy estimator

$$\widehat{\mathcal{C}}_K = \sum_a \left\{ 1 - \frac{\binom{n-N_a}{K}}{\binom{n}{K}} \right\}. \tag{14}$$

Thus one rollout panel estimates the complete in-budget visibility curve rather than privileging a manually chosen K. Appendix E connects this estimator to singleton counts and unseen-species extrapolation.

4.6 An access/execution application with a measured bridge

Equation 7 admits an exact symmetric Shapley attribution to changing access and observational execution. Replacing the latter by assigned entrance success introduces a measured bridge correction, which vanishes only under perfect enforcement and entrance consistency. Appendix L gives both identities and all compliance-aware estimands.

5 Experimental Design

Guided by the theory, we ask how finite budgets resolve checkpoint changes, how coarsening hides probability movement, and how natural access differs from assigned entrance access. All comparisons use fixed semantic auditors and matched sampling laws; detailed protocols appear in the appendix.

Exact certificates. Analytic probability vectors test Theorems 4.1 and 4.3. A zero-versus-rare grid spans $n \in \{1, 8, 32, 128, 512, 2048\}$ and masses 10^{-8} through 10^{-1}. For $(0.31, 0.17, 0.07, 0.02)$, we check eight finite-difference orders and bound order-16 discovery from moment prefixes of length 2, 3, 5, and 8 (Appendix N).

Enumerable semantic environment. Each task supplies four distinct integers from 1 through 13 and a target from 10 through 50. An exact rational solver retains tasks with 4 through 12 valid nonidentity leaf-level entrances. The panel has 128 design and 384 held-out tasks, with 2,623 held-out prompt-entrance atoms. Responses declare an entrance and a fully parenthesized expression. The fixed auditor checks the declared result, number multiset, syntax, rational value, and leaf-level entrance. Only design tasks determine the prompt and parser (Appendix O). The held-out tasks have a median of six valid entrances. This fixed universe retains modes unseen at either checkpoint, avoiding selection by observed success.

Table 1: Observed-only cells and simultaneous probability changes. Counts refer to registered fine atoms in the controlled panel and to problems in MATH-500 and GSM8K. Observed-only status is not a structural-zero certificate.

Panel	Pair	Base only	RL only	Amplified	Contracted
Controlled	Qwen3	18	9	1	0
MATH-500	Qwen3	26	16	6	34
MATH-500	DeepScaleR	25	5	0	0
GSM8K	Qwen3	14	11	5	2
GSM8K	DeepScaleR	20	19	0	0

Matched rollouts and entrance assignments. The controlled pair uses Qwen3-4B and its 500-step nonthinking mathematical RL checkpoint (Yang et al., 2025). Shared prompts, temperature 1, top-p 1, and a 256-token cap yield 256 responses per held-out task and checkpoint: 196,608 free responses. A separate decoding panel uses 64 responses at each registered alternative temperature. For every task-mode unit, an assigned prompt requests the exact entrance and draws 32 responses per checkpoint, for 167,872 intervention responses. Noncompliance receives outcome zero, so the assigned success rate is intention-to-treat. Free rollouts estimate natural access and conditional execution; assigned rollouts estimate intervention success and compliance. Appendices P and S specify the full protocols.

Real mathematical reasoning. MATH-500 provides 500 held-out problems (Hendrycks et al., 2021). Each checkpoint generates 64 worked solutions with boxed answers per problem at temperature 1, top-p 1, and a 2,048-token cap. Symbolic verification supports unbiased first-hit estimates at $K \in \{1, 2, 4, 8, 16, 32, 64\}$. The independent DeepScaleR lineage tests whether the budget effect replicates beyond Qwen3. A GSM8K transfer uses all 1,319 problems and 32 responses per checkpoint and problem (Cobbe et al., 2021). Appendices T, U, and V report all panels and lineages.

Registered inference. Fine, operator, and outcome families share a 95% simultaneous region (Appendix H); all coordinates retain their estimates, bounds, and statuses. Equation 14 supplies complete in-budget curves. Paired cluster bootstraps use 10,000 draws, with curvewise max-t bands distinguished from pointwise intervals. A zero count is always accompanied by its positive upper probability bound. Appendices Q and I detail the estimands and allocation calculations.

6 RESULTS

We organize the evidence into six observations, moving from exact identification to controlled probability flow and then to real-reasoning curves. Table 1 contrasts observed-only cells with identified probability changes across the neural panels.

Observation 1: Finite discovery evidence gives a range that narrows as moments accumulate. First, the exact certificate separates the support question from the observable curve. At 256 base rollouts, an unseen atom can still have positive mass under every nonzero confidence level. For the worked four-mode spectrum, the deliberately size-agnostic Hausdorff relaxation leaves the order-16 marginal discovery at $[0.00896, 0.43970]$. Three moments narrow it to $[0.01129, 0.08889]$, five to $[0.04138, 0.04747]$, and eight to $[0.045802, 0.045841]$. The truth, 0.045837, is contained throughout. The outer interval contracts from the measured curve and contains the exact finite-spectrum identified set; no effect threshold enters. When the size four is imposed, four exact moments recover the spectrum. Appendix F gives both primal problems and dual certificates, and Appendix N reports all exact checks.

Observation 2: Observed mode differences need not imply support differences. Turning to matched neural checkpoints, on 384 held-out enumerable tasks, the base and RL checkpoints have strict one-rollout success 0.01950 and 0.01886. Their expected discovered-mode counts increase to 0.250 and 0.227 at budget 256. The paired difference moves from -0.00064 with interval $[-0.00439, 0.00302]$ at $K = 1$ to -0.02344 with interval $[-0.05208, 0.00260]$ at $K = 256$. Every interval contains zero. The complete curve in Fig. 2 and Appendix Table 7 shows increasing point separation without converting the terminal observed count into support.

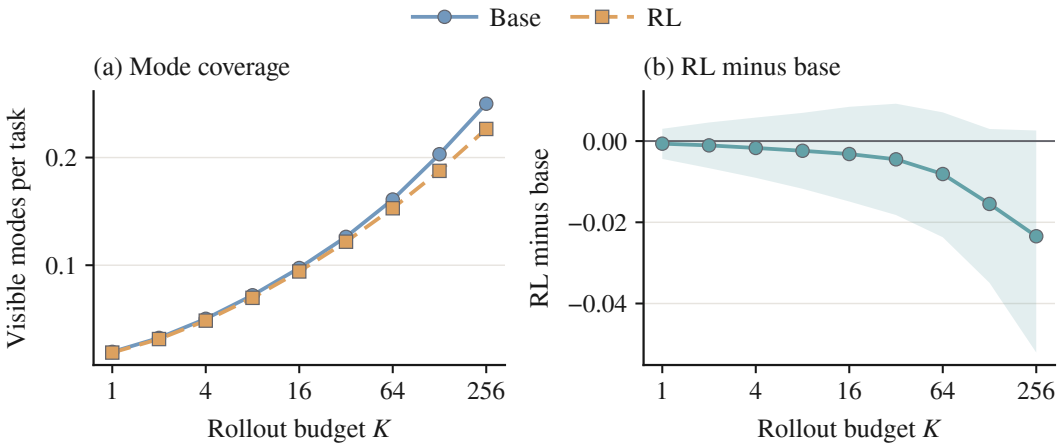

Figure 2: Complete controlled-task visibility curve. (a) Expected distinct successful modes per task for the matched Qwen3 checkpoints. (b) The paired RL-minus-base contrast with a 95% task-bootstrap interval. Every power-of-two budget supported by the 256-rollout panel is reported.

The two samples contain 105 successful fine atoms: 78 shared, 18 observed only before RL, and 9 observed only after RL. All 27 one-sided observed cells remain compatible with rare mass at the other checkpoint from counts alone. The verified full-softmax condition and enumerated witnesses additionally establish that all registered modes have positive architectural probability at both checkpoints. Thus the 27 one-sided cells are visibility differences, not literal support changes.

Observation 3: Probability amplification is identifiable, and coarse reporting hides most flow. Despite this support ambiguity, one preregistered fine atom is simultaneously amplified. Entrance $9-5$ on task 164 rises from 5 to 87 successes. Its task-conditional probability change is 0.320 with simultaneous interval $[0.080, 0.564]$. The first-hit advantage is 0.818 at budget eight with interval $[0.218, 0.995]$, then falls to 0.006 at budget 256 as both checkpoints approach visibility one. The positive base count makes this a probability-amplification certificate without a novelty claim. The distinction is operational. Finite-budget exposure rises sharply even though support membership is unchanged. Fine positive flow is 0.004924. It contracts to 0.001190 after merging operands into four operator cells and to 0.000641 after retaining only success versus failure. Outcome reporting therefore retains 13.0% of the fine point flow. Table 8 and Appendix R give all cell statuses and the complete audit. At budget 256, visibility coarsening attenuates the matched difference from -0.0234 at the fine level to -0.0104 at outcome level, with no hierarchy violation across any task or budget.

Observation 4: Interventions and decoding laws change the measurement. The same reasoning also requires a declared data-collection law. Across 2,623 assigned entrances, strict success is 0.07408 before RL and 0.07524 after RL, with compliance near 0.669. On the 302 units naturally accessed by both checkpoints, intervention-minus-observational execution gaps average -0.0671 and -0.0746. Thus the measured bridge in Equation 133 is necessary (Appendix S). Likewise, the budget-64 visibility contrast changes sign across registered temperatures, although every paired interval includes zero. Decoding is therefore part of the comparison rather than an architecture-level support boundary (Appendix W).

Observation 5: A neural checkpoint advantage can reverse with the rollout budget. We next test whether the budget dependence persists on standard mathematics. On MATH-500, one-rollout success is 0.648 for the base checkpoint and 0.573 after RL, a paired difference of -0.075 with interval $[-0.095, -0.056]$. The gap narrows monotonically with budget: at $K = 32$ it is -0.024 with interval $[-0.047, -0.002]$, and at $K = 64$ it is -0.020 with interval $[-0.044, 0.006]$. Fig. 3 shows the curves; Fig. 5 gives all 500 problem-level probability pairs.

The independent DeepScaleR lineage gives a neural crossover. RL raises one-rollout success from 0.409 to 0.438, a difference of 0.0288 with interval $[0.0177, 0.0397]$. The contrast is unresolved at $K = 2$ and reverses by $K = 4$ to -0.0129 with interval $[-0.0243, -0.0017]$. At $K = 64$, the base checkpoint leads by 0.040 with interval $[0.020, 0.060]$. This realizes the success-versus-diffusion

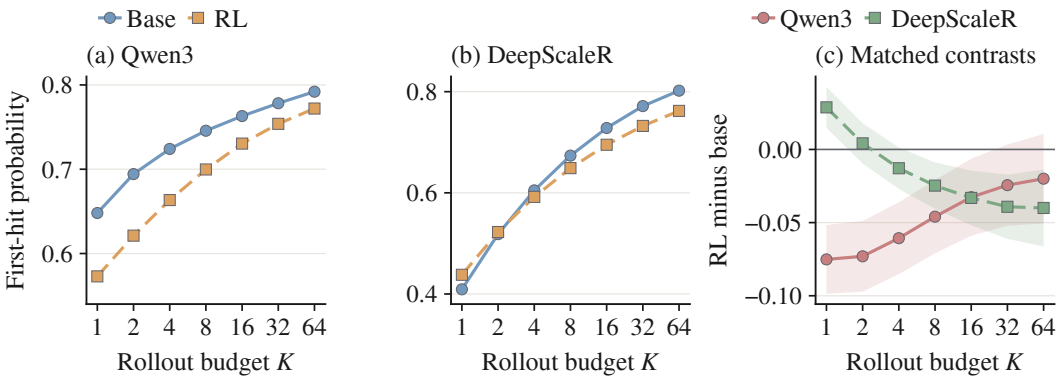

Figure 3: MATH-500 first-hit curves with 64 responses per problem. (a) The Qwen3 RL disadvantage narrows with budget. (b) DeepScaleR improves one-rollout success and reverses rank by budget four. (c) Paired intervals distinguish the crossover from the narrowing Qwen3 gap; bands are simultaneous over all seven budgets.

competition underlying Theorem 4.5 at the across-problem level: mean success can rise while repeated sampling favors a different distribution of problem difficulty. A 95% max-t band over the entire seven-budget curve is positive at $K = 1$ and negative from $K = 8$ onward, so the reversal survives simultaneous curvewise inference.

For the Qwen3 pair, 16 problems are unseen in base responses and seen after RL, while 26 show the reverse. Yet a zero base count has simultaneous upper probability 0.153, so none identifies a structural zero from rollouts. Under the verified full-softmax sampler, a finite boxed-answer witness also gives every problem positive architectural success support at both checkpoints. In contrast, the joint region identifies 6 amplified and 34 contracted problem probabilities. The same data therefore support concrete probability changes and leave absolute novelty unresolved. Appendix T gives all budgets and problem cells; Appendix V reports the second RL lineage. On GSM8K, Qwen3 RL improves one-rollout success by 0.0097 with a simultaneous curvewise lower bound of 0.0010, while later budgets and the DeepScaleR contrast remain unresolved. Appendix U reports all 1,319 problems and shows that the MATH crossover is a real phenomenon rather than a universal outcome.

Observation 6: Empirical novelty and identified probability change remain distinct across panels. Taken together, every neural panel contains atoms observed on only one side of the matched comparison. Across the controlled panel and four real checkpoint-task panels, 60 atoms are observed only after RL and 103 only before RL, while joint regions separate 48 named probability changes. The MATH and GSM8K zero-count upper bounds remain 0.153 and 0.303. Thus the same finite samples identify some probability flow, leave the rest set-valued, and never turn post-only observation into structural zero. Their curve shapes span a narrowing deficit, a certified crossover, a one-rollout gain with unresolved later order, and a fully unresolved pair. Total success mass, concentration, and budget explain these cases without a dataset-independent expansion label.

7 Conclusion

Finite rollout evaluations can resolve probability changes and budgeted visibility while leaving structural novelty unidentified from nonobservation. We formalize this distinction through support transport, Hausdorff discovery moments, simultaneous probability regions, and semantic probability flow. Together, these results explain why checkpoint rankings can depend on both the rollout budget and the reporting partition. Exact enumeration and matched neural panels then show the practical consequences: identifiable amplification, substantial flow hidden by coarsening, and a budget crossover in real mathematical reasoning. Accordingly, a capability comparison should state its semantic predicate, sampling law, probability uncertainty, and deployment budget. These quantities make the claim interpretable without converting an unseen base event into a new capability. In practice, reporting complete in-budget curves together with simultaneous mode-level intervals makes improvements and tradeoffs visible. Structural support claims additionally require explicit assumptions about the sampler and semantic universe, rather than a larger rollout budget alone.

AI USAGE STATEMENT

Generative AI tools assist selected implementation and editorial tasks, including drafting and debugging code, improving prose, reorganizing the manuscript, and adjusting figure presentation. The authors develop the mathematical framework and equations and conduct the substantive analyses and interpretation of the findings. AI feedback on mathematical expressions serves as an additional consistency check; the authors assess that feedback and retain control over every final formulation and scientific conclusion. The models evaluated in this paper generate responses as part of the stated experimental protocol. Separately, AI assistants used during preparation do not supply or replace the reported experimental measurements. The authors check assisted code against its intended computations and research outputs and review assisted text and numerical statements against the underlying results. Responsibility for the submitted work rests with the authors, including its arguments, equations, empirical claims, code, and any other materials prepared with AI assistance.

REFERENCES

Pengxiang Cai, Tianchen Fang, Xiaohan Li, Qingyuan Zeng, Guocong Li, and Jintai Chen. Curriculum reinforcement learning can incentivize reasoning capacity in LLMs beyond the base model. *arXiv preprint arXiv:2606.22317*, 2026.

Anne Chao. Nonparametric estimation of the number of classes in a population. *Scandinavian Journal of Statistics*, 11(4):265–270, 1984.

Mark Chen, Jerry Tworek, Heewoo Jun, Qiming Yuan, Henrique Ponde de Oliveira Pinto, Jared Kaplan, Harri Edwards, Yuri Burda, Nicholas Joseph, Greg Brockman, et al. Evaluating large language models trained on code. *arXiv preprint arXiv:2107.03374*, 2021.

Xiaohong Chen, Ashesh Rambachan, and Elie Tamer. Partial identification from LLM prompts. *arXiv preprint arXiv:2606.15031*, 2026.

Karl Cobbe, Vineet Kosaraju, Mohammad Bavarian, Mark Chen, Heewoo Jun, Lukasz Kaiser, Matthias Plappert, Jerry Tworek, Jacob Hilton, Reiichiro Nakano, Christopher Hesse, and John Schulman. Training verifiers to solve math word problems. *arXiv preprint arXiv:2110.14168*, 2021.

DeepSeek-AI. DeepSeek-R1: Incentivizing reasoning capability in LLMs via reinforcement learning. *arXiv preprint arXiv:2501.12948*, 2025.

Bradley Efron and Ronald Thisted. Estimating the number of unseen species: How many words did Shakespeare know? *Biometrika*, 63(3):435–447, 1976. doi: 10.1093/biomet/63.3.435.

R. A. Fisher, A. Steven Corbet, and C. B. Williams. The relation between the number of species and the number of individuals in a random sample of an animal population. *Journal of Animal Ecology*, 12(1):42–58, 1943. doi: 10.2307/1411.

I. J. Good and G. H. Toulmin. The number of new species, and the increase in population coverage, when a sample is increased. *Biometrika*, 43(1–2):45–63, 1956. doi: 10.1093/biomet/43.1-2.45.

Felix Hausdorff. Summationsmethoden und momentfolgen. I. *Mathematische Zeitschrift*, 9:74–109, 1921. doi: 10.1007/BF01378337.

Dan Hendrycks, Collin Burns, Saurav Kadavath, Akul Arora, Steven Basart, Eric Tang, Dawn Song, and Jacob Steinhardt. Measuring mathematical problem solving with the MATH dataset. In *Advances in Neural Information Processing Systems*, volume 34, 2021.

Woosuk Kwon, Zhuohan Li, Siyuan Zhuang, Ying Sheng, Lianmin Zheng, Cody Hao Yu, Joseph E. Gonzalez, Hao Zhang, and Ion Stoica. Efficient memory management for large language model serving with PagedAttention. In *Proceedings of the ACM SIGOPS 29th Symposium on Operating Systems Principles*, pp. 611–626, 2023. doi: 10.1145/3600006.3613165.

Chanuk Lee, Minki Kang, and Sung Ju Hwang. SAGE: Shaping anchors for guided exploration in RLVR of LLMs. *arXiv preprint arXiv:2605.18864*, 2026.

Michael Luo, Sijun Tan, Justin Wong, Xiaoxiang Shi, William Y. Tang, Manan Roongta, Colin Cai, Jeffrey Luo, Li Erran Li, Raluca Ada Popa, and Ion Stoica. DeepScaleR: Surpassing O1-Preview with a 1.5b model by scaling reinforcement learning, 2025. Technical report.

Charles F. Manski. *Partial Identification of Probability Distributions*. Springer, New York, 2003. doi: 10.1007/b97478.

Alon Orlitsky, Ananda Theertha Suresh, and Yihong Wu. Optimal prediction of the number of unseen species. *Proceedings of the National Academy of Sciences*, 113(47):13283–13288, 2016. doi: 10.1073/pnas.1607774113.

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 preprint arXiv:2402.03300*, 2024.

J. A. Shohat and J. D. Tamarkin. *The Problem of Moments*, volume 1 of *Mathematical Surveys*. American Mathematical Society, New York, 1943.

Changhui Sun, Lanbo Liu, Hang Lei, Tong Ling, Jiahang Xie, Zhiyong Zheng, Yujia Wang, Hao Liu, Feng Xiao, Lu Liu, Yanlong Du, Zifeng Cheng, Ziwei Jiang, and Qing Gu. Step-level on-policy distillation: Interpolating between on-policy distillation and supervised fine-tuning. *arXiv preprint arXiv:2608.16333*, 2026.

TMLS AI Research. The reliability frontier: pass^k consistency versus mean accuracy in agentic tasks. Technical Report tmls:reliability-frontier-passk:2026-06, The Machine Learning Society, June 2026.

Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc V. Le, Ed H. Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. Self-consistency improves chain of thought reasoning in language models. In *International Conference on Learning Representations*, 2023.

Xumeng Wen, Zihan Liu, Shun Zheng, Shengyu Ye, Zhirong Wu, Yang Wang, Zhijian Xu, Xiao Liang, Junjie Li, Ziming Miao, Jiang Bian, and Mao Yang. Reinforcement learning with verifiable rewards implicitly incentivizes correct reasoning in base LLMs. *arXiv preprint arXiv:2506.14245*, 2025.

Fang Wu, Weihao Xuan, Ximing Lu, Mingjie Liu, Yi Dong, Zaid Harchaoui, and Yejin Choi. The invisible leash: Why RLVR may or may not escape its origin. *arXiv preprint arXiv:2507.14843*, 2025.

An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, et al. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*, 2025.

Xinhao Yao, Lu Yu, Xiaolin Hu, Fengwei Teng, Qing Cui, Jun Zhou, and Yong Liu. The debate on RLVR reasoning capability boundary: Shrinkage, expansion, or both? a two-stage dynamic view. *arXiv preprint arXiv:2510.04028*, 2025.

Yang Yue, Zhiqi Chen, Rui Lu, Andrew Zhao, Zhaokai Wang, Yang Yue, Shiji Song, Gao Huang, et al. Does reinforcement learning really incentivize reasoning capacity in LLMs beyond the base model? In *Advances in Neural Information Processing Systems*, 2025.

Qiancheng Zhou and Ruizhe Li. Locked at the entrance, open inside: Where RLVR narrows the solution space. *arXiv preprint arXiv:2608.29188*, 2026.

A Notation and Sampling Conventions

Table 2 lists the objects used in the theory and experiments. We distinguish three indices throughout: prompt x, checkpoint j, and semantic mode z. Suppressing one of these indices always denotes an explicit mixture, never an implicit change of estimand.

Table 2: Notation used in the theory and experiments. The middle column separates mathematical parameters, observations, and experimental design objects.

Symbol	Type	Meaning
$x \in \mathcal{X}$	design	Prompt in the fixed evaluation panel
$j \in \{0, 1\}$	design	Matched checkpoint index before and after RLVR
r	observation	Generated response under the fixed decoding law
$Y \in \{0, 1\}$	observation	Registered executable success indicator
$z \in \mathcal{Z}_x$	observation	Registered semantic mode for prompt x
$a = (x, z)$	parameter index	Flattened prompt-mode atom in $\mathcal{A}$
$p_{j,x,z}$	parameter	Probability of successful mode z on prompt x
$\widetilde{p}_{j,a}$	parameter	Joint uniform-prompt mass $p_{j,x,z}/\lvert\mathcal{X}\rvert$
S_j	parameter	Positive semantic support of checkpoint j
K	design	Rollout or deployment visibility budget
$H_K(p)$	functional	First-hit probability of a mode with mass p
$\mathcal{C}_K(\boldsymbol{p}_x)$	functional	Expected modes visible within prompt x by budget K
$\overline{\mathcal{C}}_{j,K}$	functional	Panel average of prompt-conditional visibility curves
$\mathcal{D}_K(\boldsymbol{p})$	functional	Marginal discovery from budget K to $K + 1$
$\nu_{\boldsymbol{p}}$	parameter	Size-biased visibility measure $\sum_a p_a \delta_{1-p_a}$
d_a	functional	Matched joint-mass change $\widetilde{p}_{1,a} - \widetilde{p}_{0,a}$
F^+, F^-	functional	Positive and negative semantic probability flow
B	observation	Registered entrance family
$q_j(b)$	parameter	Free-rollout access probability $\mathbb{P}_j(B = b)$
$\mu_j^{\mathrm{obs}}(b)$	parameter	Observational execution conditional on access
$\mu_j^{\mathrm{int}}(b)$	parameter	Success under assignment to the entrance prompt
$\gamma_j(b)$	functional	Bridge gap $\mu_j^{\mathrm{int}}(b) - \mu_j^{\mathrm{obs}}(b)$
N_a	observation	Count of atom a in a rollout panel
α	design	Registered familywise error probability

A.1 Fixed-prompt and task-population estimands

The primary finite-panel object averages prompt-specific probabilities over the registered held-out panel:

$$\bar{p}_{j,z} = \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} p_{j,x,z}, \tag{15}$$

where a mode label shared across prompts denotes an explicitly coarsened semantic family. For the fine partition, the atom contains the prompt identity, so Equation 15 becomes $\widetilde{p}_{j,a} = p_{j,x,z}/|\mathcal{X}|$. The balanced experiment draws the same number n of independent responses for every x. Its pooled mean is unbiased for the joint mass even though individual indicators need not share a common success probability.

Visibility retains the within-prompt sampling experiment before averaging:

$$\overline{\mathcal{C}}_{j,K} = \frac{1}{|\mathcal{X}|} \sum_{x} \sum_{z \in \mathcal{Z}_x} \{1 - (1 - p_{j,x,z})^K\}. \tag{16}$$

Applying H_K directly to $\widetilde{p}_{j,a}$ would instead describe a budget of randomly interleaved prompts and is not the reported deployment estimand.

A secondary task-population estimand treats the held-out prompts as a sample from the procedural generator. Uncertainty for that object resamples prompts as clusters and keeps matched checkpoints paired. We never use task bootstrap intervals to make a statement about the exact finite panel; the simultaneous Chernoff region and the paired task bootstrap answer different questions.

A.2 The failure cell

Mode coverage concerns successful semantic atoms, but signed-flow geometry is cleanest on a probability simplex. We therefore add one common failure cell $\perp$ with mass

$$\widetilde{p}_{j,\perp} = 1 - \sum_{a \in \mathcal{A}} \widetilde{p}_{j,a}. \tag{17}$$

The augmented differences sum to zero. Positive and negative flows then agree and equal the total variation distance between the augmented checkpoint laws:

$$F^+ = F^- = \frac{1}{2} \sum_{a \in \mathcal{A} \cup \{\perp\}} |\widetilde{p}_{1,a} - \widetilde{p}_{0,a}|. \tag{18}$$

When the failure cell is omitted, $F^+ - F^-$ equals the change in total success mass. Both conventions are reported explicitly.

A.3 Support, observed support, and visibility

The mathematical support S_j is a property of the checkpoint distribution. The observed set

$$\widehat{S}_{j,n} = \{a : N_{j,a} > 0\} \tag{19}$$

is a random variable indexed by a per-prompt rollout budget. Its expected size is $|\mathcal{X}|\overline{\mathcal{C}}_{j,n}$. Calling $\widehat{S}_{j,n}$ the support would erase the distinction that motivates Theorem 4.1. We use "observed modes" for Equation 19, "visibility" for its expected budget curve, and "support" only for positive population mass.

A.4 Sampling law as part of the checkpoint measurement

A checkpoint does not induce a unique response distribution until temperature, nucleus truncation, maximum length, chat template, and stopping rules are fixed. Let λ collect these choices. Every probability in the paper is formally $p^{(\lambda)}_{j,x,z}$. The superscript is suppressed within a registered cell and restored in decoding-law robustness analyses. This convention prevents a temperature change from being interpreted as a checkpoint effect.

B Zero Versus Rare: Proofs and Extensions

B.1 Proof of Theorem 4.1

Under $p = 0$, the length-n binary vector is the all-zero vector almost surely. Under $p = \varepsilon$, that same vector has probability $(1-\varepsilon)^n$. For a point mass P at outcome x_0 and an arbitrary law Q, total variation is $1 - Q(x_0)$. Hence

$$\mathrm{TV}(P_0^n, P_\varepsilon^n) = 1 - P_\varepsilon^n(0, \ldots, 0) = 1 - (1-\varepsilon)^n. \tag{20}$$

For any randomized test $\phi \in [0,1]$,

$$\mathbb{E}_\varepsilon \phi \le \mathbb{E}_0 \phi + \mathrm{TV}(P_0^n, P_\varepsilon^n) \le \alpha + 1 - (1-\varepsilon)^n. \tag{21}$$

Taking ε to zero makes the right side converge to α. This proves the uniform-power claim.

B.2 Matched-checkpoint expansion

Let $E_a = \mathbf{1}\{p_{0,a} = 0, p_{1,a} > 0\}$ be the absolute expansion indicator for one registered atom. Fix any post-RLVR mass $p_{1,a} = c > 0$ and consider two data-generating worlds:

$$\begin{aligned} \mathsf{W}_0 &: (p_{0,a}, p_{1,a}) = (0, c), \\ \mathsf{W}_\varepsilon &: (p_{0,a}, p_{1,a}) = (\varepsilon, c). \end{aligned} \tag{22}$$

The post-RLVR observations have the same law in both worlds, while the base observations differ by Equation 8. The expansion indicator equals one in W_0 and zero in W_ε.

For an estimator $\widehat{E}_a \in \{0,1\}$, Le Cam's inequality gives

$$\begin{aligned} R_n^* &:= \inf_{\widehat{E}_a} \max_{w \in \{0,\varepsilon\}} \mathbb{P}_{\mathsf{W}_w}\{\widehat{E}_a \neq E_a(\mathsf{W}_w)\} \\ &\ge \frac{(1-\varepsilon)^{n_0}}{2}. \end{aligned} \tag{23}$$

The lower bound tends to $1/2$ as ε tends to zero. Additional post-RLVR sampling cannot reduce this ambiguity because the post-RLVR factor is identical across the two worlds.

B.3 Local alternatives and required budget

Equation 21 can be inverted. To make the two simple laws at least τ apart in total variation requires

$$n \geq \frac{\log(1-\tau)}{\log(1-\varepsilon)}. \tag{24}$$

For small ε, this is approximately $-\log(1-\tau)/\varepsilon$. The inverse dependence is not a recommended minimum-mass threshold. It quantifies what any proposed minimum-mass assumption would cost. If the assumption allows ε to approach zero with n, uniform identification disappears.

For a local alternative $\varepsilon_n = c/n$, the distance converges to

$$\lim_{n\to\infty} \{1-(1-c/n)^n\} = 1 - e^{-c}. \tag{25}$$

Thus even increasing budgets leave nontrivial ambiguity along the natural $1/n$ boundary.

B.4 Adaptive stopping does not create missing evidence

Suppose a procedure draws at most $n_{\max}$ base responses and may stop based on previous outcomes and external randomness. Couple the $p = 0$ and $p = \varepsilon$ worlds using common uniforms. The entire stopped transcripts agree whenever all of the first $n_{\max}$ uniforms exceed ε, an event of probability $(1-\varepsilon)^{n_{\max}}$. Therefore

$$\mathrm{TV}(\mathsf{Law}_0(T), \mathsf{Law}_\varepsilon(T)) \leq 1 - (1-\varepsilon)^{n_{\max}}. \tag{26}$$

Adaptive allocation can spend samples efficiently across modes, but a bounded procedure cannot uniformly separate zero from a sufficiently rare positive mass.

B.5 Multiple modes and familywise language

With M registered atoms, an observed post-RLVR-only set may contain many candidates. Applying a union bound to simple zero-versus-rare comparisons does not solve the boundary problem: each candidate alternative still approaches the corresponding zero law. What simultaneous inference can establish is a positive probability difference, not the equality $p_{0,a} = 0$. This is why the main paper uses three statuses:

$$\{\textsc{Amplified},\ \textsc{Contracted},\ \textsc{Unresolved}\}. \tag{27}$$

None of these statuses is called absolute support expansion without an additional identifying assumption from Appendix M.

B.6 Identification and testability are distinct

An ideal observer who knows the full Bernoulli sampling law knows $p = \mathbb{P}(X = 1)$ and therefore point-identifies whether $p = 0$. Theorem 4.1 addresses a different question: no test based on a finite random sample has power uniformly separated from its size over all positive alternatives. Calling a zero count an empirical nonobservation is therefore correct; treating it as a uniformly valid population-zero certificate is not.

Partial identification enters when the observable itself is a finite prefix of the population visibility moments. Even if that prefix were known exactly, multiple measures in $\mathfrak{V}_L$ can produce it and imply different future moments. Finite data add statistical uncertainty around the prefix. The paper keeps the two layers separate: Theorem 4.1 is a testing lower bound, while Theorem 4.4 is a set-identification result.

C RLVR Transport and Semantic Support

This appendix connects response-distribution updates to the semantic probability objects in the main paper. The results concern population laws and are logically separate from finite-sample identification. Under a fixed auditor, these laws determine which semantic events are possible. Finite rollouts then measure how often those events become visible under each law.

C.1 Proof of Proposition 4.2

Let g be the deterministic auditor from responses to a finite semantic space, including failure. For a semantic cell z, write

$$A_z = \{r : g(r) = z\}. \tag{28}$$

The pushforward masses are $p_{j,z} = P_j(A_z)$. Absolute continuity $P_1 \ll P_0$ means that $P_0(A) = 0$ implies $P_1(A) = 0$ for every measurable A. Applying it to A_z gives

$$p_{0,z} = 0 \Longrightarrow p_{1,z} = 0, \tag{29}$$

or equivalently $p_{1,z} > 0 \Rightarrow p_{0,z} > 0$. If $P_0 \ll P_1$ also holds, the implication runs in both directions and the semantic supports are equal.

C.2 Finite reverse divergence preserves support

If $D_{\mathrm{KL}}(P_1 \| P_0) < \infty$, then $P_1 \ll P_0$. Otherwise a set with $P_1(A) > 0 = P_0(A)$ makes the Radon-Nikodym log ratio infinite on positive P_1 mass. Data processing further gives

$$\mathrm{kl}(p_{1,z} \| p_{0,z}) \le D_{\mathrm{KL}}(P_1 \| P_0) \tag{30}$$

for every semantic event A_z. A reverse-KL-regularized optimum with finite objective therefore cannot put mass outside the reference response support (Wu et al., 2025; Lee et al., 2026).

This population fact does not imply equality of finite observed sets. Under support preservation with $p_{0,z} > 0$ and $p_{1,z} > 0$, either checkpoint may miss the mode in a finite rollout panel. Equation 163 gives the probability of one such post-only event.

C.3 Full-softmax parameter updates

Consider an autoregressive model with finite logits at every reachable prefix and sampling without token truncation. Every vocabulary token then has strictly positive conditional probability. A finite response $r = (r_1, \ldots, r_T)$ has probability

$$P_j(r) = \prod_{t=1}^{T} P_j(r_t \mid r_{<t}) > 0. \tag{31}$$

Two finite-logit checkpoints have the same support over finite token strings, up to fixed stopping rules. Any deterministic semantic map consequently has the same support whenever each nonempty semantic cell contains at least one finite response.

Nucleus, top-k, grammar, or safety filters can introduce exact zeros that depend on the checkpoint and prefix. This is why our primary measurement uses top-$p = 1$: the decoding layer does not manufacture a support difference through probability truncation. The temperature sweep still changes probabilities and therefore visibility.

C.4 Sharp KL envelope for one semantic mode

Suppose $D_{\mathrm{KL}}(P_1 \| P_0) \le \kappa$ and the base semantic mass is p_0. Define

$$\ell_\kappa(p_0) = \inf\{q \in [0,1] : \mathrm{kl}(q \| p_0) \le \kappa\}, \tag{32}$$

$$u_\kappa(p_0) = \sup\{q \in [0,1] : \mathrm{kl}(q \| p_0) \le \kappa\}. \tag{33}$$

Binary data processing yields

$$\ell_\kappa(p_0) \le p_1 \le u_\kappa(p_0). \tag{34}$$

The bounds are sharp over unrestricted response spaces. A two-cell response space with likelihood ratio constant inside A_z and its complement realizes any feasible pair (p_0, p_1).

Monotonicity of first-hit probability transports the envelope to every budget:

$$H_K\{\ell_\kappa(p_0)\} \le H_K(p_1) \le H_K\{u_\kappa(p_0)\}. \tag{35}$$

This provides a training-law sensitivity analysis without defining newness by a probability cutoff.

C.5 Bounded likelihood ratios

A stronger condition assumes

$$e^{-B} \le \frac{dP_1}{dP_0} \le e^B \quad P_0\text{-almost surely.} \tag{36}$$

Integrating over A_z gives

$$e^{-B} p_{0,z} \le p_{1,z} \le e^B p_{0,z}. \tag{37}$$

The resulting visibility envelope is

$$\begin{aligned} H_K(e^{-B} p_{0,z}) &\le H_K(p_{1,z}), \\ H_K(p_{1,z}) &\le H_K\{\min(1, e^B p_{0,z})\}. \end{aligned} \tag{38}$$

At rare base mass, a bounded likelihood ratio can create a large relative change while both finite first-hit probabilities remain small.

C.6 Semantic coarsening and information loss

For the full semantic distributions, data processing gives

$$D_{\mathrm{KL}}(T\boldsymbol{p}_1 \| T\boldsymbol{p}_0) \le D_{\mathrm{KL}}(\boldsymbol{p}_1 \| \boldsymbol{p}_0) \tag{39}$$

for every stochastic reporting channel T. This divergence contraction is the information-theoretic counterpart of signed-flow contraction in Proposition 4.6. Coarse outcome evaluation can hide both total variation and KL movement among fine reasoning modes.

C.7 Training support and empirical support

Three distinct sets are useful:

$$S_j^{\mathrm{pop}} = \{z : p_{j,z} > 0\}, \tag{40}$$

$$S_{j,n}^{\mathrm{obs}} = \{z : N_{j,z} > 0\}, \tag{41}$$

$$S_{j,K,\tau}^{\mathrm{vis}} = \{z : H_K(p_{j,z}) \ge \tau\}. \tag{42}$$

The first is a population property, the second a random finite sample, and the third a deployment-relative set for declared K and visibility probability τ. Our primary report uses the complete H_K curve rather than selecting τ. Prior uses of "empirical support" generally refer to the second or third object. Naming the object resolves apparent disagreement without deciding the population support from nonobservation.

D Visibility Spectra and Occupancy Identities

D.1 First-hit and coverage identities

For a successful mode with mass p_a, absence from K independent rollouts has probability $(1-p_a)^K$. Summing the complementary indicators proves Equation 4. Taking one forward difference gives

$$\mathcal{C}_{K+1} - \mathcal{C}_K = \sum_a \{(1-p_a)^K - (1-p_a)^{K+1}\} \tag{43}$$

$$= \sum_a p_a (1-p_a)^K, \tag{44}$$

which is Equation 9.

The coverage curve obeys $\mathcal{C}_0 = 0$ and

$$\lim_{K\to\infty} \mathcal{C}_K = |\{a : p_a > 0\}|. \tag{45}$$

This limit explains why support size is an infinite-budget property. At every finite K, replacing a zero mass by a sufficiently small positive mass changes the curve by an arbitrarily small amount.

D.2 COMPLETE MONOTONICITY

For the forward-difference operator $\Delta f_K = f_{K+1} - f_K$, repeated application to one summand gives

$$(-1)^r \Delta^r \{p(1-p)^K\} = p^{r+1}(1-p)^K. \tag{46}$$

Summing over modes proves the second identity in Equation 10. In particular, marginal discovery is nonnegative, nonincreasing, and convex as a sequence. Higher differences alternate in sign. These constraints provide deterministic checks on estimated or extrapolated curves. A fitted curve that violates them cannot be the visibility spectrum of any probability vector.

D.3 HAUSDORFF MOMENT REPRESENTATION

Define the finite positive measure

$$\nu_{\boldsymbol{p}}(A) = \sum_a p_a \mathbf{1}\{1 - p_a \in A\} \tag{47}$$

for Borel sets $A \subseteq [0,1]$. Its Kth moment is

$$\int t^K d\nu_{\boldsymbol{p}}(t) = \sum_a p_a (1-p_a)^K = \mathcal{D}_K. \tag{48}$$

For the prompt panel, averaging the prompt-specific measures gives

$$\begin{aligned} \overline{\nu}_j &= \frac{1}{|\mathcal{X}|} \sum_x \sum_z p_{j,x,z} \delta_{1-p_{j,x,z}}, \\ \overline{\mathcal{D}}_{j,K} &= \int t^K d\overline{\nu}_j(t). \end{aligned} \tag{49}$$

Thus the reported average conditional curve retains the same Hausdorff moment structure without applying the nonlinear first-hit map to joint prompt masses. Finite measures on a compact interval are determined by all polynomial moments. One proof first notes that equal moments imply equal integrals of every polynomial, then invokes uniform polynomial approximation of continuous functions and the Riesz representation theorem. Thus the infinite discovery sequence uniquely determines $\nu_{\boldsymbol{p}}$.

For an atom at $t < 1$, its associated mode mass is $p = 1 - t$. If m_p modes share that probability, the measure places weight $m_p p$ at $1 - p$, so

$$m_p = \frac{\nu_{\boldsymbol{p}}(\{1-p\})}{p}. \tag{50}$$

The positive mass spectrum and its multiplicities are therefore identified up to mode labels. The discovery curve alone does not attach a recovered mass to a semantic name. Labeled rollout counts provide that additional information.

D.4 FINITE-UNIVERSE POWER-SUM INVERSION

When the registered semantic universe has known size M, pad the mass vector with zeros to length M and define power sums

$$S_r = \sum_{a=1}^{M} p_a^r. \tag{51}$$

The binomial theorem gives the triangular relation

$$\mathcal{C}_K = \sum_{r=1}^{K} (-1)^{r+1} \binom{K}{r} S_r. \tag{52}$$

Starting from $S_1 = \mathcal{C}_1$, this system recursively recovers $S_1, \ldots, S_M$ from $\mathcal{C}_1, \ldots, \mathcal{C}_M$.

Let e_r be the elementary symmetric polynomial of degree r in the M masses, with $e_0 = 1$. Newton's identities give

$$r e_r = \sum_{i=1}^{r} (-1)^{i-1} e_{r-i} S_i, \qquad r = 1, \ldots, M. \tag{53}$$

The recovered e_r define the monic polynomial

$$P(u) = u^M - e_1 u^{M-1} + e_2 u^{M-2} - \cdots + (-1)^M e_M, \tag{54}$$

whose multiset of roots is exactly $(p_1, \ldots, p_M)$. Thus M exact population coverage values identify every mass, including structural zeros.

This finite algebraic result and Theorem 4.1 concern different information. Exact real-valued population coverages are not observed from a finite rollout sample. Estimation errors are amplified by alternating binomial coefficients and polynomial root recovery, especially near zero or repeated roots. The moment identified set propagates that uncertainty instead of plugging noisy estimates into Equation 54.

D.5 Conditioning of direct inversion

Perturbing coverage values by δ_K perturbs the recursively recovered power sum by a signed combination whose coefficients grow with binomial terms. A crude bound is

$$|\widetilde{S}_K - S_K| \le |\delta_K| + \sum_{r=1}^{K-1} \binom{K}{r} |\widetilde{S}_r - S_r|. \tag{55}$$

Even small coordinate errors can therefore grow quickly with K. Near a multiple root, the inverse map from polynomial coefficients to individual masses is also non-Lipschitz. These conditioning facts explain why finite sampling produces partial rather than stable point identification in the rare tail despite the finite population algebra.

D.6 Unbiased in-budget occupancy estimation

Let N_a be the number of occurrences of mode a among n independent rollouts. Draw a uniformly random subset of $K \le n$ positions. Conditional on N_a, the probability that this subset contains no occurrence of a is

$$\frac{\binom{n-N_a}{K}}{\binom{n}{K}}. \tag{56}$$

Averaging first over the random subset and then over the original rollouts is equivalent to drawing K fresh exchangeable positions. Its unconditional absence probability is $(1-p_a)^K$. Hence each summand of Equation 14 is unbiased for $H_K(p_a)$, proving unbiasedness of the full curve.

Two endpoint checks are useful. At $K = 1$,

$$\widehat{\mathcal{C}}_1 = \sum_a \frac{N_a}{n}, \tag{57}$$

which is the sample success probability. At $K = n$,

$$\widehat{\mathcal{C}}_n = \sum_a \mathbf{1}\{N_a > 0\}, \tag{58}$$

the observed number of successful modes. Intermediate budgets interpolate between accuracy and observed breadth using the same recorded panel.

D.7 Answer Pass@K and semantic discovery answer different questions

Let $s = \sum_a p_a$ be total success probability. The probability that at least one of K rollouts succeeds is

$$\text{Pass@}\,K = 1 - (1-s)^K. \tag{59}$$

This functional depends on $\boldsymbol{p}$ only through s. In contrast, $\mathcal{C}_K$ depends on every mode mass. Two checkpoints can therefore have identical Pass@K at every budget and different semantic visibility curves. At equal s, the difference begins at $K = 2$ because

$$\mathcal{C}_2 = 2s - \sum_a p_a^2. \tag{60}$$

The collision term $\sum_a p_a^2$ is invisible to answer-level Pass@K and is the first spectral quantity revealed by semantic occupancy.

E Frequency Profiles and Unseen Modes

The visibility spectrum can also be expressed through the empirical frequency-of-frequencies used in unseen-species estimation (Good & Toulmin, 1956; Efron & Thisted, 1976; Orlitsky et al., 2016). This appendix gives the identities and clarifies which are used in our reporting.

E.1 Occupancy frequency profile

For n rollouts, let

$$\Phi_r = \sum_a \mathbf{1}\{N_a = r\} \tag{61}$$

be the number of successful semantic modes observed exactly r times. Under a multinomial sample,

$$\mathbb{E}\Phi_r = \sum_a \binom{n}{r} p_a^r (1-p_a)^{n-r}. \tag{62}$$

The observed successful-mode count is

$$|\widehat{S}_n| = \sum_{r=1}^{n} \Phi_r, \tag{63}$$

while the number of registered but unobserved modes is Φ_0. Because the semantic universe is enumerated in the controlled task, Φ_0 is observable as a count of zero cells. Its population probability mass is not zero and must be inferred.

E.2 Singletons estimate next-step discovery

The singleton count satisfies

$$\frac{\mathbb{E}\Phi_1}{n} = \sum_a p_a(1-p_a)^{n-1} = \mathcal{D}_{n-1}. \tag{64}$$

Thus Φ_1/n is unbiased for the probability that rollout n reveals a mode not seen in the previous $n-1$ positions. This is the Good-Turing missing-mass identity shifted by one index. A panel with many singleton successful modes can have a large marginal discovery rate even when its observed mode count appears stable.

More generally, falling-factorial combinations of N_a estimate polynomial functionals of the spectrum. Equation 14 packages these functionals into the complete in-budget coverage curve and remains stable for every $K \le n$.

E.3 Future new-mode count

Given n initial rollouts and m fresh rollouts, the expected number of modes that are absent initially and appear later is

$$U_{n,m}(\boldsymbol{p}) = \sum_a (1-p_a)^n \{1-(1-p_a)^m\}. \tag{65}$$

It is a linear combination of discovery moments:

$$U_{n,m} = \sum_{h=0}^{m-1} \mathcal{D}_{n+h}. \tag{66}$$

Moment bounds on the future discovery sequence therefore give bounds on the number of modes expected from a larger evaluation budget.

Under Poissonized sampling with $m = tn$, the classical Good-Toulmin estimator is

$$\widehat{U}(t) = \sum_{r=1}^{\infty} (-1)^{r+1} t^r \Phi_r. \tag{67}$$

Its alternating coefficients grow when $t > 1$, creating high variance and motivating smoothing or restricted extrapolation. Our main report stays within the generated budget using Equation 14; beyond-budget claims use the identified interval in Equation 11 instead of an unconstrained point extrapolation.

E.4 Two checkpoints have two frequency profiles

Let $\Phi_{j,r}$ be the profile at checkpoint j. Comparing $\Phi_{1,0}$ and $\Phi_{0,0}$ measures a finite observed-set difference, but it mixes probability changes across the entire rare tail. A mode can move from $N_{0,a} = 0$ to $N_{1,a} > 0$ under equal population support, while another moves in the opposite direction. The signed coordinate region retains their identities; the frequency profile summarizes their rarity structure.

A useful matched table cross-classifies counts:

$$\Phi^{01}_{r,s} = \sum_a \mathbf{1}\{N_{0,a} = r, N_{1,a} = s\}. \tag{68}$$

The post-only empirical modes are $\sum_{s\geq 1} \Phi^{01}_{0,s}$, and the base-only modes are $\sum_{r\geq 1} \Phi^{01}_{r,0}$. We report both, together with the number observed at both checkpoints.

E.5 Known semantic universe versus unknown species

Classical unseen-species problems often leave the number of possible categories unknown. The controlled environment supplies all logically valid entrances, so the zero-count category list is known. This removes uncertainty about the denominator but leaves the probability problem intact. Even with a known finite universe, a zero count is compatible with every sufficiently small positive mass. Enumeration enables simultaneous coordinate bounds and intervention on a named unseen mode; it does not invalidate Theorem 4.1.

E.6 Lower bounds on support size

Every observed mode certifies at least one positive-support atom, so

$$|S| \geq |\widehat{S}_n|. \tag{69}$$

Frequency-based estimators such as Chao's use singletons and doubletons to raise this lower estimate (Chao, 1984). They still rely on a statistical model or asymptotic target and do not prove that an individual zero-count mode has zero or positive mass. Our problem requires both aggregate coverage and labeled matched changes, so we retain coordinate regions beside any aggregate species estimate.

F Sharp Truncated-Moment Identification

F.1 The finite-spectrum identified set

Let the first L discovery moments lie in registered intervals $[\ell_k, u_k]$. For a known registered universe of size M, define

$$\begin{aligned}\mathfrak{P}_{M,L} := \{\boldsymbol{p} : \quad & \boldsymbol{p} \in [0,1]^M, \ \sum_a p_a \leq 1, \\ & \ell_k \leq \mathcal{D}_k(\boldsymbol{p}) \leq u_k, \ 0 \leq k < L\}.\end{aligned} \tag{70}$$

This set is compact, and $\mathcal{D}_T$ is continuous, so both extrema in Equation 11 are attained. Its image need not be convex, but its minimum and maximum are sharp endpoints: every feasible finite spectrum has future discovery between them, and each endpoint is achieved by one.

The constraints are nonlinear in $\boldsymbol{p}$. They retain the defining relation between an atom location $1 - p_a$ and its measure weight p_a, as well as the known number of registered modes. This is the exact finite-spectrum partial-identification problem.

F.2 Hausdorff measure relaxation

For a tractable outer program, define $m_k(\nu) = \int_0^1 t^k d\nu(t)$ and

$$\begin{aligned}\mathfrak{V}_L = \{\nu \geq 0 : & \ell_k \leq m_k(\nu) \leq u_k, \\ & 0 \leq k < L\}.\end{aligned} \tag{71}$$

Every feasible spectrum induces $\nu_{\boldsymbol{p}} \in \mathfrak{V}_L$, but the converse need not hold because an arbitrary measure weight at t need not be an integer multiple of $1 - t$. Therefore

$$\begin{aligned} \inf_{\nu \in \mathfrak{V}_L} m_T(\nu) &\le \underline{D}_T, \\ \overline{D}_T &\le \sup_{\nu \in \mathfrak{V}_L} m_T(\nu). \end{aligned} \tag{72}$$

The measure endpoints are sharp for the relaxed Hausdorff class and valid outer bounds for the checkpoint spectrum. Their interval is convex: every interior value is achieved by a convex combination of endpoint measures.

The set is weak-star compact whenever $u_0 < \infty$. Tightness is immediate on the compact domain $[0, 1]$, and the moment inequalities define a closed set. The objective is weak-star continuous. Both extrema are therefore attained.

F.3 Finite atomic extremizers

Under L exact measure equalities, map a point t to the moment vector

$$v_L(t) = (1, t, \dots, t^{L-1}) \in \mathbb{R}^L. \tag{73}$$

The normalized relaxed moment vector lies in the convex hull of this curve. Standard extreme-point results for moment sets imply that a linear-functional extremizer can be represented by at most L atoms. For interval constraints, introducing slack variables and retaining only active faces gives a finite atomic extremizer with at most one atom per active independent constraint. This is why the numerical program can be audited through a small set of support points even when its discretization grid is dense.

F.4 Dual polynomial certificates

For exact moments $m_0, \dots, m_{L-1}$, the upper-bound dual searches for a polynomial of degree at most $L - 1$ that dominates t^T on $[0, 1]$:

$$\begin{aligned} \overline{D}_T = \inf_{\lambda \in \mathbb{R}^L} & \sum_{k=0}^{L-1} \lambda_k m_k, \\ \text{s.t.} \quad & \sum_{k=0}^{L-1} \lambda_k t^k \ge t^T \quad \forall t \in [0, 1]. \end{aligned} \tag{74}$$

The lower dual reverses the inequality and takes a supremum. A feasible polynomial therefore certifies a bound independently of the primal optimizer. At optimum, contact points between the dual polynomial and t^T contain the atoms of a complementary primal solution.

For interval moments, the dual splits each coefficient into nonnegative upper and lower multipliers. The certificate remains a polynomial inequality over a compact interval and can be checked by root isolation or a sufficiently dense grid followed by interval arithmetic.

F.5 Grid linear program used in the exact certificate

The numerical certificate uses a registered grid $0 = t_1 < \cdots < t_G = 1$ and nonnegative weights w_g. It solves

$$\begin{aligned} \min / \max_{\boldsymbol{w} \ge 0} \quad & \sum_{g=1}^{G} w_g t_g^T, \\ \text{s.t.} \quad & \ell_k \le \sum_{g=1}^{G} w_g t_g^k \le u_k, \quad 0 \le k < L. \end{aligned} \tag{75}$$

When the true atoms belong to the grid and moments are exact, feasibility is guaranteed. Grid bounds converge monotonically to their continuous counterparts when one combines primal refinement with dual envelope checks. The experiment uses the grid only to illustrate the identified set; no neural claim depends on a hand-chosen grid resolution.

Table 3: Hausdorff-relaxation grid bounds for the worked spectrum. The true value is 0.0458373. Longer observed prefixes narrow the compatible future without an effect threshold.

Prefix	Lower	Upper	Width
2	0.008962	0.439700	0.430738
3	0.011293	0.088890	0.077597
5	0.041382	0.047475	0.006092
8	0.045802	0.045841	0.000039

F.6 Worked spectrum

For the four-mode spectrum

$$\boldsymbol{p} = (0.31, 0.17, 0.07, 0.02), \tag{76}$$

we constrain progressively longer exact prefixes and bound $\mathcal{D}_{16}$. Table 3 reports the deterministic certificate generated by the released program.

The first two moments permit radically different relaxed tails even though total success mass and the initial discovery slope agree. Five moments already remove most of the ambiguity, and eight nearly determine the target for this spectrum. This is a property of the observed moment geometry, not a universal guarantee: other spectra can remain weakly identified for longer prefixes.

The relaxation deliberately withholds the fact that the generating spectrum has four modes. If $M = 4$ is imposed, $\mathcal{D}_0$ through $\mathcal{D}_3$ recover power sums S_1 through S_4, and Newton's identities recover the exact four masses. The finite-spectrum bounds therefore collapse after four exact moments, while Table 3 remains wider because it is valid for the larger Hausdorff measure class. Reporting both shows precisely what the known semantic-universe size contributes.

F.7 From sampling uncertainty to moment uncertainty

In neural data, $\mathcal{C}_K$ and $\mathcal{D}_K$ are estimated rather than known. Simultaneous lower and upper bounds on the coverage curve induce moment intervals through

$$\mathcal{D}_K = \mathcal{C}_{K+1} - \mathcal{C}_K. \tag{77}$$

A valid outer interval uses the lower bound of $\mathcal{C}_{K+1}$ minus the upper bound of $\mathcal{C}_K$, and conversely for the upper endpoint. Solving Equation 75 over these intervals propagates finite-rollout uncertainty into future visibility. We report the identified interval rather than a point extrapolation whenever it remains broad.

G Budget Crossovers and Concentration Order

G.1 Proof of Schur-concavity

For $K \geq 2$, define $f_K(p) = 1 - (1-p)^K$ on $[0, 1]$. Its first two derivatives are

$$\begin{aligned} f'_K(p) &= K(1-p)^{K-1}, \\ f''_K(p) &= -K(K-1)(1-p)^{K-2} \leq 0. \end{aligned} \tag{78}$$

Thus f_K is increasing and concave. The symmetric separable sum $\mathcal{C}_K(\boldsymbol{p}) = \sum_a f_K(p_a)$ is Schur-concave. If $\boldsymbol{p}$ majorizes $\boldsymbol{q}$ and their sums agree, Karamata's inequality gives

$$\mathcal{C}_K(\boldsymbol{p}) \leq \mathcal{C}_K(\boldsymbol{q}). \tag{79}$$

At $K = 1$, $f_1(p) = p$, so the two coverages equal their common total success mass. Strict concavity makes Equation 79 strict whenever the spectra are not permutations and the majorization step redistributes positive mass.

G.2 An exact crossover witness

Consider

$$\boldsymbol{p} = (0.60), \qquad \boldsymbol{q} = (0.20, 0.20, 0.10). \tag{80}$$

At one rollout, $\mathcal{C}_1(\boldsymbol{p}) = 0.60$ and $\mathcal{C}_1(\boldsymbol{q}) = 0.50$. At two rollouts,

$$\mathcal{C}_2(\boldsymbol{p}) = 0.84, \qquad \mathcal{C}_2(\boldsymbol{q}) = 0.91. \tag{81}$$

The diffuse spectrum leads thereafter and converges to three visible modes, while the concentrated spectrum converges to one. This proves the existence claim in Theorem 4.5.

The witness separates two mechanisms that a single number mixes. Increasing total success mass favors small budgets. Redistributing that mass across more modes favors large-budget discovery. RLVR can change both at once, so the sign of a coverage contrast is allowed to depend on K without any contradiction.

G.3 A SUFFICIENT CONDITION FOR EVENTUAL REVERSAL

Let $s_p > s_q$ but let $m_p < m_q$ denote the numbers of positive modes. Then

$$\mathcal{C}_1(\boldsymbol{p}) - \mathcal{C}_1(\boldsymbol{q}) = s_p - s_q > 0, \tag{82}$$

while

$$\lim_{K\to\infty} \{\mathcal{C}_K(\boldsymbol{p}) - \mathcal{C}_K(\boldsymbol{q})\} = m_p - m_q < 0. \tag{83}$$

Because the difference is continuous in real-valued K after extending $(1-p)^K$, it has a real crossing. On integer budgets, the sign must change between some consecutive values unless it equals zero exactly at an integer. This argument establishes eventual reversal but not uniqueness.

G.4 MULTIPLE CROSSINGS ARE POSSIBLE

The difference between two discovery sequences is a finite signed mixture of exponentials:

$$\begin{aligned}\mathcal{D}_K(\boldsymbol{p}) - \mathcal{D}_K(\boldsymbol{q}) = &\sum_a p_a(1-p_a)^K \\ &- \sum_b q_b(1-q_b)^K.\end{aligned} \tag{84}$$

Signed exponential mixtures can change sign more than once. Therefore neither one early budget nor one late budget determines the ordering at all intermediate budgets. The complete curve is the smallest direct report that does not conceal such reversals.

G.5 DECODING-LAW DEPENDENCE

The mass vector itself depends on the registered decoding law λ. Temperature and truncation can change total success, concentration, or both. For two laws λ and λ', the checkpoint contrast is

$$\Delta_K^{(\lambda)} = \mathcal{C}_K(\boldsymbol{p}_1^{(\lambda)}) - \mathcal{C}_K(\boldsymbol{p}_0^{(\lambda)}). \tag{85}$$

There is no reason for its sign to equal that of $\Delta_K^{(\lambda')}$. Our decoding sweep treats this dependence as part of the measurement map. It does not select the law whose contrast is largest.

G.6 RELATION TO MAJORIZATION DIAGNOSTICS

Exact majorization is a strong partial order. Empirically, one may compare sorted cumulative masses

$$M_r(\boldsymbol{p}) = \sum_{a=1}^{r} p_{(a)}, \tag{86}$$

where $p_{(1)} \geq p_{(2)} \geq \cdots$. Crossing Lorenz prefixes make the two spectra incomparable by majorization, but the visibility curve remains defined. We therefore use majorization as a sufficient mechanistic explanation for coverage ordering and report the full curve as the primary observable. This comparison requires equal total mass. Unequal success totals instead require the budget-specific analysis above, because the fixed-sum majorization theorem does not apply directly.

H Simultaneous Probability Regions

H.1 Panel-average Chernoff bound

For a registered cell, let $X_i \in \{0, 1\}$ be independent indicators with possibly different means p_i. The fixed-panel target is

$$\bar{p} = \frac{1}{n}\sum_{i=1}^{n} p_i, \qquad \widehat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i. \tag{87}$$

For $\lambda > 0$, concavity of $p \mapsto \log\{1 + p(e^\lambda - 1)\}$ gives for $S_n = \sum_i X_i$

$$\log \mathbb{E}e^{\lambda S_n} = \sum_i \log\{1 + p_i(e^\lambda - 1)\} \tag{88}$$

$$\leq n \log\{1 + \bar{p}(e^\lambda - 1)\}. \tag{89}$$

The right side is the log moment-generating function of a $\mathrm{Binomial}(n, \bar{p})$ variable. Chernoff optimization therefore yields, for $u > \bar{p}$,

$$\mathbb{P}(\widehat{p} \geq u) \leq \exp\{-n\,\mathrm{kl}(u\|\bar{p})\}, \tag{90}$$

and for $\ell < \bar{p}$,

$$\mathbb{P}(\widehat{p} \leq \ell) \leq \exp\{-n\,\mathrm{kl}(\ell\|\bar{p})\}. \tag{91}$$

The binary divergence is

$$\mathrm{kl}(u\|p) = u \log \frac{u}{p} + (1 - u) \log \frac{1 - u}{1 - p}. \tag{92}$$

This proof justifies pooling a balanced panel without pretending that all prompts have the same cell probability.

H.2 Inversion

For a one-sided tail allocation η, define the lower endpoint as zero when $N = 0$ and otherwise as the solution $p < \widehat{p}$ to

$$n\,\mathrm{kl}(\widehat{p}\|p) = \log(1/\eta). \tag{93}$$

The upper endpoint is defined analogously above $\widehat{p}$, with endpoint one when $N = n$. At a zero count, the upper solution has the closed form

$$U(0, n, \eta) = 1 - \eta^{1/n}. \tag{94}$$

Equation 94 is positive for every finite n and every $\eta > 0$. A zero count therefore produces an upper bound, not an equality claim.

H.3 Familywise allocation

We register three nested granularities: fine prompt-entrance atoms, four operator families plus failure, and success versus failure. Let M_g be the number of cells at granularity g. The total familywise error α is split equally across granularities, and each granularity allocates

$$\eta_g = \frac{\alpha}{3 \cdot 4M_g} \tag{95}$$

to each checkpoint, cell, and tail. A union bound then gives simultaneous coverage of all coordinate intervals at probability at least $1 - \alpha$.

The allocation is fixed by the reporting family rather than tuned to observed counts. Other valid allocations change interval width but not the inferential distinction between zero mass and a finite upper bound. We provide the complete cell count and allocation in each result manifest. In particular, cells with no observed successes remain in the family. Dropping them after observing the data would change the registered multiplicity correction.

H.4 MATCHED DIFFERENCES AND VISIBILITY TRANSFORMS

If checkpoint intervals are $[L_{0,a}, U_{0,a}]$ and $[L_{1,a}, U_{1,a}]$, simultaneous bounds for the probability change are

$$L_a^{\Delta} = L_{1,a} - U_{0,a}, \qquad U_a^{\Delta} = U_{1,a} - L_{0,a}. \tag{96}$$

For a fine atom $a = (x, z)$, these are joint-mass endpoints. Multiplying all four by $|\mathcal{X}|$ restores conditional mode probabilities; endpoints are capped at one. Below, L and U denote those conditional endpoints. Because H_K is increasing, write the cell's first-hit contrast as $\Delta H_K = H_K(p_{1,a}) - H_K(p_{0,a})$. Then

$$\begin{aligned} L_{K,a}^{H} &= H_K(L_{1,a}) - H_K(U_{0,a}), \\ U_{K,a}^{H} &= H_K(U_{1,a}) - H_K(L_{0,a}), \\ L_{K,a}^{H} &\le \Delta H_K \le U_{K,a}^{H}. \end{aligned} \tag{97}$$

The same joint event validates Equation 97 for every reported K because no new random interval is constructed after the coordinate region.

H.5 OUTER BOUNDS FOR SIGNED FLOW

On the joint rectangular region, coordinatewise lower bounds on positive flow and valid outer upper bounds are

$$\sum_a [L_a^{\Delta}]_+ \le F^+ \le \min\left\{1, \sum_a [U_a^{\Delta}]_+\right\}, \tag{98}$$

$$\sum_a [-U_a^{\Delta}]_+ \le F^- \le \min\left\{1, \sum_a [-L_a^{\Delta}]_+\right\}. \tag{99}$$

These intervals are deliberately labeled outer bounds. Imposing the simplex constraints can sharpen them through a linear program, but the coordinatewise form is transparent and remains valid under every registered coarsening.

H.6 TASK-POPULATION UNCERTAINTY

For generalization beyond the fixed panel, we compute each checkpoint's occupancy curve separately for every held-out prompt and bootstrap prompts as paired clusters. If $c_{j,x,K}$ is the per-prompt estimator, one bootstrap draw uses common sampled indices $x_1^*, \ldots, x_{|\mathcal{X}|}^*$ and records

$$\widehat{\Delta}_K^* = \frac{1}{|\mathcal{X}|} \sum_i (c_{1,x_i^*,K} - c_{0,x_i^*,K}). \tag{100}$$

The reported percentile interval uses 10,000 registered draws. This interval describes prompt heterogeneity and is not substituted for the simultaneous cellwise region.

H.7 SIMULTANEOUS BANDS OVER THE BUDGET CURVE

A neural crossover requires two signed statements on the same estimated curve. Pointwise 95% intervals do not give a 95% joint statement across all budgets. We therefore construct a paired bootstrap max-t band over the registered grid $\mathcal{K}$. Let $\widehat{\Delta}_K$ be the observed task-average contrast, $\widehat{\Delta}_K^{*(b)}$ its value in bootstrap draw b, and $\widehat{\sigma}_K$ the bootstrap standard deviation. The draw-level maximum is

$$T^{*(b)} = \max_{K \in \mathcal{K}} \left| \frac{\widehat{\Delta}_K^{*(b)} - \widehat{\Delta}_K}{\widehat{\sigma}_K} \right|. \tag{101}$$

If $c_{0.95}$ is its empirical 95th percentile, the reported simultaneous band is

$$\widehat{\Delta}_K \pm c_{0.95} \widehat{\sigma}_K, \qquad K \in \mathcal{K}. \tag{102}$$

All budgets use the same resampled prompt indices, so the procedure retains the strong correlation induced by nested occupancy estimators.

For Qwen3 MATH-500, $c_{0.95} = 2.382$; for DeepScaleR, $c_{0.95} = 2.458$. The DeepScaleR band is positive at $K = 1$, unresolved at $K = 2$ and $K = 4$, and negative at $K \in \{8, 16, 32, 64\}$. The crossover claim therefore holds as a joint curve statement between budgets one and eight, while the point estimate first changes sign at budget four.

The GSM8K max-t grid contains six budgets through 32. Qwen3 has a positive simultaneous lower endpoint at $K = 1$ and unresolved later budgets; DeepScaleR is unresolved throughout. These classifications use the registered grid and do not search for a crossing location.

H.8 Why confidence is not an effect threshold

The confidence level controls a repeated-sampling error probability. It does not declare that changes below a chosen magnitude are scientifically absent. Every point estimate and interval is reported on its natural probability or visibility scale. The labels in Equation 27 depend only on whether the joint region separates a signed contrast from zero.

I Rollout-Budget and Allocation Design

I.1 Zero-count resolution as a curve

For one cell with zero occurrences in n independent trials, the one-sided upper endpoint at tail probability η is

$$u_0(n, \eta) = 1 - \eta^{1/n}. \tag{103}$$

This curve is the exact inversion of the all-zero probability, apart from the conservative familywise allocation. It decays approximately as $\log(1/\eta)/n$. Reporting u_0 states what the experiment resolves without declaring an effect below a chosen value irrelevant.

If a reader proposes a target resolution δ, the required budget follows by inversion:

$$n(\delta, \eta) = \left\lceil \frac{\log \eta}{\log(1 - \delta)} \right\rceil. \tag{104}$$

We report Equation 104 as a sensitivity calculation, not as the paper's definition of capability.

I.2 Matched probability-difference allocation

Suppose a total of $N = n_0 + n_1$ independent rollouts can be divided between two checkpoints for one Bernoulli cell. The variance of the difference estimator is

$$V(n_0, n_1) = \frac{p_0(1 - p_0)}{n_0} + \frac{p_1(1 - p_1)}{n_1}. \tag{105}$$

When p_0, p_1 are known design values, differentiation gives Neyman allocation

$$\frac{n_0}{n_1} = \sqrt{\frac{p_0(1 - p_0)}{p_1(1 - p_1)}}. \tag{106}$$

The probabilities are precisely the unknown objects in our setting. Equal allocation is minimax under checkpoint exchange and gives paired curves with the same maximum supported budget. We therefore use 256 versus 256 in the controlled panel and 64 versus 64 in each real-reasoning panel.

I.3 Support claims allocate asymmetrically

Once a post-RL mode is observed, additional post samples establish its positive mass more precisely but cannot resolve whether the base probability is zero. In the two-world construction of Equation 22, total variation depends only on n_0:

$$\mathrm{TV}(\mathsf{W}_0, \mathsf{W}_\varepsilon) = 1 - (1 - \varepsilon)^{n_0}. \tag{107}$$

An evaluation devoted specifically to a support claim should therefore spend remaining budget on the base checkpoint after confirming post visibility. The conclusion still requires a minimum-mass or structural assumption because every finite n_0 leaves arbitrarily small alternatives contiguous.

I.4 Allocation across prompts

For a fixed-panel mean $|\mathcal{X}|^{-1}\sum_x p_x$, unequal rollout counts produce the unbiased estimator

$$\widehat{p} = \frac{1}{|\mathcal{X}|}\sum_x \frac{N_x}{n_x}. \tag{108}$$

Its variance is minimized, for known p_x, by $n_x \propto \sqrt{p_x(1-p_x)}$. Adaptive estimates of those variances can improve accuracy estimation, but they complicate simultaneous rare-mode inference and make the maximum available visibility budget differ across prompts. Balanced allocation preserves a common curve through $K = 256$ and avoids result-dependent sampling.

I.5 One panel, many budgets

The occupancy estimator reuses a size-n sample for every $K \le n$. These are correlated estimators, which is desirable: the same underlying response panel defines a coherent curve. Pointwise reruns at different budgets would waste responses and add independent Monte Carlo variation. Paired cluster bootstrap draws retain this joint curve dependence.

I.6 Compute accounting

The controlled free panel contains

$$384 \times 256 \times 2 = 196{,}608 \tag{109}$$

responses. The entrance panel contains

$$2{,}623 \times 32 \times 2 = 167{,}872. \tag{110}$$

MATH-500 contains $500 \times 64 \times 2 = 64{,}000$ responses per matched pair. The temperature cells each contain $384 \times 64 \times 2 = 49{,}152$. All counts refer to generated responses; no hidden candidate-screening queries are used. GSM8K contributes 168,832 responses across both pairs, and the five MATH decoding-law replications contribute 160,000. The complete neural total is 1,067,072 generated responses.

I.7 Checkpoint and GPU sharding

Sharding changes wall-clock time and not sample allocation. Every shard receives task indices congruent to its shard index and all registered sample indices for those tasks. A complete merge requires the Cartesian task-sample identifier set. This design permits recovery from one failed process without repeating completed shards or changing the probability law.

J Signed Flow and Coarsening Geometry

J.1 Positive and negative parts

For any real vector $\boldsymbol{d}$, the identities

$$d_a = [d_a]_+ - [-d_a]_+, \qquad |d_a| = [d_a]_+ + [-d_a]_+ \tag{111}$$

imply

$$F^+ - F^- = \sum_a d_a, \qquad F^+ + F^- = \sum_a |d_a|. \tag{112}$$

For successful cells alone, the first quantity is the change in success mass. After adding failure, it is zero and each flow equals total variation as shown in Equation 18.

This decomposition prevents a common cancellation. If one mode gains mass δ and another loses the same amount, aggregate success does not change, while $F^+ = F^- = \delta$. Reporting both flows therefore distinguishes balanced redistribution from unchanged probabilities, even when the success contrast alone is exactly zero.

J.2 Proof of Proposition 4.6

Let g map fine atoms to coarse cells. The difference in cell c is

$$d'_c = \sum_{a:g(a)=c} d_a. \tag{113}$$

Subadditivity of the positive part gives

$$[d'_c]_+ \le \sum_{a:g(a)=c} [d_a]_+. \tag{114}$$

Summing over c proves the positive-flow inequality. Applying the same argument to $-\boldsymbol{d}$ proves the negative-flow inequality.

Equality in Equation 114 holds if and only if all nonzero d_a within the cell have the same sign. Equality of both flow inequalities therefore requires no coarse cell to mix a positive and a negative fine change.

J.3 Exact cancellation witness

Consider two successful fine modes and a shared failure cell:

$$\begin{aligned} \boldsymbol{p}_0 &= (0.10, 0.30, 0.60), \\ \boldsymbol{p}_1 &= (0.30, 0.10, 0.60). \end{aligned} \tag{115}$$

At the fine level, $F^+ = F^- = 0.20$. Merging the two success modes produces the same coarse law $(0.40, 0.60)$ at both checkpoints, so both coarse flows are zero. The coarse success rate and every answer-level Pass@K are identical even though the reasoning-mode law moves by total variation 0.20.

J.4 Stochastic semantic channels

A noisy semantic auditor can be represented by a Markov kernel T from fine atoms to reported labels. The coarsened difference is $T\boldsymbol{d}$. Markov kernels contract total variation:

$$\frac{1}{2}\|T\boldsymbol{d}\|_1 \le \frac{1}{2}\|\boldsymbol{d}\|_1. \tag{116}$$

When the augmented laws have equal total mass, positive and negative flows each equal these total variations. Thus probabilistic annotation noise can only hide net detectable movement at this level; it cannot create more total variation than exists in the underlying partition.

For successful cells without failure, a stochastic channel that maps every success atom to a success label preserves total signed mass. Combining this identity with Equation 116 gives contraction of both positive and negative flow.

J.5 Flow versus coverage

Signed flow compares checkpoint probability laws at one rollout. Coverage compares nonlinear budget transforms of those laws. They are related but not interchangeable. A first-order bound follows from the Lipschitz constant of H_K:

$$|H_K(p) - H_K(q)| \le K|p - q|. \tag{117}$$

Summing gives

$$|\mathcal{C}_K(\boldsymbol{p}_1) - \mathcal{C}_K(\boldsymbol{p}_0)| \le K(F^+ + F^-) \tag{118}$$

on successful cells. The bound is intentionally one-sided: substantial flow can cancel in coverage at a particular budget, while a large coverage change requires nontrivial underlying flow. At $K = 1$, the contrast reduces to $F^+ - F^-$. Larger budgets reweight each mode through its first-hit probability, so the two flow totals alone do not determine the coverage ranking.

J.6 Proof of visibility contraction

Within one coarse cell c, the event that the cell appears at least once is the union of its fine-mode events. The expected indicator of this union is $H_K(\sum_{a\in c} p_a)$, while the expected number of distinct fine modes is $\sum_{a\in c} H_K(p_a)$. The union bound gives

$$H_K\left(\sum_{a\in c} p_a\right) \le \sum_{a\in c} H_K(p_a). \tag{119}$$

Summing over cells proves visibility contraction. For $K \ge 2$, it is strict whenever a coarse cell contains at least two positive fine masses, because two distinct modes then have positive probability of both appearing.

Writing $L_K(\boldsymbol{p})$ for the fine-minus-coarse difference and subtracting the two checkpoints gives

$$\Delta\mathcal{C}_K^{\mathrm{coarse}} = \Delta\mathcal{C}_K^{\mathrm{fine}} + L_K(\boldsymbol{p}_0) - L_K(\boldsymbol{p}_1). \tag{120}$$

Although each collision loss is nonnegative, their matched difference has either sign. An exact reversal at $K = 2$ is

$$\boldsymbol{p}_0 = (0.60, 0), \qquad \boldsymbol{p}_1 = (0.25, 0.25), \tag{121}$$

when both modes are merged. Fine coverages are 0.84 and 0.875, so RL leads by 0.035. Coarse coverages are 0.84 and 0.75, so RL loses by 0.09. Each checkpoint curve contracts, yet their ranking reverses.

J.7 Hierarchy as an empirical diagnostic

Let F_g^+ denote positive flow at successive coarsenings. Proposition 4.6 yields the monotone chain

$$F^+_{\mathrm{outcome}} \le F^+_{\mathrm{operator}} \le F^+_{\mathrm{entrance}}. \tag{122}$$

The gaps localize cancellation. A large entrance-to-operator gap means opposing changes occur among entrances using the same operator. A large operator-to-outcome gap means different operators move in opposing directions while total success remains stable. The complete empirical hierarchy is reported rather than selecting the level with the largest contrast.

K Registered Semantic Partitions

K.1 Why the partition is fixed before held-out generation

Semantic coverage is meaningful only relative to a map from responses to modes. Searching many maps after observing checkpoint differences would create an unreported selection problem. We therefore construct the full arithmetic solution set, define the hierarchy, and freeze the parser on 128 design tasks before generating any of the 384 held-out tasks.

For each puzzle, a valid response uses every supplied integer exactly once, uses only binary $+, -, \times, /$ and parentheses, and evaluates exactly to the target over rational arithmetic. Floating-point tolerance never enters the verifier.

K.2 Fine entrance modes

A leaf-level entrance is an operation whose two children are original supplied integers. Its canonical label is

$$z = (a, \circ, b), \qquad \circ \in \{+, -, \times, /\}. \tag{123}$$

For commutative operators, operands are sorted. Ordered operands are retained for subtraction and division. The parser requires the declared entrance to occur unchanged as a leaf-level subtree in the final expression.

Identity multiplications and divisions by one are excluded from the registered fine partition. This is a semantic definition, not a probability cutoff: the same deterministic rule is applied to every task, checkpoint, and rollout. Appendix W reports an expanded partition that includes these operations.

K.3 Complete enumeration

For a multiset of current rational values, the solver recursively chooses an unordered pair, applies every valid ordered binary operation, and replaces the pair by its result. A branch succeeds if the final value equals the target. Conditioning the first recursion on a candidate entrance yields the complete set $\mathcal{Z}_x$ of registered modes with at least one valid continuation.

Memoization is not required at four operands, so the reference implementation keeps expression witnesses and exact fractions at every branch. The generation pipeline independently re-evaluates each witness through the response parser. All 512 released tasks pass this cross-check.

K.4 Coarse operator and outcome partitions

The operator map forgets operands:

$$g_{\mathrm{op}}(a, \circ, b) = \circ. \tag{124}$$

The outcome map merges all registered successful modes and retains failure:

$$g_{\mathrm{out}}(z) = \textsc{Success}, \qquad g_{\mathrm{out}}(\bot) = \textsc{Failure}. \tag{125}$$

These deterministic maps instantiate the nesting assumed by Proposition 4.6.

K.5 Failure reasons are diagnostic, not alternative modes

The parser records missing tags, unavailable entrance operands, an incorrect declared intermediate result, invalid syntax, a mismatched number multiset, a declared entrance absent from the expression, a wrong target, and an unregistered identity entrance. All map to the common failure cell for the primary semantic law. Appendix R reports their rates to show whether a checkpoint contrast is driven by formatting or arithmetic, but they are not searched as post hoc semantic modes.

K.6 Prompt identity in the fine atom

The same surface entrance, such as $9 - 5$, can occur on different puzzles. It does not denote the same solution family because the remaining numbers and target differ. The fine atom is therefore (x, z), while the operator analysis deliberately merges across prompts. This distinction also makes the complete denominator available: the held-out set contains 2,623 registered prompt-entrance atoms before the failure cell is added.

L Access, Execution, and the Intervention Bridge

L.1 Observational factorization

For a registered entrance b, define the event $B = b$ when the response declares the correct intermediate value and uses that entrance as a leaf-level subtree with the supplied number multiset. The response need not reach the target. Access and observational execution are

$$\begin{aligned} q_j(b) &= \mathbb{P}_j(B = b), \\ \mu_j^{\mathrm{obs}}(b) &= \mathbb{P}_j(Y = 1 \mid B = b). \end{aligned} \tag{126}$$

Their product equals the successful mode probability by the definition of conditional probability.

L.2 Symmetric product decomposition

The exact symmetric decomposition is

$$\begin{aligned} p_1 - p_0 = &\frac{\mu_0^{\mathrm{obs}} + \mu_1^{\mathrm{obs}}}{2}(q_1 - q_0) \\ &+ \frac{q_0 + q_1}{2}(\mu_1^{\mathrm{obs}} - \mu_0^{\mathrm{obs}}). \end{aligned} \tag{127}$$

Expanding the two midpoint terms in Equation 127 gives

$$\frac{\mu_0+\mu_1}{2}(q_1-q_0)+\frac{q_0+q_1}{2}(\mu_1-\mu_0) \tag{128}$$

$$=q_1\mu_1-q_0\mu_0. \tag{129}$$

The first term changes access while averaging execution across the endpoints; the second changes execution while averaging access. It equals the Shapley allocation obtained by averaging the two orders in which the factors may be changed. This removes the residual that appears if one arbitrarily changes access first or execution first.

The identity requires both conditional executions to be defined. If one checkpoint never accesses b in the population, its observational conditional is not identified. In finite data, a zero access count is treated as unresolved rather than assigned execution zero.

L.3 Registered entrance assignment

For every held-out puzzle and every registered nonidentity entrance, the intervention prompt adds one instruction: the final expression must contain the specified leaf operation and the declaration tag must state its exact rational result. All other examples, response tags, decoding settings, and verifier rules are unchanged.

Let $I_b=1$ denote assignment to this prompt and define

$$\mu_j^{\mathrm{int}}(b)=\mathbb{P}_j(Y=1\mid I_b=1). \tag{130}$$

Because generation remains unconstrained, assignment does not guarantee $B=b$. A response that solves the puzzle through another entrance is counted as noncompliant and unsuccessful for this estimand. We report the compliance rate

$$c_j(b)=\mathbb{P}_j(B=b\mid I_b=1) \tag{131}$$

beside μ_j^{int}.

L.4 Bridge identity

Define $\gamma_j=\mu_j^{\mathrm{int}}-\mu_j^{\mathrm{obs}}$. Substituting $\mu_j^{\mathrm{obs}}=\mu_j^{\mathrm{int}}-\gamma_j$ into $p_j=q_j\mu_j^{\mathrm{obs}}$ gives

$$p_1-p_0=q_1\mu_1^{\mathrm{int}}-q_0\mu_0^{\mathrm{int}}-(q_1\gamma_1-q_0\gamma_0). \tag{132}$$

Applying the symmetric product identity to the first difference proves the exact bridge identity

$$\begin{aligned}p_1-p_0=&\frac{\mu_0^{\mathrm{int}}+\mu_1^{\mathrm{int}}}{2}(q_1-q_0)\\&+\frac{q_0+q_1}{2}(\mu_1^{\mathrm{int}}-\mu_0^{\mathrm{int}})\\&-(q_1\gamma_1-q_0\gamma_0).\end{aligned} \tag{133}$$

The bridge gap combines prompt-assignment effects, imperfect compliance, and selection in the naturally accessed population. It is not labeled causal execution unless a stronger design structurally fixes the mediator B and satisfies consistency. Its scientific role is to stop an unjustified substitution: a prefix or instruction experiment need not equal $\mathbb{P}(Y=1\mid B=b)$.

L.5 Compliance-aware secondary quantities

Among compliant intervention responses, one may report

$$\mu_j^{\mathrm{pp}}(b)=\mathbb{P}_j(Y=1\mid I_b=1,B=b), \tag{134}$$

but conditioning on the post-assignment mediator is not a randomized causal comparison. We use it only to diagnose whether intervention failures arise from refusing the requested entrance or from failing after entering it.

The identity

$$\mu_j^{\mathrm{int}}(b)=c_j(b)\mu_j^{\mathrm{pp}}(b) \tag{135}$$

holds under the registered scoring rule because noncompliant responses receive outcome zero. It provides an exact audit of the scorer and separates instruction following from downstream arithmetic.

L.6 Finite-sample reporting

Free-rollout access is estimated from 256 responses per task. Observational execution is reported only where at least one access is observed and the main two-checkpoint decomposition only where both are observed. Entrance assignment uses 32 responses for every one of the 2,623 task-mode units. No unit is chosen because its free-rollout probability or intervention effect is large.

M Assumptions That Can Identify Support

The impossibility theorem is assumption-relative. This appendix records four ways to strengthen the model and the exact inferential price of each. None is used for the headline result.

M.1 Known minimum positive mass

Suppose every positive registered mass satisfies

$$p_a \in \{0\} \cup [p_{\min}, 1] \tag{136}$$

for known $p_{\min} > 0$. A zero count then has probability at most $(1 - p_{\min})^n$ under any positive mass. Testing zero against the restricted alternative becomes possible, and a familywise budget for M cells is

$$n \geq \frac{\log(\alpha/M)}{\log(1 - p_{\min})}. \tag{137}$$

The conclusion identifies support only at the declared resolution $p_{\min}$. Choosing that value from observed effects would reintroduce a manual threshold, so we instead report Equation 137 as a sensitivity curve.

M.2 Finite precision and deterministic truncation

If generation is executed by a finite-state sampler with exact zeroing, some sequences can have literal zero probability. A complete symbolic audit of all sampler states could then identify support. Standard autoregressive evaluation does not enumerate that state space, and floating-point underflow or nucleus truncation makes support depend on the decoding implementation. This assumption therefore identifies the support of a particular executable sampler, not an architecture-level capability.

M.3 Parametric mode spectra

Assume mode masses follow a known finite-dimensional family $p_a = p_a(\theta)$ with an identifiable parameter θ. Sufficient data can identify θ and hence zeros imposed by the family. The conclusion is only as credible as the structural zeros. A softmax parameterization with finite logits, for example, assigns positive probability to every token sequence allowed by the decoder and makes many literal support questions vacuous.

M.4 Continuity across related prompts or checkpoints

One may link an unseen cell to observed neighbors through a Lipschitz or smooth model. If a metric d and constant L satisfy

$$|p_a(x) - p_a(x')| \leq L d(x, x'), \tag{138}$$

nearby observed probabilities bound an unseen prompt. Such assumptions can be valuable for task-population generalization, but they do not turn a finite count into direct evidence of exact zero. The metric and constant must be externally justified.

M.5 Teacher or solver access

An exhaustive solver identifies which semantic modes are logically valid, as our controlled environment does. It does not identify whether a checkpoint assigns those modes positive probability. Likewise, a teacher can demonstrate a valid mode but cannot establish that the base model assigns it zero mass. Solver support and model support are distinct sets:

$$S_j \subseteq S_{\text{valid}}. \tag{139}$$

M.6 Sensitivity rather than hidden assumption

For any proposed identifying restriction, we recommend reporting conclusions as a function of its parameter. For the minimum-mass model, the relevant curve is the largest mass compatible with a zero count at error level η:

$$p_{\max}(n, \eta) = 1 - \eta^{1/n}. \tag{140}$$

This quantity comes from the sampling law and makes the assumption visible. It is not used to redefine "new" at a convenient empirical cutoff.

N Exact Finite-System Certificates

N.1 Zero-versus-rare grid

The exact script evaluates

$$\tau(n, p) = 1 - (1 - p)^n \tag{141}$$

with stable logarithmic arithmetic over rollout budgets $n \in \{1, 8, 32, 128, 512, 2048\}$ and probabilities $p \in \{10^{-8}, 10^{-7}, \ldots, 10^{-1}\}$. For each cell it records the total variation distance, the power ceiling at size 0.05, and first-hit probabilities at the registered deployment budgets.

For small np, the first-order approximation is

$$\tau(n, p) = np + O(n^2 p^2). \tag{142}$$

This makes the ambiguity scale visible: with 256 base rollouts, a mass of 10^{-8} changes the entire data law by less than 2.6×10^{-6}.

N.2 Random coarsening census

We draw two independent vectors of positive exponential variates, normalize each to the simplex, and assign fine cells uniformly to a smaller set of coarse labels. For 100,000 registered trials, the script checks

$$F^{+}_{\text{coarse}} \le F^{+}_{\text{fine}}, \qquad F^{-}_{\text{coarse}} \le F^{-}_{\text{fine}}. \tag{143}$$

The maximum floating-point excess is 2.78×10^{-16}. Strict contraction occurs in 94,150 trials, while equality cases correspond to partitions without opposing signs inside a cell.

N.3 Access/execution algebra census

Each random trial also samples q_0, q_1, μ_0, μ_1 independently from the unit interval and computes both sides of Equation 127. The maximum absolute residual over 100,000 trials is 2.22×10^{-16}. This test covers boundaries near zero and one and is independent of neural response parsing.

N.4 Complete-monotonicity certificate

For the worked spectrum in Equation 76, we compute $\mathcal{D}_0$ through $\mathcal{D}_{16}$ and form signed forward differences through order eight. The minimum value at every order is positive. The order zero through eight minima are

$$\begin{aligned}
(\eta_0, \eta_1) &= (4.58 \times 10^{-2}, 4.08 \times 10^{-3}),\\
(\eta_2, \eta_3) &= (6.57 \times 10^{-4}, 1.58 \times 10^{-4}),\\
(\eta_4, \eta_5) &= (4.92 \times 10^{-5}, 1.81 \times 10^{-5}),\\
(\eta_6, \eta_7) &= (7.37 \times 10^{-6}, 3.15 \times 10^{-6}),\\
\eta_8 &= 1.39 \times 10^{-6}.
\end{aligned} \tag{144}$$

The values agree with the closed form $\sum_a p_a^{r+1}(1 - p_a)^K$. Thus the certificate checks the full sign pattern implied by the spectrum, rather than only the monotone decrease of discovery with budget. Agreement at higher orders also checks numerical cancellation in the finite-difference implementation, where a decreasing discovery sequence alone would be insufficient.

N.5 Crossover certificate

The script evaluates Equation 80 for every integer budget from 1 through 64. The concentrated spectrum leads at budget one and the diffuse spectrum first leads at budget two. It remains ahead through the end of the grid. A second equal-mass pair, (0.50) and $(0.25, 0.25)$, verifies strict Schur-concavity at every budget from two onward.

N.6 Moment-program certificate

The primal grid contains 2,001 equally spaced points in $[0, 1]$ together with the four true atoms. HiGHS solves both directions of Equation 75. Every optimum satisfies all registered moment equalities to solver tolerance, contains no negative weight, and brackets the known target. The active-atom counts for prefix lengths 2, 3, 5, and 8 are 2, 3, 5, and 7 at both extrema, consistent with finite atomic extremizers.

N.7 Independent unit checks

Nineteen deterministic tests cover the total variation identity, limiting power, first-hit transformation, coarsening contraction, access/execution algebra, exact expression parsing, semantic normalization, incorrect declaration rejection, complete monotonicity, occupancy endpoints, and equal-mass diffusion. All tests pass before neural generation begins.

O Enumerable Expression Environment

O.1 Task construction

The deterministic generator enumerates candidate tuples of four distinct integers in $\{1, \ldots, 13\}$ and targets in $\{10, \ldots, 50\}$ under seed 20260910. For each candidate, it enumerates all valid registered entrances and retains the task if their number lies between 4 and 12 inclusive. Generation stops after 512 eligible tasks. The first 128 form the design split and the remaining 384 form the held-out split.

The range restriction prevents the benchmark from consisting solely of either unique-solution puzzles or combinatorially redundant puzzles. It is applied before any checkpoint response is sampled. Table 4 summarizes the resulting fixed panel.

O.2 Recursive exact solver

Each solver state is a tuple (v_i, e_i) of a rational value and its symbolic expression. For every pair of entries, the recursion considers addition, multiplication, both subtraction orders, and both valid division orders. It replaces the pair by the result and recurses until one value remains. A branch is valid exactly when that value equals the integer target.

Table 4: Deterministic expression-task panel.

Quantity	Design	Held-out
Tasks	128	384
Registered modes	882	2,623
Minimum per task	4	4
Median per task	6	6
Mean per task	6.891	6.831
Maximum per task	12	12

To enumerate entrances, the outer loop fixes one operation on two original numbers, inserts its rational result, and asks whether the remaining three-value state has a valid continuation. A witness expression is stored for each canonical entrance. Exact fractions avoid tolerance-dependent mode membership.

O.3 Prompt contract

Two examples establish the response grammar. A task then requests exactly

```
<first>a OP b = result</first>,
<expression>expression</expression>.
```
(145)

The expression must be fully parenthesized and use all numbers exactly once. The declared operation must appear unchanged at a leaf. The model is free to choose any registered or unregistered response; the tags make semantic auditing deterministic rather than constraining decoding.

O.4 Parser order

The parser applies checks in the following fixed order:

1. locate both tags and parse integer entrance operands;
2. verify that both operands occur in the supplied multiset;
3. compare the declared result with exact rational arithmetic;
4. parse an abstract syntax tree containing only integer leaves and binary arithmetic;
5. verify the complete leaf multiset;
6. locate the canonical entrance among leaf-level subtrees;
7. compare the final rational value with the target;
8. require membership in the preregistered nonidentity mode set.

The access indicator becomes one after checks 1 through 6, even if the final target is wrong. Successful execution requires all eight checks.

O.5 Canonicalization and duplicate numbers

Although the released tasks use distinct integers, the parser treats inputs as multisets. It removes both declared operands separately and compares sorted leaf lists. Addition and multiplication sort operands in their semantic label; subtraction and division retain order. Equivalent higher-level association is allowed because only the declared leaf entrance defines the mode.

O.6 Examples

For numbers $(1, 3, 4, 6)$ and target 24, the entrance $3/4$ has witness

$$6/\{1 - (3/4)\} = 24. \tag{146}$$

The declaration $3/4 = 4/3$ fails before expression evaluation. An expression that contains $3/4$ and uses all numbers but evaluates to 9 records access to the mode and failed execution. A correct expression whose declaration names a different leaf operation fails the semantic consistency check.

O.7 Intervention expansion

The held-out intervention manifest is the Cartesian expansion of each task and its registered modes. Its 2,623 rows carry task identity, forced mode, numbers, target, and a prompt produced by one deterministic transformation. The transformation inserts the requested leaf operation and its exact result before the unchanged response contract. This manifest is frozen before intervention generation.

P Checkpoint Lineage and Generation Protocol

P.1 Matched lineages

Table 5 records the checkpoint comparisons used in the final panels. Each pair shares architecture and tokenizer, and the second checkpoint is a documented reinforcement-learning continuation of the first. The analysis concerns the public artifacts under a fixed executable sampler; it does not reconstruct their training data.

Qwen3 supplies a dense 4B architecture with a switchable nonthinking template (Yang et al., 2025). The public mathematical RL artifact is evaluated at its reported step 500 and follows the documented rule-reward construction used in the corresponding nonthinking math setup (Sun et al., 2026). DeepScaleR continues the 1.5B DeepSeek-R1 distilled Qwen checkpoint through scaled reinforcement learning (DeepSeek-AI, 2025; Luo et al., 2025). Model configuration, vocabulary size, special tokens, and tokenizer hashes are checked within each pair before generation.

Table 5: Matched checkpoint lineages. Controlled semantic enumeration uses the Qwen3 pair; both pairs enter MATH-500 and GSM8K.

Pair	Before RL	After RL
Qwen3	Qwen3-4B	Qwen3-4B, RL step 500
DeepScaleR	DeepSeek-R1-Distill-Qwen-1.5B	DeepScaleR-1.5B-Preview

P.2 Primary controlled decoding law

The controlled panel fixes

$$\lambda_{\mathrm{ctrl}} = (T = 1,\ \text{top-p} = 1,\ L_{\max} = 256,\ \text{nonthinking template}). \tag{147}$$

Every task receives 256 responses from each checkpoint. Sampling uses no top-k truncation, repetition penalty, beam search, or best-of selection. The fixed seed constructs reproducible pseudorandom streams, while independent output samples remain the observational units.

The temperature sweep changes only T to 0.5, 0.7, 1.3, or 1.6 and generates 64 responses per task and checkpoint. The nucleus control changes only top-p to 0.9. These cells remain separate distributions $p^{(\lambda)}_{j,x,z}$.

P.3 Real-reasoning decoding law

The MATH-500 panel uses

$$\lambda_{\mathrm{math}} = (T = 1,\ \text{top-p} = 1,\ L_{\max} = 2048). \tag{148}$$

The prompt requests a worked solution followed by a boxed final answer. Chat templates are applied from each matched tokenizer. Qwen3 runs in nonthinking mode for consistency with its RL continuation; the DeepSeek-derived pair uses its native reasoning template. Truncated responses are scored from any boxed answer already present and otherwise count as failures.

MATH robustness repeats 32 responses per problem for Qwen3 at temperature 0.7 and top-$p = 0.9$, and for DeepScaleR at temperatures 0.7 and 1.3 and top-$p = 0.9$. Each cell changes one coordinate from Equation 148.

P.4 Sharding and record identity

Task indices are assigned by deterministic modular sharding. Each process writes a distinct JSONL shard, and merging sorts by

$$\texttt{checkpoint::task-id::sample-index}. \tag{149}$$

The merger rejects duplicate identifiers and verifies the expected record count. Scoring operates only on complete merged files. The Qwen3 controlled panel has 98,304 records per checkpoint; the corresponding intervention panel has 83,936.

P.5 Inference runtime

Generation uses vLLM with bfloat16 model weights and one process per GPU (Kwon et al., 2023). The fallback Transformers path is used only for small compatibility smoke tests and is excluded from reported aggregate cells. Unique compiler caches prevent concurrent processes from sharing incomplete artifacts. All scientific outputs are plain response text and exact scorer records, so GPU type changes scheduling rather than the estimand.

P.6 Checkpoint exclusions

A Qwen2.5-7B GRPO lineage passed configuration and model-card checks but its base artifact did not reliably follow the registered two-tag grammar on the design split. It is excluded from the controlled semantic panel before held-out generation. The exclusion is based on measurement validity, not the sign of a checkpoint contrast. No held-out score from that pair enters a reported result.

Q Statistical Estimands and Decision Rules

Q.1 Primary finite-panel estimands

For each checkpoint, the primary fine probability vector contains the 2,623 joint held-out prompt-entrance masses and one failure mass. The registered functionals are:

$$s_j = 1 - \widetilde{p}_{j,\perp}, \tag{150}$$

$$\overline{\mathcal{C}}_{j,K} = |\mathcal{X}|^{-1} \sum_x \mathcal{C}_K(\boldsymbol{p}_{j,x}), \tag{151}$$

$$\overline{\mathcal{D}}_{j,K} = \overline{\mathcal{C}}_{j,K+1} - \overline{\mathcal{C}}_{j,K}, \tag{152}$$

$$F_g^+ = \sum_{c \in g} [\widetilde{p}_{1,c} - \widetilde{p}_{0,c}]_+, \tag{153}$$

$$F_g^- = \sum_{c \in g} [\widetilde{p}_{0,c} - \widetilde{p}_{1,c}]_+. \tag{154}$$

Here g ranges over fine entrance, operator, and outcome partitions. The reported budgets are powers of two from 1 through 256 for the primary panel and through 64 for the decoding and real-reasoning panels.

Q.2 Coordinate status

Every fine or coarse cell receives one of the three statuses from Equation 27. Amplification requires the simultaneous lower bound on $\widetilde{p}_{1,c} - \widetilde{p}_{0,c}$ to exceed zero. Contraction requires its simultaneous upper bound to fall below zero. Otherwise the cell remains unresolved. No status is changed using a point-estimate magnitude, relative percentage, or selected rollout budget.

An empirical base-unseen and RL-seen indicator is recorded separately:

$$O_c = \mathbf{1}\{N_{0,c} = 0, N_{1,c} > 0\}. \tag{155}$$

This describes the sample and never substitutes for the support indicator in Equation 22.

Q.3 Visibility-curve uncertainty

Equation 14 is computed per task before averaging. This retains the prompt as the population unit and prevents tasks with more successful samples from receiving more weight. Paired prompt-bootstrap intervals use the same resampled task indices for both checkpoints and all budgets, preserving curve dependence.

For exact finite-panel coordinate claims, the Chernoff region from Appendix H is primary. For task-population curve claims, the paired bootstrap is primary. Tables label which interval is shown.

Q.4 Moment identified sets

The exact certificate supplies moment equalities. Neural extrapolation supplies simultaneous intervals $[\ell_k, u_k]$ propagated from coverage uncertainty. The reported future-discovery interval is the entire feasible image of Equation 11. If sampling noise makes the naive interval constraints infeasible, we enlarge all intervals by the smallest common slack that restores feasibility and report that slack. The current controlled panel requires no such repair for in-budget curves.

Q.5 Entrance quantities

For every intervention unit, the free panel estimates access q_j, successful mode mass p_j, and observational execution p_j/q_j when access is observed. The intervention panel estimates intention-to-treat success μ_j^{int}, compliance c_j, and compliant execution. The symmetric observational decomposition is reported only on units with observed access at both checkpoints. Its numerical residual is an implementation audit and must remain below floating-point tolerance.

Q.6 Real-reasoning quantities

For MATH-500 problem x, let $s_{j,x}$ be symbolic answer success probability. We report

$$A_{j,K} = \frac{1}{500} \sum_{x} \{1 - (1 - s_{j,x})^K\} \tag{156}$$

using the unbiased hypergeometric estimator. We also report counts of problems in the four observed cells formed by zero or positive counts at each checkpoint, plus simultaneous probability-change statuses. The former are empirical coverage; the latter are population-probability statements.

Q.7 No result-dependent inclusion

All 384 controlled tasks, 2,623 modes, 500 real problems, checkpoints, decoding cells, and budgets are defined before their held-out scores are examined. Main text space determines aggregation level, while every registered cell remains in the appendix tables or compact result files.

R Complete Controlled-Task Results

R.1 Response audit

Table 6 summarizes every response in the primary controlled panel. Both checkpoints access a syntactically and semantically valid entrance on more than half of the rollouts, while exact target completion is near two percent. This separation supplies many observed entrances with failed execution, which is required for the access analysis.

Table 6: Qwen3 controlled-panel audit; 98,304 responses per checkpoint.

Quantity	Base	RL
Access responses	58,016	55,537
Access rate	0.5902	0.5649
Exact successes	1,917	1,854
Success rate	0.01950	0.01886
Tasks with success	82	78
Observed modes	96	87

The largest failure class is a valid declared entrance followed by an expression with the wrong target: 56,099 base responses and 53,680 RL responses. The declared entrance is absent from the expression in 26,081 and 24,111 responses, and the supplied-number multiset is violated in 13,164 and 17,892. All remaining parser failures together account for 1,043 and 767 responses. This audit locates the low exact rate in arithmetic execution rather than missing output tags.

R.2 Complete visibility curve

Table 7 reports the unbiased expected number of distinct successful modes per prompt. At $K = 1$ the quantity equals success probability. At $K = 256$ it equals the mean observed mode count, 0.250 for the base checkpoint and 0.227 after RL. Every paired task-bootstrap interval for the checkpoint difference contains zero.

The widening point difference reflects accumulation across budgets. Its uncertainty also grows because relatively few prompts expose multiple successful modes. Reporting only $K = 256$ would hide that the one-rollout success contrast is nearly zero and would make the terminal empirical mode count look like a support estimate.

R.3 Observed-only modes and simultaneous status

Across checkpoints, 105 fine successful atoms appear at least once. Seventy-eight appear in both samples, 18 only in the base sample, and 9 only in the RL sample. The latter nine are empirical base-unseen and RL-seen modes. Their simultaneous base probability intervals all have positive upper endpoints, so none identifies absolute support expansion from rollout counts. Because each registered mode has a solver witness shorter than the generation cap and both samplers assign positive

Table 7: Complete primary visibility curve. Intervals are paired 95% task-bootstrap intervals for RL minus base.

K	Base	RL	Difference	95% interval
1	0.01950	0.01886	−0.00064	[−0.00439, 0.00302]
2	0.03249	0.03143	−0.00106	[−0.00667, 0.00458]
4	0.05023	0.04853	−0.00170	[−0.00905, 0.00579]
8	0.07209	0.06970	−0.00239	[−0.01175, 0.00695]
16	0.09740	0.09421	−0.00319	[−0.01489, 0.00844]
32	0.12633	0.12184	−0.00450	[−0.01821, 0.00920]
64	0.16102	0.15290	−0.00812	[−0.02368, 0.00707]
128	0.20311	0.18763	−0.01548	[−0.03491, 0.00298]
256	0.25000	0.22656	−0.02344	[−0.05208, 0.00260]

probability to every token at every prefix, the structural model gives the complementary conclusion $S_0 = S_1 = S_{\text{valid}}$ for this panel. The observed-only cells arise entirely from finite visibility.

One mode is simultaneously amplified after correction across all 2,624 fine cells. On task 164, entrance $9 - 5$ appears in 5 base successes and 87 RL successes. Its pooled atom mass changes by

$$\begin{aligned} \widehat{\Delta} p &= 8.34 \times 10^{-4}, \\ \mathcal{I}_{0.95} &= [2.09 \times 10^{-4}, 1.47 \times 10^{-3}]. \end{aligned} \tag{157}$$

Equation 157 uses the joint uniform-prompt mass needed for signed flow. Multiplying by the known 384-prompt weight gives the task-conditional change 0.320 with interval $[0.080, 0.564]$. Monotone transport gives first-hit change 0.818 at budget eight with interval $[0.218, 0.995]$. At budget 256 the point change falls to 0.006 because both first-hit probabilities approach one. The mode was observed before RL, so this is a probability-amplification certificate without a structural-zero claim.

R.4 Flow at every semantic granularity

Table 8 applies the same augmented probability law, including failure, at all three granularities. Point positive and negative flows are equal, as required by Equation 18. Coarsening contracts them monotonically, as visualized in Fig. 4.

The fine-to-operator contraction removes 75.8% of point flow, and the operator-to-outcome contraction removes another 46.2%. Aggregate exact success therefore retains only 13.0% of the fine semantic movement. The result is a direct empirical instance of Proposition 4.6.

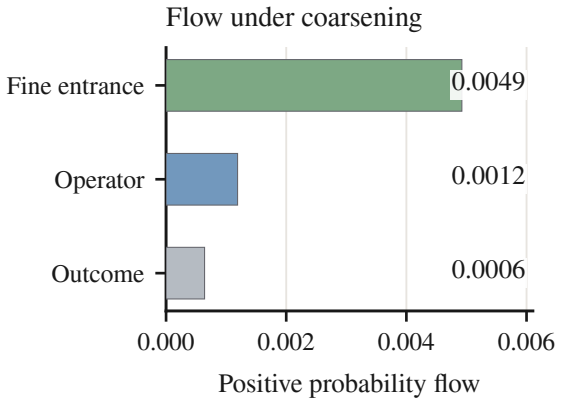

Figure 4: Positive probability flow grows as the semantic partition becomes finer. The negative flow is identical after including failure.

R.5 Visibility contraction across the same hierarchy

No task, checkpoint, or budget violates the ordering predicted by Proposition 4.7. At budget 256, fine visibility is 0.2500 before RL and 0.2266 after RL. Operator visibility is 0.2214 and 0.2109, while outcome visibility is 0.2135 and 0.2031. The fine matched difference is -0.0234; both coarser differences are -0.0104. Coarsening preserves the sign in this panel but removes more than half its endpoint magnitude, consistent with different collision losses. Complete curves for all levels accompany the compact hierarchy result. The flow and visibility hierarchies therefore describe different aspects of the same laws. Flow contracts by cancellation of opposing probability changes, whereas visibility contracts when distinct fine modes become indistinguishable within a coarse cell.

Table 8: Signed probability-flow hierarchy. The last two columns count cells separated by the simultaneous region.

Level	$F^+ = F^-$	**Amp.**	**Contr.**
Fine entrance	0.004924	1	0
Operator	0.001190	0	0
Outcome	0.000641	0	0

R.6 TASK-LEVEL DISTRIBUTION

Most held-out puzzles remain unsolved in a 256-response panel: the base checkpoint succeeds on 82 of 384 and the RL checkpoint on 78. Among solved tasks, a small number dominate the success count. The largest base atom has 216 occurrences and the largest RL atom has 203. Many observed atoms are singletons, which is exactly the regime in which an observed-set cardinality is most sensitive to rollout budget. Complete per-task counts accompany the compact result tables.

R.7 ALL OBSERVED FINE MODES

Table 9 lists the complete 105-cell union of successful modes observed at either checkpoint, in task order. The other 2,518 registered prompt-entrance atoms have zero success counts at both checkpoints and remain present in the compact simultaneous-region table. Listing the observed union makes the empirical set difference auditable without selecting modes by effect magnitude.

Table 9: All 105 observed successful fine modes, in task order. The difference is RL minus base probability multiplied by 10^5.

Task	**Mode**	**Base**	**RL**	$10^5\Delta p$	**Task**	**Mode**	**Base**	**RL**	$10^5\Delta p$
129	`12-1`	101	59	-42.72	133	`12+13`	1	3	2.03
136	`9-11`	1	0	-1.02	137	`10-9`	1	1	0.00
137	`11-10`	1	0	-1.02	144	`9-3`	2	2	0.00
146	`2+3`	1	1	0.00	146	`3*4`	116	33	-84.43
148	`6-4`	0	1	1.02	154	`7-6`	4	1	-3.05
156	`5*11`	143	164	21.36	158	`10+11`	45	5	-40.69
158	`11+13`	1	0	-1.02	158	`12+13`	1	0	-1.02
159	`10-9`	46	123	78.33	161	`11-8`	3	0	-3.05
164	`9-5`	5	87	83.41	169	`12-9`	4	1	-3.05
169	`9-7`	1	0	-1.02	173	`9-2`	4	13	9.16
175	`12-11`	2	1	-1.02	186	`4*7`	1	0	-1.02
200	`10+11`	2	3	1.02	203	`12-4`	1	0	-1.02
206	`12-9`	3	3	0.00	206	`13-12`	1	0	-1.02
210	`9-8`	5	6	1.02	213	`11-3`	13	1	-12.21
214	`13-1`	1	0	-1.02	216	`5-1`	0	1	1.02
222	`2*4`	1	1	0.00	225	`6*7`	1	1	0.00
226	`13-1`	2	1	-1.02	228	`9-3`	1	1	0.00
233	`12/2`	61	61	0.00	234	`5*6`	4	6	2.03
234	`9-7`	3	2	-1.02	239	`13-8`	2	1	-1.02
240	`3*6`	2	0	-2.03	246	`3*9`	32	6	-26.45
250	`12-11`	83	66	-17.29	251	`8-3`	57	68	11.19
255	`10-6`	26	22	-4.07	256	`10-8`	0	19	19.33
259	`11-3`	40	34	-6.10	266	`2*11`	1	1	0.00
267	`11-6`	10	0	-10.17	270	`8/2`	0	2	2.03
277	`7-4`	1	8	7.12	290	`10-7`	216	203	-13.22
292	`10-9`	0	2	2.03	292	`9-4`	3	7	4.07
294	`5+13`	0	1	1.02	297	`12-11`	3	5	2.03
297	`12-13`	6	11	5.09	298	`11-6`	1	1	0.00
298	`6-4`	22	3	-19.33	305	`12-7`	2	2	0.00
308	`8-2`	100	127	27.47	312	`12-11`	3	3	0.00

Continued on next page

Table 9: All 105 observed successful fine modes, in task order (continued).

Task	Mode	Base	RL	$10^5\Delta p$	Task	Mode	Base	RL	$10^5\Delta p$
314	9-7	42	46	4.07	316	8/2	41	37	-4.07
317	7-2	3	2	-1.02	330	4*13	8	4	-4.07
333	13-6	1	0	-1.02	341	13-8	0	1	1.02
346	13-1	1	8	7.12	346	13-2	34	25	-9.16
361	3*8	19	18	-1.02	365	12-11	31	14	-17.29
366	5*7	16	2	-14.24	366	7-5	2	8	6.10
378	10-6	14	12	-2.03	379	9-5	10	4	-6.10
394	8-2	28	12	-16.28	396	13-5	17	3	-14.24
404	13-11	2	2	0.00	406	12-9	4	36	32.55
412	5*9	26	29	3.05	414	6/2	0	1	1.02
415	5+8	4	6	2.03	416	9-2	8	15	7.12
417	13-5	1	1	0.00	421	5-1	1	0	-1.02
424	13-10	3	7	4.07	427	8-7	1	0	-1.02
432	6*7	1	3	2.03	440	12-11	8	6	-2.03
450	13-7	1	1	0.00	455	12-6	35	23	-12.21
456	12/2	16	15	-1.02	460	5-2	16	34	18.31
460	8-4	1	0	-1.02	468	12-2	2	0	-2.03
470	11-7	74	82	8.14	471	10-4	88	105	17.29
472	5*8	10	33	23.40	482	10-7	1	1	0.00
482	10-2	2	2	0.00	483	13-12	4	11	7.12
484	5+11	0	1	1.02	504	8-6	3	0	-3.05
508	11-9	136	65	-72.22	508	9-7	3	4	1.02
510	7-1	6	1	-5.09					

R.8 ALL HELD-OUT TASK ENDPOINTS

Table 10 reports every held-out task in deterministic order. "Valid" is the solver-enumerated mode count. Base and RL columns are the numbers of distinct successful modes observed by budget 256, which equals the endpoint of the unbiased occupancy curve. This complete table shows the sparse task distribution behind the average curve and retains every zero endpoint.

Table 10: Budget-256 visibility for every held-out task. Valid is the solver-enumerated number of modes; the difference is RL minus base observed modes.

Task	Valid	Base	RL	Δ	Task	Valid	Base	RL	Δ
128	4	0	0	0	129	11	1	1	0
130	6	0	0	0	131	5	0	0	0
132	4	0	0	0	133	5	1	1	0
134	12	0	0	0	135	8	0	0	0
136	4	1	0	-1	137	9	2	1	-1
138	8	0	0	0	139	8	0	0	0
140	6	0	0	0	141	6	0	0	0
142	4	0	0	0	143	6	0	0	0
144	12	1	1	0	145	9	0	0	0
146	9	2	2	0	147	5	0	0	0
148	10	0	1	1	149	6	0	0	0
150	9	0	0	0	151	6	0	0	0
152	10	0	0	0	153	9	0	0	0
154	5	1	1	0	155	6	0	0	0
156	8	1	1	0	157	7	0	0	0
158	8	3	1	-2	159	8	1	1	0
160	7	0	0	0	161	5	1	0	-1
162	9	0	0	0	163	6	0	0	0
164	9	1	1	0	165	5	0	0	0

Continued on next page

Table 10: Budget-256 visibility for every held-out task (continued).

Task	Valid	Base	RL	Δ	Task	Valid	Base	RL	Δ
166	6	0	0	0	167	5	0	0	0
168	6	0	0	0	169	5	2	1	-1
170	7	0	0	0	171	4	0	0	0
172	5	0	0	0	173	4	1	1	0
174	7	0	0	0	175	5	1	1	0
176	4	0	0	0	177	5	0	0	0
178	5	0	0	0	179	4	0	0	0
180	12	0	0	0	181	7	0	0	0
182	7	0	0	0	183	4	0	0	0
184	4	0	0	0	185	5	0	0	0
186	5	1	0	-1	187	4	0	0	0
188	6	0	0	0	189	4	0	0	0
190	8	0	0	0	191	4	0	0	0
192	7	0	0	0	193	7	0	0	0
194	12	0	0	0	195	4	0	0	0
196	7	0	0	0	197	5	0	0	0
198	11	0	0	0	199	5	0	0	0
200	8	1	1	0	201	9	0	0	0
202	8	0	0	0	203	10	1	0	-1
204	5	0	0	0	205	4	0	0	0
206	11	2	1	-1	207	9	0	0	0
208	4	0	0	0	209	5	0	0	0
210	6	1	1	0	211	4	0	0	0
212	6	0	0	0	213	11	1	1	0
214	9	1	0	-1	215	8	0	0	0
216	8	0	1	1	217	10	0	0	0
218	10	0	0	0	219	4	0	0	0
220	9	0	0	0	221	4	0	0	0
222	9	1	1	0	223	4	0	0	0
224	6	0	0	0	225	12	1	1	0
226	9	1	1	0	227	11	0	0	0
228	5	1	1	0	229	5	0	0	0
230	5	0	0	0	231	6	0	0	0
232	7	0	0	0	233	6	1	1	0
234	8	2	2	0	235	6	0	0	0
236	6	0	0	0	237	6	0	0	0
238	5	0	0	0	239	12	1	1	0
240	5	1	0	-1	241	5	0	0	0
242	5	0	0	0	243	6	0	0	0
244	4	0	0	0	245	9	0	0	0
246	7	1	1	0	247	5	0	0	0
248	9	0	0	0	249	5	0	0	0
250	8	1	1	0	251	6	1	1	0
252	5	0	0	0	253	5	0	0	0
254	4	0	0	0	255	8	1	1	0
256	10	0	1	1	257	9	0	0	0
258	12	0	0	0	259	7	1	1	0
260	6	0	0	0	261	5	0	0	0
262	5	0	0	0	263	9	0	0	0
264	12	0	0	0	265	12	0	0	0
266	8	1	1	0	267	5	1	0	-1
268	7	0	0	0	269	4	0	0	0
270	7	0	1	1	271	5	0	0	0
272	10	0	0	0	273	7	0	0	0
274	11	0	0	0	275	11	0	0	0

Continued on next page

Table 10: Budget-256 visibility for every held-out task (continued).

Task	Valid	Base	RL	Δ	Task	Valid	Base	RL	Δ
276	12	0	0	0	277	10	1	1	0
278	6	0	0	0	279	12	0	0	0
280	5	0	0	0	281	4	0	0	0
282	4	0	0	0	283	4	0	0	0
284	4	0	0	0	285	6	0	0	0
286	10	0	0	0	287	7	0	0	0
288	4	0	0	0	289	9	0	0	0
290	5	1	1	0	291	4	0	0	0
292	5	1	2	1	293	11	0	0	0
294	4	0	1	1	295	4	0	0	0
296	5	0	0	0	297	5	2	2	0
298	7	2	2	0	299	7	0	0	0
300	9	0	0	0	301	4	0	0	0
302	8	0	0	0	303	10	0	0	0
304	4	0	0	0	305	5	1	1	0
306	5	0	0	0	307	10	0	0	0
308	6	1	1	0	309	8	0	0	0
310	5	0	0	0	311	12	0	0	0
312	5	1	1	0	313	4	0	0	0
314	7	1	1	0	315	9	0	0	0
316	9	1	1	0	317	10	1	1	0
318	11	0	0	0	319	10	0	0	0
320	5	0	0	0	321	7	0	0	0
322	9	0	0	0	323	6	0	0	0
324	4	0	0	0	325	7	0	0	0
326	9	0	0	0	327	5	0	0	0
328	4	0	0	0	329	10	0	0	0
330	8	1	1	0	331	5	0	0	0
332	7	0	0	0	333	6	1	0	-1
334	6	0	0	0	335	4	0	0	0
336	5	0	0	0	337	10	0	0	0
338	5	0	0	0	339	10	0	0	0
340	10	0	0	0	341	5	0	1	1
342	6	0	0	0	343	7	0	0	0
344	4	0	0	0	345	4	0	0	0
346	11	2	2	0	347	4	0	0	0
348	4	0	0	0	349	4	0	0	0
350	4	0	0	0	351	5	0	0	0
352	8	0	0	0	353	10	0	0	0
354	7	0	0	0	355	6	0	0	0
356	9	0	0	0	357	11	0	0	0
358	7	0	0	0	359	10	0	0	0
360	6	0	0	0	361	12	1	1	0
362	4	0	0	0	363	8	0	0	0
364	6	0	0	0	365	12	1	1	0
366	6	2	2	0	367	10	0	0	0
368	5	0	0	0	369	9	0	0	0
370	9	0	0	0	371	5	0	0	0
372	11	0	0	0	373	8	0	0	0
374	4	0	0	0	375	8	0	0	0
376	5	0	0	0	377	5	0	0	0
378	5	1	1	0	379	9	1	1	0
380	9	0	0	0	381	5	0	0	0
382	4	0	0	0	383	4	0	0	0
384	6	0	0	0	385	6	0	0	0

Continued on next page

Table 10: Budget-256 visibility for every held-out task (continued).

Task	Valid	Base	RL	Δ	Task	Valid	Base	RL	Δ
386	11	0	0	0	387	10	0	0	0
388	4	0	0	0	389	4	0	0	0
390	4	0	0	0	391	6	0	0	0
392	10	0	0	0	393	4	0	0	0
394	5	1	1	0	395	5	0	0	0
396	12	1	1	0	397	7	0	0	0
398	6	0	0	0	399	6	0	0	0
400	5	0	0	0	401	5	0	0	0
402	5	0	0	0	403	5	0	0	0
404	5	1	1	0	405	4	0	0	0
406	11	1	1	0	407	9	0	0	0
408	10	0	0	0	409	11	0	0	0
410	7	0	0	0	411	5	0	0	0
412	9	1	1	0	413	9	0	0	0
414	4	0	1	1	415	11	1	1	0
416	9	1	1	0	417	4	1	1	0
418	6	0	0	0	419	5	0	0	0
420	6	0	0	0	421	4	1	0	-1
422	4	0	0	0	423	9	0	0	0
424	10	1	1	0	425	5	0	0	0
426	4	0	0	0	427	12	1	0	-1
428	9	0	0	0	429	5	0	0	0
430	4	0	0	0	431	5	0	0	0
432	4	1	1	0	433	9	0	0	0
434	4	0	0	0	435	10	0	0	0
436	4	0	0	0	437	5	0	0	0
438	5	0	0	0	439	4	0	0	0
440	5	1	1	0	441	4	0	0	0
442	6	0	0	0	443	4	0	0	0
444	4	0	0	0	445	5	0	0	0
446	12	0	0	0	447	9	0	0	0
448	9	0	0	0	449	4	0	0	0
450	4	1	1	0	451	10	0	0	0
452	11	0	0	0	453	5	0	0	0
454	6	0	0	0	455	4	1	1	0
456	7	1	1	0	457	6	0	0	0
458	12	0	0	0	459	4	0	0	0
460	4	2	1	-1	461	4	0	0	0
462	10	0	0	0	463	5	0	0	0
464	4	0	0	0	465	11	0	0	0
466	5	0	0	0	467	9	0	0	0
468	9	1	0	-1	469	8	0	0	0
470	4	1	1	0	471	6	1	1	0
472	4	1	1	0	473	5	0	0	0
474	6	0	0	0	475	6	0	0	0
476	9	0	0	0	477	4	0	0	0
478	11	0	0	0	479	4	0	0	0
480	11	0	0	0	481	4	0	0	0
482	11	2	2	0	483	5	1	1	0
484	9	0	1	1	485	5	0	0	0
486	4	0	0	0	487	8	0	0	0
488	5	0	0	0	489	6	0	0	0
490	10	0	0	0	491	11	0	0	0
492	6	0	0	0	493	10	0	0	0
494	7	0	0	0	495	6	0	0	0

Continued on next page

Table 10: Budget-256 visibility for every held-out task (continued).

Task	Valid	Base	RL	Δ	Task	Valid	Base	RL	Δ
496	4	0	0	0	497	12	0	0	0
498	7	0	0	0	499	4	0	0	0
500	11	0	0	0	501	5	0	0	0
502	5	0	0	0	503	5	0	0	0
504	9	1	0	-1	505	5	0	0	0
506	7	0	0	0	507	12	0	0	0
508	10	2	2	0	509	5	0	0	0
510	6	1	1	0	511	8	0	0	0

S Entrance-Assignment Results

S.1 Complete response counts

The intervention panel contains 83,936 responses per checkpoint: 32 responses for each of 2,623 task-mode units. Table 11 reports the complete intention-to-treat audit.

Table 11: Entrance-assignment summary over all registered units.

Quantity	Base	RL
Compliant responses	56,217	56,122
Compliance rate	0.66976	0.66863
Exact successes	6,218	6,315
Intervention success	0.07408	0.07524
Units with success	601	595
Units with compliance	2,566	2,561

Assigned entrances raise mean success from roughly 0.019 under free generation to 0.075. The unit-average RL-minus-base change is 0.00116. Giving every task equal weight yields a success change of 0.00064 and a compliance change of -0.00799. These summaries distinguish the gain from assigning an entrance from the contrast between checkpoints. The former is large relative to free generation, while the latter is much smaller.

For the task-weighted comparison, the paired 95% bootstrap interval is $[-0.00152, 0.00283]$ for success and $[-0.01580, -0.00016]$ for compliance. Thus the success contrast remains unresolved, whereas the small compliance decrease is separated from zero.

S.2 Failure decomposition

For the base checkpoint, the assigned entrance is correctly used but the target is wrong in 49,999 responses; the corresponding RL count is 49,807. Number multiset violations account for 20,840 and 21,760. The remaining failures are primarily invalid expression forms or absence of the requested leaf operation. Only 10 base and 16 RL responses omit the two-tag protocol, confirming that intervention outcomes are dominated by execution rather than response format.

S.3 Natural observational decomposition

Both checkpoints observe nonzero access for 302 of the 2,623 registered units. On these common-access units, the mean symmetric access component is -0.00179 and the mean observational execution component is 0.00103. Their sum is the mean successful-mode probability change on the same units. The maximum algebraic residual across all 302 cells is

$$6.59 \times 10^{-17}. \tag{158}$$

The opposing signs show why labeling the matched change only as access or only as execution would be incomplete: natural access decreases on average while conditional completion partially offsets it within the common-access subset. These terms are descriptive averages over the 302 shared units. Extending the attribution to all registered entrances requires information about conditional execution on cells that free generation does not expose.

S.4 THE BRIDGE GAP IS MATERIAL

Across common-access cells, the mean gaps $\mu^{\mathrm{int}} - \mu^{\mathrm{obs}}$ are -0.0671 for the base checkpoint and -0.0746 for the RL checkpoint. The registered entrance prompt therefore does not reproduce the naturally selected conditional population. Compliance is about 0.669 rather than one, and assigned units include entrances that are rare under free generation. Equation 133 is needed empirically, not only as a formal caveat.

S.5 CELLS WITHOUT NATURAL ACCESS

Most registered entrances have zero observed natural access at one or both checkpoints, so their observational execution probabilities remain unresolved. The assigned prompt can still measure an operational completion probability for every unit. It cannot retroactively determine whether natural access probability was zero. This division mirrors the paper's central distinction: intervention may reveal executable continuation while finite free rollouts leave spontaneous support unidentified.

S.6 PER-PROTOCOL DIAGNOSTIC

Because noncompliant assigned responses receive outcome zero, exact Equation 135 holds in every unit. We additionally tabulate success conditional on compliance to distinguish refusal of the requested entrance from arithmetic failure after compliance. This diagnostic is descriptive because compliance is a post-assignment event. Aggregated conditional success is 0.11061 for the base checkpoint and 0.11252 after RL, a difference of 0.00192. Multiplying by compliance rates 0.66976 and 0.66863 exactly reproduces the intention-to-treat rates 0.07408 and 0.07524. The small overall increase therefore accompanies slightly better completion among compliant responses despite slightly lower compliance; it is not a causal complier effect.

T REAL MATHEMATICAL-REASONING PANEL

T.1 DATASET AND SYMBOLIC SCORING

MATH-500 contains 500 held-out problems spanning algebra, counting and probability, geometry, intermediate algebra, number theory, precalculus, and prealgebra. The released answer string is parsed once per problem. Each model response is independently parsed for a boxed final answer and compared by the standard symbolic equivalence checker. Parser exceptions receive outcome zero and are counted separately from wrong answers. The checker canonicalizes equivalent exact forms before comparison, so differences in boxed formatting do not create checkpoint contrasts. Prompts, references, parser version, and the failure rule are fixed within each matched pair. Parser exceptions remain a separate audit count rather than being silently merged with ordinary mathematical errors.

T.2 FIRST-HIT ESTIMATOR

For $c_{j,x}$ successes among 64 responses, the unbiased first-hit estimate at budget $K \le 64$ is

$$\widehat{H}_{j,x,K} = 1 - \frac{\binom{64-c_{j,x}}{K}}{\binom{64}{K}}. \tag{159}$$

We average Equation 159 across the 500 problems and bootstrap problems as matched clusters. Per-problem probability changes use a simultaneous KL-Chernoff region over all problems and both checkpoints.

T.3 QWEN3 COMPLETE FIRST-HIT CURVE

Table 12 gives the complete power-of-two curve. The base checkpoint leads at every budget, while the magnitude of the difference decreases from 0.075 at one rollout to 0.020 at 64. The paired interval excludes zero through budget 32 and includes zero at budget 64. The simultaneous max-t band excludes zero through budget 16 and includes zero at budgets 32 and 64.

Table 12: Qwen3 MATH-500 first-hit curve. Intervals are paired 95% problem-bootstrap intervals for RL minus base.

K	Base	RL	Difference	95% interval
1	0.6480	0.5729	-0.0752	$[-0.0948, -0.0561]$
2	0.6942	0.6212	-0.0730	$[-0.0932, -0.0535]$
4	0.7241	0.6634	-0.0607	$[-0.0810, -0.0409]$
8	0.7457	0.6997	-0.0460	$[-0.0665, -0.0257]$
16	0.7631	0.7304	-0.0326	$[-0.0542, -0.0110]$
32	0.7783	0.7539	-0.0244	$[-0.0469, -0.0016]$
64	0.7920	0.7720	-0.0200	$[-0.0440, \ 0.0060]$

Table 13: DeepScaleR MATH-500 first-hit curve. Intervals are paired 95% problem-bootstrap intervals for RL minus base.

K	Base	RL	Difference	95% interval
1	0.4092	0.4379	0.0288	$[\ 0.0177, \ 0.0397]$
2	0.5183	0.5225	0.0042	$[-0.0068, \ 0.0150]$
4	0.6048	0.5920	-0.0129	$[-0.0243, -0.0017]$
8	0.6736	0.6488	-0.0247	$[-0.0376, -0.0122]$
16	0.7282	0.6951	-0.0331	$[-0.0484, -0.0185]$
32	0.7715	0.7323	-0.0392	$[-0.0569, -0.0220]$
64	0.8020	0.7620	-0.0400	$[-0.0600, -0.0200]$

The changing gap is a probability-spectrum effect even at the problem level. RL lowers mean one-attempt success substantially, but repeated sampling recovers many problems from both distributions. The $K = 64$ endpoint is an empirical boundary at that budget, not the limit $K \to \infty$.

T.4 DeepScaleR realizes a neural budget crossover

Table 13 reports the independent 1.5B lineage. RL improves one-rollout success from 0.4092 to 0.4379. Its advantage nearly vanishes by budget two and becomes a statistically separated disadvantage at budget four. The base advantage then grows through budget 64.

At $K = 1$, increased total success mass dominates. At larger K, the base checkpoint exposes a broader set of solvable problems and overtakes the RL checkpoint. The separated signs at $K = 1$ and $K = 4$ provide a direct neural witness for Theorem 4.5 at the pointwise level. The registered max-t band over all seven budgets gives $[0.0147, 0.0428]$ at $K = 1$ and $[-0.0406, -0.0089]$ at $K = 8$, establishing the reversal simultaneously across the curve. Every registered power of two appears in Table 13.

T.5 Observed problem cells

For each matched pair, we report problems seen at both checkpoints, only before RL, only after RL, or at neither checkpoint. These are empirical first-hit cells at budget 64. A base-unseen and RL-seen problem is accompanied by the simultaneous upper bound on its base success probability and is never labeled a structural support expansion.

For the Qwen3 pair, 370 problems appear at both checkpoints, 26 only at the base checkpoint, 16 only after RL, and 88 at neither. The 16 post-only problems are the sample pattern most often described as new capability. With 64 responses per problem and familywise allocation across 500 problems, a zero count retains the simultaneous upper bound

$$p_{0,x} \leq 0.1526. \tag{160}$$

Thus the experiment allows appreciable rare base probability for every post-only problem.

The same joint region is informative about differences. Six problems have strictly positive lower bounds on $s_{1,x} - s_{0,x}$, 34 have strictly negative upper bounds, and 460 remain unresolved. Fig. 5 displays all probability pairs and statuses. The coexistence of 6 amplified problems, 34 contracted

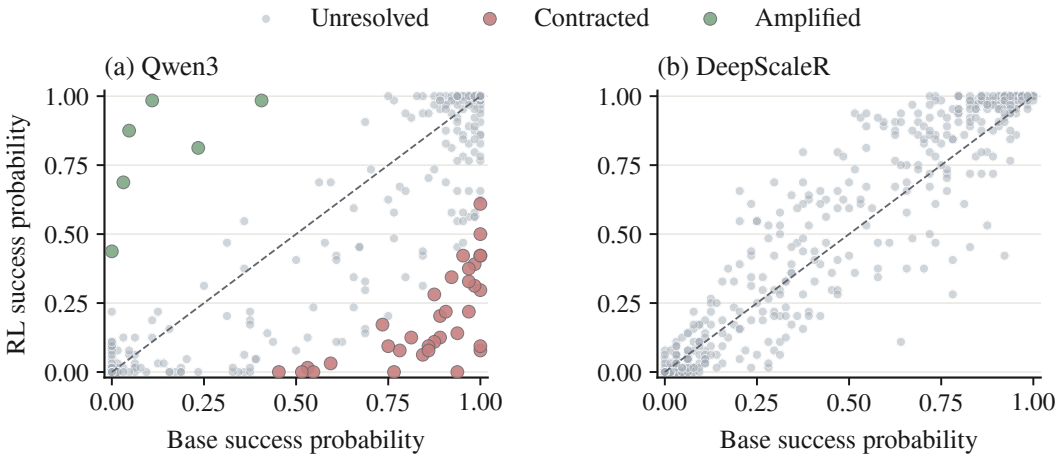

Figure 5: Per-problem empirical success probabilities and simultaneous statuses on MATH-500. (a) Qwen3 contains 6 amplified, 34 contracted, and 460 unresolved problems. (b) Every DeepScaleR problem remains unresolved after familywise correction even though its paired mean curve reverses rank.

problems, and 16 empirically post-only problems is the intended inferential separation: probability movement can be identified even when absolute support cannot.

For DeepScaleR, 376 problems appear at both checkpoints, 25 only before RL, 5 only after RL, and 94 at neither. All 500 per-problem simultaneous difference intervals remain unresolved even though the paired aggregate curve separates at six of seven budgets. This is expected: familywise coordinate inference asks which named problem probability changed, while the paired curve asks whether their panel average changed. The five post-only problems retain the same 0.1526 simultaneous base upper bound and do not identify absolute expansion. Pooling both lineages yields 21 post-only and 51 base-only problem cells. This count asymmetry changes empirical visibility, not structural status: every one-sided cell still admits positive probability at the checkpoint where it was unobserved.

Fig. 5 separates point clouds from inferential status. Many points lie far from the diagonal, yet only changes whose full simultaneous interval clears zero receive an amplified or contracted label.

T.6 Subject and difficulty breakdowns

The compact tables retain subject and nominal MATH difficulty. Each breakdown uses all problems in its cell and the same rollout law. Small categories are reported with intervals rather than pooled into a favorable aggregate.

Fig. 6 connects each subject's one-rollout and 64-rollout contrast. Qwen3 is negative at one rollout in every subject, while some late contrasts approach or cross zero. DeepScaleR has positive one-rollout point contrasts in six of seven subjects and negative budget-64 contrasts in all seven. The aggregate crossover is therefore distributed across the task taxonomy rather than produced by one category.

Tables 14 and 15 provide the paired intervals underlying the visualization. DeepScaleR's one-rollout advantage is positive for every difficulty level and is separated from zero for levels 1, 3, 4, and 5. At budget 64, every level has a negative point contrast, with levels 2 and 5 separated. This repeated sign change across difficulty supports the spectrum-level explanation.

U GSM8K Transfer Panel

U.1 Protocol

The GSM8K test split contains 1,319 problems (Cobbe et al., 2021). Each matched checkpoint produces 32 responses per problem under the MATH-500 symbolic parser, temperature one, and top-p one: 42,208 responses and a complete first-hit curve through budget 32. The prompt, parser,

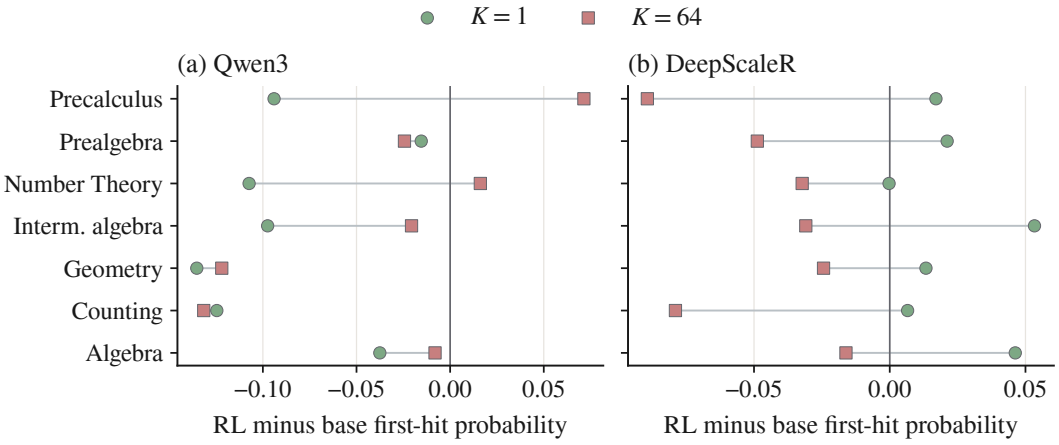

Figure 6: MATH-500 subject contrasts at one and 64 rollouts. Connecting lines show how the same subject changes with deployment budget.

Table 14: Subject-level RL-minus-base first-hit differences. Brackets contain paired 95% problem-bootstrap intervals.

Qwen3 **Subject**	Δ_1	**95% interval**	Δ_{64}	**95% interval**
Algebra	$-.038$	$[-.073, -.003]$	$-.008$	$[-.040, .016]$
Counting	$-.125$	$[-.214, -.046]$	$-.132$	$[-.237, -.026]$
Geometry	$-.135$	$[-.210, -.069]$	$-.122$	$[-.220, -.024]$
Interm. algebra	$-.097$	$[-.140, -.056]$	$-.021$	$[-.093, .052]$
Number theory	$-.107$	$[-.165, -.055]$	$.016$	$[.000, .048]$
Prealgebra	$-.015$	$[-.061, .029]$	$-.024$	$[-.073, .024]$
Precalculus	$-.094$	$[-.155, -.042]$	$.071$	$[-.036, .179]$
DeepScaleR **Subject**	Δ_1	**95% interval**	Δ_{64}	**95% interval**
Algebra	$.046$	$[.026, .067]$	$-.016$	$[-.040, .000]$
Counting	$.007$	$[-.044, .056]$	$-.079$	$[-.184, .000]$
Geometry	$.013$	$[-.009, .040]$	$-.024$	$[-.122, .049]$
Interm. algebra	$.053$	$[.028, .079]$	$-.031$	$[-.072, .010]$
Number theory	$.000$	$[-.043, .041]$	$-.032$	$[-.081, .000]$
Prealgebra	$.021$	$[-.008, .050]$	$-.049$	$[-.098, -.012]$
Precalculus	$.017$	$[-.005, .041]$	$-.089$	$[-.214, .018]$

and sampling coordinates are matched within each pair. This isolates checkpoint differences while retaining every supported power-of-two budget.

U.2 Qwen3 transfer

Qwen3 RL improves one-rollout success from 0.9107 to 0.9204. The point contrast is 0.0097 with paired interval $[0.0027, 0.0166]$ and simultaneous max-t band $[0.0010, 0.0184]$. At budget two the curves nearly coincide, and later point contrasts are small and negative. Their intervals and simultaneous bands all include zero. Table 16 gives every budget.

The early improvement is identified jointly over the curve. The later negative point estimates do not form a second simultaneously separated sign, so GSM8K does not provide another certified budget crossover. It instead shows that probability improvement at one rollout can coexist with an unresolved large budget relation. For a fixed deployment budget, we therefore report its own contrast and uncertainty. The positive one-rollout certificate cannot be carried over to later budgets, even though all estimates use the same generated responses.

Table 15: Difficulty-level RL-minus-base first-hit differences. Brackets contain paired 95% problem-bootstrap intervals.

Qwen3				
Level	Δ_1	**95% interval**	Δ_{64}	**95% interval**
1	$-.017$	$[-.044, .003]$	$-.023$	$[-.070, .000]$
2	$-.011$	$[-.039, .015]$	$.022$	$[.000, .056]$
3	$-.087$	$[-.133, -.043]$	$.000$	$[-.048, .048]$
4	$-.098$	$[-.144, -.053]$	$-.016$	$[-.063, .031]$
5	$-.106$	$[-.145, -.069]$	$-.067$	$[-.134, .000]$
DeepScaleR				
Level	Δ_1	**95% interval**	Δ_{64}	**95% interval**
1	$.032$	$[.001, .061]$	$-.023$	$[-.070, .000]$
2	$.026$	$[-.002, .052]$	$-.044$	$[-.089, -.011]$
3	$.037$	$[.006, .069]$	$-.029$	$[-.067, .010]$
4	$.032$	$[.009, .053]$	$-.031$	$[-.078, .008]$
5	$.020$	$[.003, .038]$	$-.060$	$[-.112, -.015]$

Table 16: Qwen3 GSM8K first-hit curve. Intervals are paired 95% problem-bootstrap intervals for RL minus base.

K	**Base**	**RL**	**Difference**	**95% interval**
1	0.9107	0.9204	0.0097	[0.0027, 0.0166]
2	0.9407	0.9413	0.0006	[−0.0060, 0.0071]
4	0.9571	0.9538	−0.0033	[−0.0098, 0.0031]
8	0.9668	0.9627	−0.0041	[−0.0106, 0.0023]
16	0.9731	0.9696	−0.0034	[−0.0102, 0.0033]
32	0.9765	0.9742	−0.0023	[−0.0099, 0.0053]

Table 17: DeepScaleR GSM8K first-hit curve. Intervals are paired 95% problem-bootstrap intervals for RL minus base.

K	**Base**	**RL**	**Difference**	**95% interval**
1	0.7134	0.7072	−0.0062	[−0.0132, 0.0007]
2	0.8199	0.8168	−0.0031	[−0.0093, 0.0030]
4	0.8853	0.8821	−0.0032	[−0.0088, 0.0026]
8	0.9245	0.9212	−0.0033	[−0.0093, 0.0026]
16	0.9481	0.9451	−0.0030	[−0.0100, 0.0041]
32	0.9629	0.9621	−0.0008	[−0.0099, 0.0083]

U.3 DeepScaleR transfer

DeepScaleR changes one-rollout success from 0.7134 to 0.7072. The point difference is -0.0062 with interval $[-0.0132, 0.0007]$. Every later contrast is smaller in magnitude, and all intervals contain zero. Table 17 reports the full curve.

Fig. 7 makes the task contrast visible. MATH-500 exposes a large Qwen3 deficit and a DeepScaleR crossover, while GSM8K is close to saturation by modest budgets. The theory permits all three curve shapes because they arise from different problem-level mass spectra.

U.4 Observed-only problem cells

For Qwen3, 1,274 problems are observed at both checkpoints, 14 only before RL, 11 only after RL, and 20 at neither. Five problem probabilities are simultaneously amplified, two are contracted, and 1,312 remain unresolved. For DeepScaleR, 1,250 appear at both checkpoints, 20 only before RL, 19 only after RL, and 30 at neither; all 1,319 coordinate differences remain unresolved.

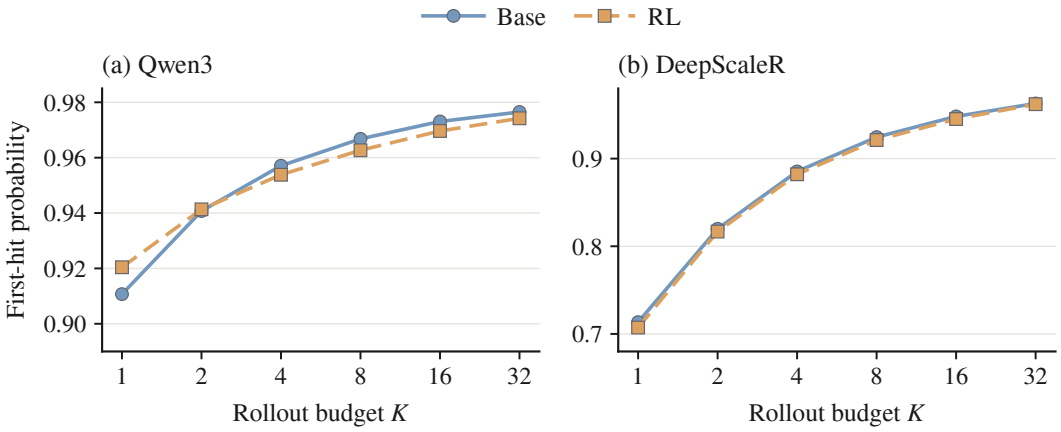

Figure 7: GSM8K first-hit curves for both matched checkpoint lineages. Qwen3 has a small one-rollout RL advantage that vanishes by budget two; DeepScaleR remains nearly matched across the full budget range.

With 32 rollouts and familywise correction across 1,319 problems, a zero count retains simultaneous upper probability 0.3033. The post-only empirical counts are therefore especially far from structural-zero certificates. The Qwen3 coordinate region still identifies seven named probability changes, again showing that difference inference and support inference have distinct power.

U.5 Task transfer conclusion

GSM8K adds a higher-accuracy, shorter-reasoning regime. Its first-hit curves saturate more quickly than MATH-500, leaving less room for late-budget ordering changes. This task dependence is predicted by the visibility spectrum: when most problem masses are already high, $(1-p)^K$ decays rapidly and later discovery becomes small for both checkpoints.

V Checkpoint and Model Replications

V.1 Qwen3 controlled lineage

The Qwen3 pair supplies the complete semantic panel because both checkpoints reliably follow the fixed two-tag grammar on design tasks. The main results use 196,608 natural responses and 167,872 assigned-entrance responses. Every reported fine atom is defined by the solver before held-out generation.

V.2 DeepScaleR real-reasoning lineage

DeepScaleR provides a different parameter scale and a documented scaled-RL continuation of DeepSeek-R1-Distill-Qwen-1.5B. Its native reasoning responses do not reliably obey the artificial two-tag expression grammar, so the replication uses the ordinary MATH-500 boxed-answer protocol. A design smoke of eight problems and two responses per problem produces five correct answers from each checkpoint, confirming parser and prompt compatibility before the full panel.

The full panel changes one-rollout success from 0.4092 to 0.4379, then reverses the ordering by budget four. At budget 64, expected problem visibility is 0.802 before RL and 0.762 after RL. Appendix Table 13 contains the full curve, while Fig. 3 compares it directly with the Qwen3 lineage.

V.3 What replication means here

The theorem package is checkpoint-agnostic. Neural replication asks whether the same inferential pattern appears: many observed-only empirical cells, positive zero-count upper bounds, a complete first-hit curve, and possible separation of some probability changes without a structural-zero conclusion. It does not require the two training runs to have the same accuracy effect.

Table 18: Budget-64 controlled visibility across temperature. Intervals are paired 95% task-bootstrap intervals for RL minus base.

T	Base	RL	Difference	95% interval
0.5	0.0885	0.0938	0.0052	$[-0.0130, 0.0234]$
0.7	0.1146	0.1250	0.0104	$[-0.0078, 0.0286]$
1.0	0.1610	0.1529	-0.0081	$[-0.0237, 0.0071]$
1.3	0.1979	0.2161	0.0182	$[-0.0078, 0.0443]$
1.6	0.2500	0.2396	-0.0104	$[-0.0417, 0.0208]$

V.4 Lineage-specific and common objects

Architecture, scale, base instruction tuning, and RL recipe differ across lineages. The common objects are the fixed prompt panel, within-pair tokenizer, within-pair decoding law, exact answer verifier, rollout budgets, and statistical rules. We report each matched contrast separately before any pooled summary.

W Decoding and Semantic Robustness

W.1 Temperature sweep

Table 18 reports expected mode coverage at budget 64 under five temperatures. The $T = 1$ row uses the primary 256-response panel through its unbiased $K = 64$ estimator; other rows use independent 64-response panels. All sampling laws use top-$p = 1$. Fig. 8 shows the matched contrast and its intervals.

Coverage rises with temperature for both checkpoints, while the checkpoint contrast changes sign across registered laws and remains unresolved in every cell. The pattern supports treating the decoding law as part of the measurement definition. Selecting one temperature after comparing signs would replace one ambiguity with a selection artifact.

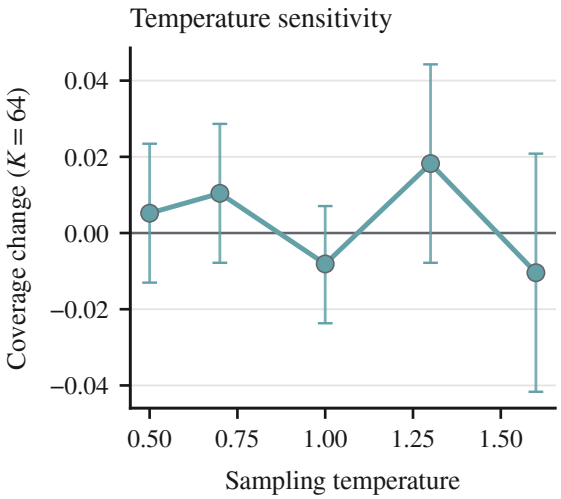

Figure 8: The budget-64 visibility ordering changes across registered temperatures. Every paired task-bootstrap interval includes zero.

W.2 Nucleus truncation

The top-$p = 0.9$ control repeats 64 responses per held-out task at temperature one. It tests whether removing the lowest-probability token tail changes the observed-only mode counts or the sign of the visibility contrast. At budget 64, both checkpoints expose 0.10677 modes per task, with paired difference 0 and interval $[-0.01823, 0.01823]$. Each sample contains 41 distinct successful fine atoms. Under full-softmax sampling, the corresponding expected coverages are 0.16102 and 0.15290. Nucleus truncation therefore removes visible tail modes and erases the small point contrast without identifying any population support change. Table 19 gives the matched values.

W.3 MATH-500 decoding-law replication

Fig. 9 compares the complete matched first-hit contrasts under every registered MATH decoding law. Qwen3's negative curve is nearly unchanged at lower temperature and under nucleus truncation. DeepScaleR is more sensitive: the default and top-$p = 0.9$ laws show certified crossovers, lower temperature identifies the early RL advantage but leaves the late sign unresolved, and temperature 1.3 leaves the aggregate curve unresolved.

Table 19: Budget-64 visibility under nucleus truncation at temperature one.

Top-p	Base	RL	Δ	95% interval
1.0	.161	.153	−.008	[−.024, .007]
0.9	.107	.107	.000	[−.018, .018]

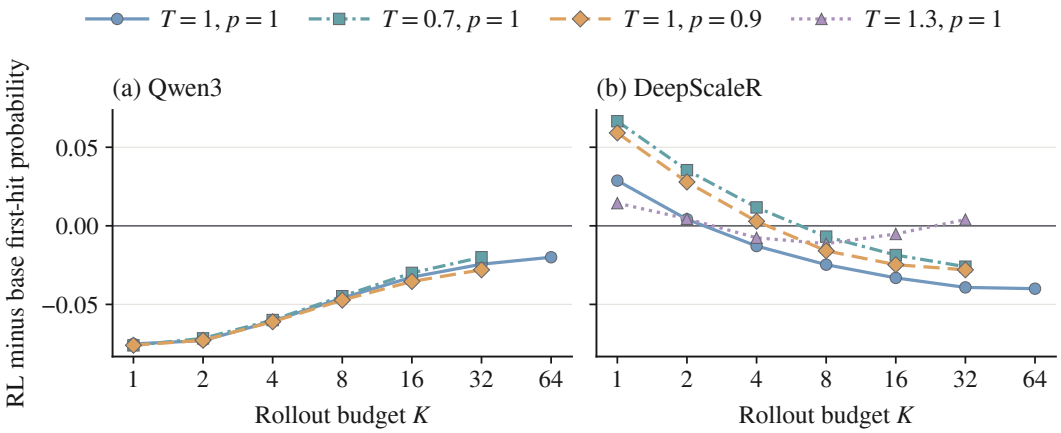

Figure 9: MATH-500 RL-minus-base first-hit curves under registered decoding laws (p in the legend denotes top-p). Qwen3 preserves its narrowing deficit. DeepScaleR preserves the point crossover except at temperature 1.3, while simultaneous certification depends on the law as reported in Table 20.

Table 20: MATH-500 decoding robustness. "Separated" lists budgets whose 95% max-t band excludes zero, with the sign of RL minus base. Primary rows have 64 responses per problem; other rows have 32.

Pair	T	Top-p	Δ_1	Δ_{last}	Separated budgets
Qwen3	1.0	1.0	−0.0752	−0.0200	− at 1, 2, 4, 8, 16
Qwen3	0.7	1.0	−0.0761	−0.0200	− at 1, 2, 4, 8, 16
Qwen3	1.0	0.9	−0.0760	−0.0280	− at 1, 2, 4, 8, 16
DeepScaleR	1.0	1.0	0.0288	−0.0400	+ at 1; − at 8, 16, 32, 64
DeepScaleR	0.7	1.0	0.0666	−0.0260	+ at 1, 2
DeepScaleR	1.3	1.0	0.0144	0.0040	none
DeepScaleR	1.0	0.9	0.0591	−0.0280	+ at 1, 2; − at 16

The Qwen3 agreement shows that its deficit is not produced by one truncation choice. The DeepScaleR top-p control supplies an independent crossover certificate: the simultaneous band is positive at budgets one and two and negative at budget 16. The lower-temperature curve has the same point reversal but its late simultaneous interval includes zero. At temperature 1.3, both checkpoints lose substantial one-rollout accuracy and the matched curve oscillates near zero, locating a weakly identified sampling regime.

W.4 Fine and coarse semantics

The primary hierarchy treats nonidentity leaf operations as fine modes. A secondary parser includes multiplication and division by one. Another coarsening retains only the operator, and the outcome map retains only exact success. Proposition 4.6 predicts monotone contraction across these maps. The coarsening guarantees concern nested maps; expanding the parser's accepted mode universe is a separate robustness check.

The expanded partition changes absolute observed counts but leaves the central inference unchanged: zero observed base count retains a positive upper probability bound. Keeping the parser fixed within each comparison makes this a change of semantic resolution, rather than a change in the generated response sample.

W.5 BUDGET GRID

All powers of two supported by the rollout panel are reported. Adding every integer budget produces a visually denser curve but no new underlying statistic, because Equation 14 determines every in-budget value. The complete integer grid is retained in plotting data, while tables show powers of two for readability.

W.6 PROMPT FORMAT

The 128-task design split compares raw few-shot continuation and the model's chat template. The Qwen3 pair follows the two-tag grammar more reliably with the nonthinking chat template, which is frozen before held-out generation. The comparison is a measurement-validity decision based on design data and never uses a held-out checkpoint contrast.

X ADDITIONAL CONSTRUCTIONS AND BOUNDARY CASES

The following constructions separate the probability objects used throughout the paper and identify cases in which finite-prefix ambiguity disappears.

Identical Pass@K, different visibility. Consider

$$\boldsymbol{p} = (0.4, 0.1), \qquad \boldsymbol{q} = (0.25, 0.25). \tag{161}$$

Both spectra have success mass 0.5 and hence identical answer Pass@K (Equation 59). Nevertheless,

$$\mathcal{C}_2(\boldsymbol{p}) = 0.84, \qquad \mathcal{C}_2(\boldsymbol{q}) = 0.875. \tag{162}$$

The less concentrated spectrum exposes more distinct modes.

Identical observed sets, different futures. A sample containing each of two modes once is compatible with both $(0.5, 0.5)$ and $(0.99, 0.01)$. The observed cardinality is two in either case, but future discovery differs. Counts and their uncertainty therefore contain information lost by an observed-set summary.

Post-only observation with unchanged support. For base mass ε and RL mass $c\varepsilon$, where $c > 1$ and $c\varepsilon \leq 1$, independent samples of size n produce zero base counts and at least one RL count with probability

$$(1-\varepsilon)^n\{1-(1-c\varepsilon)^n\}. \tag{163}$$

This event can be common despite identical support. Conversely, a genuinely new mode of mass ε appears by budget K with probability $H_K(\varepsilon) \to 0$ as $\varepsilon \to 0$. Thus unchanged support can accompany apparent novelty, while actual expansion can have negligible finite-budget impact.

Coarse improvement with fine contraction. For two successful modes followed by failure, let

$$\boldsymbol{p}_0 = (0.15, 0.15, 0.70), \qquad \boldsymbol{p}_1 = (0.29, 0.01, 0.70). \tag{164}$$

Success is unchanged, although one mode gains mass and the other loses it; concentration reduces coverage for $K \geq 2$. If failure instead drops to 0.65 and the first mode rises to 0.34, one-rollout success improves while coverage can still fall at larger budgets, as in Theorem 4.5.

Conditioning and assignment can reverse order. Suppose $(\mu^{\mathrm{obs}}, \mu^{\mathrm{int}})$ changes from $(0.8, 0.3)$ to $(0.7, 0.4)$. Natural conditional execution falls while assigned success rises because the selected populations differ. The bridge gaps, -0.5 and -0.3, reconcile these changes through Equation 133.

Finite moments can determine a boundary tail. If $\mathcal{D}_0 = 1$ and $\mathcal{D}_1 = 0$, the representing measure is concentrated at zero and all later moments vanish. More generally, exposed boundary moment vectors can identify a finite atomic measure from a short prefix. Accordingly, the bounds in Theorem 4.4 may coincide; finite-prefix ambiguity is not universal.