When Is Test-Time Allocation Actionable? Exchangeability and Breadth Dominance in Hierarchical Generation

Anonymous authors
Paper under double-blind review

Abstract

Retrospective oracle gains and accurate value prediction do not by themselves establish that adaptive test-time allocation can improve generation. We audit when information that distinguishes actions becomes available. Under joint conditional exchangeability of future child observations and costs, unseen child labels reduce to one expand action per parent. With hidden fresh randomness and no candidate observation before spawn, policies cannot select a particular unseen parent realization. These conditions restrict action identity but do not establish breadth optimality. In a frozen two-stage TRELLIS generator, value models achieve pooled Spearman correlations of 0.935–0.970, while only 0.108–0.534% of retrospective target variation lies within decision sets. For the primary input-silhouette proxy, the empirical within-parent sum-of-squares share is 4.038% and the ANOVA one-child ICC is 0.959. Three development attempts fail to improve on the selected heuristic comparator. A one-shot 80-unit test of a common spawn-first controller with learned branch selection yields a normalized-utility gain of 0.004281 (95% CI [0.000903, 0.008504]) over uniform branching. Raw silhouette-Chamfer improves by 0.409%, with an interval including zero. A zero-parameter rule attains 92.2% of the learned gain's point estimate; the incremental benefit of the 76-feature Q model remains unresolved. These are finite-table replay results under an estimated cost contract, not end-to-end latency or perceptual quality improvements. The study provides an actionability audit and evidence of small tested-policy gains in one breadth-dominant hierarchy.

1 Introduction

Test-time scaling improves a frozen generator by spending additional compute on sampling, search, or verification. Best-of-N widens the candidate pool; verifier-guided methods instead concentrate computation on candidates that look promising (Snell et al., 2025; Ma et al., 2025; Chen et al., 2025). A common motivation for adaptivity is a large retrospective gap between a simple policy and an oracle that sees every sampled outcome. A second is high predictive accuracy of a learned verifier. Both are incomplete diagnostics: the oracle may exploit information that was unavailable when an action had to be chosen, and pooled accuracy may be dominated by differences that are shared by all actions in a decision set.

Hierarchical generators make this distinction concrete. A first-stage *spawn* action creates a reusable parent, while a cheaper *expand* action samples another child from an observed parent. The controller must trade breadth against repeated sampling. Yet the realized quality of a fresh parent does not exist before spawn, and future children of a fixed parent are sampled by the same conditional generator. The relevant question is therefore not only whether outcomes vary, but whether that variation is distinguishable *before* the policy commits compute.

We call this distinction *actionability*. Section 3 defines the policy-visible information contract. Under conditional child exchangeability, naming different unobserved child slots cannot create distinct actions. Under a constant pre-spawn candidate payload, a policy cannot rank specific unseen parent realizations. These statements do not eliminate the decision problem: a policy may still decide how

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.

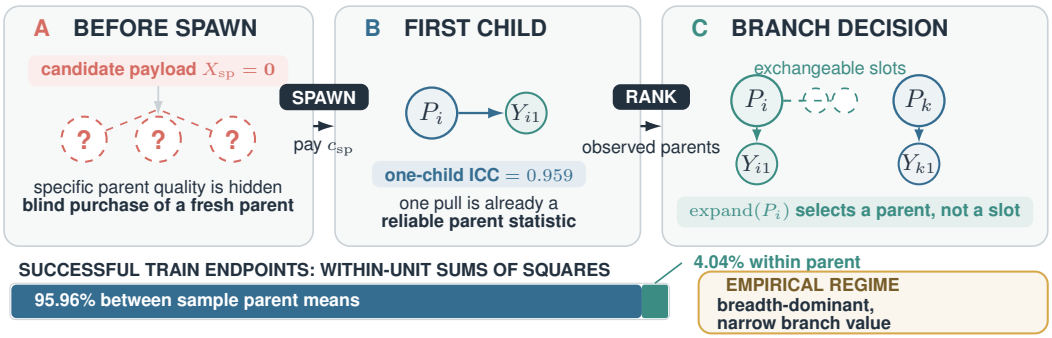

Figure 1: **Information timing and the observed regime.** Under the pre-sampling contract, a fresh parent has no candidate observation (27 payload fields are zero). Observed children permit comparisons across parents; unseen slot labels remain exchangeable. The bar partitions empirical sums of squares, not latent variance components. The separately estimated ICC is 0.959. Breadth dominance and small tested branch gains are empirical findings, not logical consequences of exchangeability or a high ICC.

many parents to spawn and may rank parents after observing a first child. They identify which action identities the contract permits, but do not identify the size of the remaining learnable gain.

We study this boundary in a frozen two-stage 3D generator, TRELLIS (Xiang et al., 2025), using 200 conditioning units split physically into 80 development, 40 calibration, and 80 test units. Each unit contains four stage-1 parents and four independently seeded stage-2 children per parent. The primary endpoint is an input-silhouette Chamfer proxy; three frozen structural proxies test metric sensitivity. Comparisons use recorded heterogeneous action costs, equal nominal budgets, explicit failure conventions, and conditioning-unit bootstrap intervals. We distinguish the online-cost B0 replay from legacy retrospective action-support restrictions at B1.

Our contributions are fourfold.

- We formalize an information-contract audit for hierarchical test-time allocation: an action-space quotient for exchangeable child slots, a non-ranking result for unseen parents without candidate observations, and an exact decomposition of oracle headroom into actionable and hindsight-only components.
- We expose a value–decision dissociation. Pooled Q-label rank correlation reaches 0.935–0.970, while only 0.108–0.534% of total label variance lies within decision sets. On the primary metric, within-parent raw-endpoint sum-of-squares share is 4.038% [2.226%, 6.552%] among successful children; a separate ANOVA one-child ICC estimate is 0.959.
- We report the complete chronological attempt sequence. One-step expected improvement, a continuation-only retrospective correction, and finite-horizon Q fail in different ways when allowed to choose both action type and branch. A factorial replay localizes the dominant loss to spawn/expand choice within the tested replay rather than to an absence of all branch signal.
- We confirm the resulting narrow intervention on held-out data. With composition fixed, Q-based branch selection improves the precommitted normalized primary over equal-budget uniform breadth, but a zero-parameter current-utility rule captures 92.2% of that gain's point estimate. Its own comparison is a prespecified secondary result; neither equivalence to Q nor the absence of a useful learned increment is established.

2 Related Work

Adaptive test-time compute. Best-of-N, verifier-guided search, and prompt-adaptive allocation spend inference compute non-uniformly across candidates or inputs (Snell et al., 2025; Raman et al., 2025; Qu, 2026; Bilal et al., 2026). Verification granularity interpolates between terminal selection and frequent intermediate pruning (Chen et al., 2025). In visual generation, inference-time search

selects promising diffusion noises (Ma et al., 2025), and group inference uses intermediate predictions to prune candidate pools (Parmar et al., 2026). Relevant to breadth and cost, Flash-BoN creates inexpensive independent drafts and selectively refines them, showing that wall-clock and verifier overhead can favor breadth (Rawal et al., 2026). Our contribution is not the first adaptive best-of-N method or the first breadth-favoring result. We instead audit a *reusable-parent* hierarchy. Flash-BoN drafts expose candidate-dependent outputs before refinement; by contrast, a TRELLIS spawn has no candidate-level observation before purchasing the parent. Children then share the realized parent, and exchangeability quotients their unnamed future slots. Our held-out test measures the gain of specific policies within this information contract and a restricted replay cost model, not all attainable allocation value or full wall-clock efficiency.

Hierarchical and many-armed sampling. Knowledge-gradient methods value a measurement by its expected effect on the terminal decision and can share information through hierarchical attributes (Mes et al., 2011). Infinitely-many-armed bandits explicitly trade discovering new arms against repeated sampling of observed arms (Wang et al., 2008); many-arm analyses show regimes in which simple greedy rules are unexpectedly strong (Bayati et al., 2020). Progressive widening likewise controls the addition of actions and stochastic successors in tree search (Couëtoux et al., 2011). These settings include cumulative regret and Bayesian terminal selection; knowledge-gradient methods already address the latter. Our objective instead selects the best realized child in a finite generated table, not the arm with the best posterior mean. Rather than proposing another bandit algorithm, we audit which candidate observations the generator exposes before a decision. Discovery versus repeated pulls and the strength of greedy breadth are therefore not our novelty; the delta is deciding when an uninstantiated generator parent becomes an identifiable arm and separating a clairvoyant terminal gap from pre-action policy value. Our ICC estimate is descriptive: neither a high nor a low ICC alone determines adaptive allocation value.

Selection-induced error. Verifier imperfection has been related to the geometry of its ROC curve rather than to a single accuracy number, and low-compute behavior need not predict high-compute best-of-N performance (Dorner et al., 2026). Our audit is complementary: even a high pooled rank correlation can be dominated by hierarchy levels shared by every action in a choice set. We therefore audit information timing and conditional order before asking how verifier error scales with N. Maximizing noisy estimates creates the optimizer's curse (Smith & Winkler, 2006); max-induced value overestimation also motivates Double Q-learning (van Hasselt, 2010). We observe the same selection geometry in a non-RL replay, but do not transfer Double-Q guarantees: our split forests share training data, so this experiment does not isolate independent estimation noise or implement a Double-Q learning algorithm. More broadly, our results distinguish predictive accuracy from conditional choice quality: an argmax consumes within-decision-set order, not pooled between-instance correlation.

3 Problem Setting and Information Contract

Let x be a conditioning input. Stage 1 samples a reusable parent $P_i \sim p_1(\cdot \mid x)$. Conditional on P_i, stage 2 samples terminal children

$$Y_{ij} = u(G_2(P_i, \epsilon_{ij})), \qquad \epsilon_{ij} \overset{\text{iid}}{\sim} p_\epsilon, \tag{1}$$

where larger Y is better. A *spawn* action draws a fresh parent and its first child at cost c_{sp}; *expand*(i) draws one additional child from an observed parent at cost c_{ex}. Given nominal budget B, a policy maximizes the best credited terminal utility,

$$\mathcal{V}(\pi) = \mathbb{E}_\pi\left[\max_{(i,j)\in\mathcal{O}_T} Y_{ij}\right], \qquad \mathcal{O}_T = \text{children credited within the replay budget.} \tag{2}$$

For fixed known costs this reduces to the usual constraint $\sum_t c(a_t) \le B$. With uncertain costs, our replay instead admits an action when $C_t + \widehat{c}(a_t) \le B$, where C_t is recorded completed cost. If the attempted action finishes beyond B, its endpoint is not credited and the replay stops at the pre-action state. Reported completed utilization omits that final attempted action, so it is not a measurement of all elapsed work.

The history $\mathcal{H}_t$ contains the conditioning input, realized parents, observed child utilities and morphology, completed action costs, remaining nominal budget, and failure records. It excludes future seeds, noise realizations, children, and realized future costs. We denote by Π_{pre} the class of policies measurable with respect to this pre-action history. This policy class is essential: a controller that peeks at a partial denoising state receives realization-specific information and is outside Π_{pre}.

Our TRELLIS instantiation uses a sparse stage-1 structure as P_i and a 25-step structured-latent sample as its child. The sampler, CFG strength, and decoder are fixed; child seed is the only varying stage-2 sampling input. Joint exchangeability of the resulting observations and costs is a modeling assumption; nondeterministic replay error is measured separately. The two online action costs are heterogeneous. Development medians are 13.7518 s for spawn and 5.7676 s for expand; confirmatory replay uses the more conservative, frozen q95 contract of 19.1849 s and 11.2277 s. All policies in a comparison receive the same nominal budget and the same cost rule. Costs include sampling, Gaussian decoding and RGB rendering, but omit CPU metric extraction, feature computation and forest prediction, as well as conditioning, model loading and serialization. They do not establish an end-to-end deployment budget. B1 additionally inherits a retrospective manifest restriction, detailed in Section 6.

4 Actionability Under Conditional Exchangeability

4.1 What can be proved from the generation contract

Proposition 1 (Exchangeable-slot quotient)**.** *Fix an observed parent P and history $\mathcal{H}_t$. Suppose its unobserved children's complete observable outcomes, failure indicators and costs are jointly conditionally exchangeable given $(P, \mathcal{H}_t)$. Assume feasibility, stopping, side observations and the terminal objective are invariant to child labels, and no action exposes label-specific side information. For any permutation of the unobserved slots, a policy and its counterpart with all present and future slot references permuted induce the same distribution over label-free histories, costs, and terminal utility. Thus all unobserved child slots of P form one decision-equivalence class:* expand(P).

Proof. Apply a permutation that swaps the two named slots and leaves every observed variable fixed, and apply the same permutation to all later policy references to those slots. Joint exchangeability leaves the full future observation and cost law invariant, and feasibility and the objective ignore labels. The two relabelled trajectories therefore have the same distribution. Iterating this construction quotients all unseen slots of the parent into one action. □

Proposition 2 (No ranking of a specific unseen realization)**.** *Let X_{sp} be the candidate-level observation available before spawning a fresh parent. If $X_{\mathrm{sp}} = x_0$ almost surely and the new parent realization is drawn only after the action using fresh randomness conditionally independent of the history and policy randomization, then any Π_{pre}-measurable rule has the same conditional distribution over the quality of that specific new parent as an unlabelled fresh draw. The rule may decide* whether *to spawn, but cannot rank which unseen parent realization will be obtained.*

Proof. An admissible pre-spawn action is independent of the fresh realization conditional on its observed history. The conditioning argument is given in the appendix; it does not constrain history-dependent scheduling. □

These propositions are structural, not distributional claims about an optimal depth. In particular, exchangeability alone does not imply pure breadth: the tail of $Y \mid P$, the distribution of parent means, costs, budget, and the best-terminal objective all affect the value of repeating a parent.

4.2 Oracle headroom is only an upper bound on actionable value

Let π_0 be a deployable baseline, let $V^*_{\mathrm{pre}} = \sup_{\pi \in \Pi_{\mathrm{pre}}} \mathcal{V}(\pi)$, and let π_Ω be a retrospective oracle that observes the complete realized endpoint table before choosing among the same feasible actions under the same cost contract. Then the familiar oracle gap has the exact decomposition

$$\underbrace{\mathcal{V}(\pi_\Omega) - \mathcal{V}(\pi_0)}_{\text{retrospective headroom}} = \underbrace{\mathcal{V}(\pi_\Omega) - V^*_{\mathrm{pre}}}_{\text{hindsight-only gap}} + \underbrace{V^*_{\mathrm{pre}} - \mathcal{V}(\pi_0)}_{\text{actionable headroom}} . \tag{3}$$

The identity is elementary but changes the empirical burden: a large left-hand side establishes outcome dispersion, not a large actionable term. Candidate information and conditional ranking must be audited separately. We do not identify V^*_{pre} or either right-hand term over the unrestricted policy class; empirically we estimate only values and contrasts for the tested frozen policies. Equation 3 therefore supplies an upper-bound warning and an audit target, not a numerical decomposition of our TRELLIS results.

4.3 Auditing the TRELLIS contract

Exchangeability and index drift. For a fixed parent, every child uses the same 25-step sampler and CFG strength; only an independent seed changes. We nevertheless test for implementation drift. Across 318 complete Train parents, the parent-fixed raw-Chamfer trend is 2.20×10^{-5} per child index, with unit-clustered 95% interval $[-8.12 \times 10^{-5}, 1.25 \times 10^{-4}]$. Within-parent permutation tests give $p = 0.714$ for a linear trend and $p = 0.982$ for any index effect. Each index has 318 successes and two failures, and the other three proxies also show no detectable trend. These checks detect no index drift at this resolution; failure to reject does not prove equality or independence. The common sampler motivates the exchangeability assumption, which also requires invariant runtime and observation behavior.

Action-space consequence. Proposition 1 predicts that two distinct legal expand actions must refer to different parents. The implementation audit confirms this for all 954/954 aligned multi-branch decision sets: the cross-parent action-pair share is exactly one. This is an action-space audit, not an estimated dose response.

Spawn-side information. The runner represents a prospective spawn with 27 candidate morphology and quality fields. All 27 are exactly zero in both fitting and replay, closing the train–replay mismatch alternative for that payload. The additional fresh-noise and no-side-information assumptions in Proposition 2 are separate from this feature check. Input- and state-level features may describe the distribution of a fresh parent, but no feature identifies the quality of the particular realization that will be drawn.

4.4 Variance–information mismatch

Let X denote the conditioning unit. The within-unit outcome variance obeys

$$\mathbb{E}_X[\mathrm{Var}(Y \mid X)] = \underbrace{\mathbb{E}_X\left[\mathrm{Var}_{P|X}\{\mathbb{E}(Y \mid P, X)\}\right]}_{\sigma_b^2} + \underbrace{\mathbb{E}_{X,P}[\mathrm{Var}(Y \mid P, X)]}_{\sigma_w^2}. \tag{4}$$

We report an empirical sum-of-squares share, which is not the population component ratio in Equation 4. For the 1,272 successful Train children from 318 complete parents and 80 units, let $\bar{Y}_{ui}$ and $\bar{Y}_u$ be sample parent and unit means. Then

$$\widehat{s}_w^{\mathrm{SS}} = \frac{\sum_{u,i,j}(Y_{uij} - \bar{Y}_{ui})^2}{\sum_{u,i,j}(Y_{uij} - \bar{Y}_u)^2} = 4.038\% \quad [2.226\%, 6.552\%], \tag{5}$$

The complementary 95.962% is between sample parent means, which themselves contain child noise. Separately, a nested-ANOVA moment estimate gives $\widehat{\sigma}_w = 0.002400$, $\widehat{\sigma}_b = 0.011650$ and one-child ICC 0.959291; its within-component ratio is 4.071%, not exactly 4.038%. Both summaries condition on successful endpoints; failure-inclusive policy utility is a different estimand. The other frozen proxies have empirical shares 0.472%, 0.537%, and 13.449%. We use these values to describe the observed variance structure, not as a sufficient predictor of policy value; policy value is not monotone across the four metrics. These are raw-endpoint shares and are not invariant to the nonlinear survival normalization used by the held-out primary; we do not use Equation 5 as a numerical explanation or upper bound for that normalized effect.

The audit exposes a variance–information mismatch. Within a conditioning unit, differences between sample parent means account for most of the primary outcome variation, while specific fresh realizations lack candidate observations before purchase. Under the random-effects summary, one child is a reliable measurement of parent mean quality, not a sufficient statistic for its tail or future

improvement probability. This is consistent with the observed breadth-dominant regime, while still leaving genuine choices over the number of parents and which observed parent to expand. Consistently, the retrospective oracle's mean positive branch depth is only 1.34–1.53. We treat that depth as empirical scope evidence, not as a numerical prediction identified by Equation 5 alone.

Scope. The propositions cover pre-sampling policies with hidden seed/noise. Partial denoising observations reveal realization-specific information and fall outside this contract. An unexecuted peek protocol is documented in the appendix; it supplies no positive or negative evidence about interruptible policies.

5 Experimental Protocol and Development Evidence

5.1 Data, endpoints, and budgets

We use 200 license-eligible Google Scanned Objects conditioning images (Downs et al., 2022), partitioned at the conditioning-unit level into 80 Train, 40 calibration, and 80 test units. For each unit, the frozen schedule contains four stage-1 parents and four children per parent, for 3,200 main endpoints before replay controls. Test content remains sealed until the calibration gate. Failed endpoints are retained at utility zero and are never silently resampled.

The four lower-is-better proxies are input-silhouette Chamfer (primary), silhouette compactness, cross-view area coefficient of variation, and adjacent view area jump. A Train-frozen empirical survival transform maps a raw value y to

$$U(y) = \frac{1 + \#\{v > y\} + 0.5\#\{v = y\}}{n + 1}, \tag{6}$$

with failed endpoints assigned zero. This bounded utility was frozen as the confirmatory estimand; raw Chamfer is always reported beside it to expose the effect scale of this dimensionless 2D proxy. It measures the best match among eight rendered silhouettes, not 3D Chamfer distance to a ground-truth surface (Appendix B.1).

Development uses budgets $B0 = 55.0073$ s and $B1 = 78.0778$ s. Held-out replay uses predeclared Train q95 action-cost estimates: spawn costs 19.1849 s and expand 11.2277 s. Q and uniform receive the same nominal budget and admission rule. A post-hoc source audit finds that inherited manifest support is inactive at B0 but can filter actions using future costs at B1; the latter results are retained as support-restricted replay (Appendix B.4).

5.2 Policies and chronological attempt sequence

The representation contains 73 base features describing the input, observed parents and children, and action type, plus three horizon features, for **76 Q inputs**. Horizon features use remaining recorded budget and the frozen development median costs even when q95 controls action admission. Future endpoints and costs are excluded from the feature vector. Extremely randomized trees (Geurts et al., 2006) are fit in five conditioning-grouped folds. At held-out inference, the five fold forests for a metric are averaged; no model is refit after calibration.

We report every allocator family in the order attempted:

1. **One-step EI-per-cost** predicts the next-child utility distribution and greedily maximizes expected improvement over the incumbent per nominal action cost.
2. **Continuation-only correction** replaces only the spawn credit by the best feasible realized child prefix. It is a retrospective mechanism ablation, not a deployable policy.
3. **Finite-horizon Q** labels each state–action row by exact retrospective terminal utility under the remaining budget, then learns that label with the same grouped forest family. This clairvoyant continuation label, denoted Q_Ω, is not the value of an online optimal policy.
4. **Fixed-composition branch selection** removes spawn/expand choice from Q. A frozen uniform breadth controller supplies the action-type rule (spawn whenever admissible, otherwise expand); Q only ranks legal observed parents. Branch choices can still alter subsequent

recorded cost and therefore later feasibility. The mandatory zero-parameter comparator chooses the parent with highest current observed utility at the first legal expand and follows uniform thereafter. A second mandatory rule selects the parent with lowest observed stage-1-plus-first-child cost.

The first three are development attempts, not a multiple-model contest on test. Only item 4 was frozen for calibration and one-shot test confirmation.

The confirmatory baseline is equal-nominal-budget uniform breadth. A development "strongest" baseline is reported only descriptively: it was selected on Train and used realized future cost for feasibility at 62/621 checks, so it is neither an unbiased held-out opponent nor deployable under the online cost contract.

5.3 Why accurate Q prediction did not yield an unconstrained policy

Finite-horizon Q is learnable in pooled prediction space: Spearman correlation is 0.935–0.970 across eight metric–budget cells. Most target variation is at hierarchy levels shared within a decision; only 0.108–0.534% lies within an actual decision set. After removing fold, unit, and state means, residual Spearman is 0.102–0.209. On Q's visited non-degenerate sets it becomes -0.0875 (Kendall -0.0930), while 82.3% of visited multi-action sets have numerically degenerate retrospective Q_Ω labels. These ties do not imply that all online continuations have equal value. Enumeration-set macro Spearman is 0.3685; correlations are defined only on nondegenerate multi-action sets, as specified in Appendix B.5.

Table 1: Development-only chronology and diagnostic outcome. "Wins" count metric–budget cells against the stated comparator; these are not independent trials.

Attempt	Comparator	Outcome
One-step EI/cost	frozen strongest; score permutation	0/8 wins against each; selected score error is optimistic in 8/8
Continuation-only replacement	frozen strongest	improves the one-step policy in 8/8, flips 0/8
Finite-horizon Q	frozen strongest	0/8 wins; 3/8 intervals show harm; improves one-step in 8/8
Fixed composition + Q branch	fixed composition + uniform branch	branch contribution positive in 8/8

This dissociation is not simply "no signal." A 2×2 replay separates action-type composition from branch choice: Q's Shapley composition contribution is negative in 8/8 cells, whereas its Shapley branch contribution is positive in 8/8. Winner-error-excess point estimates are positive in 7/8 finite-Q cells and 8/8 one-step-score cells, consistent with selection-induced optimism (Smith & Winkler, 2006). Ranking is strongest at the first branch decision ($\rho = 0.5722$) and falls later ($\rho = 0.2060$). A first-Q-only policy retains 98.24% of the primary B0 development gain, and the current-utility rule retains 97.9%. These development findings motivated exactly one constrained held-out test; they are not counted as confirmatory evidence.

6 Held-out Confirmation: Value Without a Resolved Learned Increment

The model, features, five-forest ensemble, q95 admission rule, primary cell, failure handling and analysis were frozen before held-out access. A one-shot 40-unit calibration sign gate passed, authorizing the one-shot 80-unit test. No test-driven model or policy change was made.

Table 2: One-shot 80-unit test, precommitted primary cell. Values are paired differences in normalized utility with unit-cluster bootstrap 95% intervals. Only the first contrast is primary; rule contrasts are unadjusted secondaries.

Contrast	Mean [95% CI]
Q branch − matched uniform	$0.004281\ [0.000903, 0.008504]$
Current-utility rule − uniform	$0.003947\ [0.000943, 0.007551]$
Q branch − current-utility rule	$0.000334\ [-0.002534, 0.003977]$

Primary result. For input-silhouette Chamfer at budget B0, the frozen Q branch selector improves normalized utility over matched uniform breadth by 0.004281, with a unit-cluster bootstrap 95% confidence interval of $[0.000903, 0.008504]$. Both policies receive the same nominal $55.0073\,\mathrm{s}$ budget and the same q95 online cost model. Their credited completed-budget utilization is 85.55% and 85.82%, respectively; the paired utilization difference is not distinguishable from zero. Thus the comparison is neither realized-spend matching nor a larger budget for Q. There are two attempted overruns for Q and one for uniform; the frozen replay discards those endpoints and stops. Completed utilization does not include the discarded action's work or CPU selector overhead.

The corresponding raw Chamfer improvement is only 0.000199 (0.409% relative to matched uniform), with a 95% interval of $[-0.000005, 0.000433]$. Thus the effect is statistically resolved under the predeclared normalized-utility estimand, while the raw proxy difference is small on its numerical scale and its interval includes zero. We report these two statements together: normalized utility was the predeclared primary, but its confirmation does not imply a perceptually large effect.

A simple rule and an unresolved learned increment. A mandatory secondary policy uses only the currently observed utility at the first branch decision and follows the same frozen uniform composition thereafter. It improves normalized utility over matched uniform by $0.003947\ [0.000943, 0.007551]$. Q exceeds this rule by only $0.000334\ [-0.002534, 0.003977]$. The zero-parameter rule therefore captures 92.2% of the Q-over-uniform primary point estimate. This ratio is descriptive, not an equivalence bound. Hence the held-out result confirms positive B0 branch-selection value under replay, but does not resolve the incremental benefit of the 76-input Q function over the current-utility rule. There was no prespecified equivalence margin or test of whether that increment is practically negligible.

Across all eight metric–budget cells, Q beats matched uniform in seven point estimates and five confidence intervals; the current-utility rule does so in seven and six, respectively. Q has a positive point estimate over the rule in all eight cells, but only one interval excludes zero. These are descriptive counts under the recorded replay support, including the B1 limitation below. They are not eight independent signs: every cell evaluates the same 80 units (pairwise and global unit-set Jaccard $= 1$), and the analysis reuses one unit-cluster bootstrap index matrix across cells. An eight-trial sign test is therefore invalid. The concordant point estimates are compatible with a positive learned increment, but its size and practical relevance remain uncertain. The prespecified simple-rule secondary has a positive unadjusted B0 interval; this does not establish that it recovers the entire Q effect.

Online-cost scope of the confirmation. A post-hoc source audit reconstructs all 17,654 test manifest records from the saved cost ledger. The builder excludes actions whose recorded next-state cost exceeds B1 before applying the q95 filter. At B0 this removes none of 1,779 otherwise admissible actions across all states, so the primary is unaffected by this restriction. At B1 it removes 13 actions across seven units; six occur in states reachable by some branch sequence under the common controller. These are support counts, not measured changes in policy value. We retain B1 estimates as retrospective-support secondary results, not online deployment confirmations. No test policy was refit or rerun to repair them; Appendix B.4 gives the reproducible audit.

Failure-inclusive execution. Eight of 1,280 test main endpoints failed; nine replay controls are excluded from endpoint analyses. Failures retain zero utility with no retry or GPU migration. Audit chronology and reproduction limits are in the appendix.

7 Discussion and Limitations

A practical actionability audit. Before training an allocator, list the observations available before each action; quotient conditionally exchangeable sample labels; decompose outcome variation at the policy's hierarchy; and measure ranking inside actual decision sets, including the policy-induced state distribution. Pooled value accuracy is not a substitute. Cost admission, candidate support and endpoint credit must also respect the same information contract. These checks expose information restrictions and evaluation errors, but are not a universal test that rules out learning.

The TRELLIS hierarchy is consistent with a breadth-dominant regime, not proof that the bandit problem disappears. Spawn remains genuine exploration, and a high-value observed parent can merit expansion. The held-out B0 branch gain is positive in normalized utility under the replay contract. Raw silhouette-Chamfer improves by 0.000199, or 0.409%, with an interval including zero. This is a small measured gain, not a numerical consequence of the 4.038% raw sum-of-squares share or evidence of perceptually visible improvement.

What is and is not learned. The current-utility rule attains 92.2% of Q's primary gain point estimate. Q-minus-rule is unresolved, not proven zero. The eight cell point estimates are positive, but reuse the same 80 units and include B1 support-restricted results. The rule's own positive unadjusted secondary interval does not establish equivalence, noninferiority or that learning is unnecessary. The three failed unrestricted methods also have different failure modes; their comparisons are development evidence against selected heuristics, not an upper bound over all admissible policies.

Scope and cost limitations. The evidence covers one frozen generator, a four-by-four hierarchy, two budgets and 80 test units, using automatic 2D silhouette proxies without human evaluation. The recorded action budget omits CPU metric and selector overhead, imputes failed-child costs, and excludes a discarded overrun action from completed utilization. End-to-end deployment savings are unmeasured. B1 inherits future-cost candidate filtering and cannot support an online confirmation claim; the post-hoc audit finds no such filtering at B0. Neither the successful-child variance analysis nor the test intervals identify performance outside this hierarchy or across new model fits.

The propositions apply to systems satisfying their full observation, cost and label-invariance assumptions. The numerical variance share, ICC and effect sizes do not transfer automatically. ICC describes parent-mean reliability; conditional tails, costs, budget and covariates can still make adaptation useful at high ICC. No universal ICC threshold follows. Partial denoising observations can reveal realization-specific information; the unexecuted peek protocol supplies no evidence about that policy class. The development strongest heuristic also uses future realized cost at some feasibility checks and remains a descriptive comparator.

8 Conclusion

Adaptive test-time allocation depends on the location and timing of information, not only on outcome variation or pooled verifier accuracy. Conditional exchangeability identifies redundant child labels, and hidden fresh randomness prevents selecting a specific unseen parent realization. In TRELLIS, these information restrictions accompany a breadth-dominant hierarchy and a large value–decision dissociation. A common spawn-first controller exposes a small held-out B0 branch gain under the replay contract; a zero-parameter rule attains most of its point estimate, with an unresolved learned increment. The contribution is a diagnostic workflow and a bounded empirical case, rather than a theorem that adaptive allocation is unnecessary.

REPRODUCIBILITY STATEMENT

The appendix specifies policy definitions, feature dimensions, forest hyperparameters, normalization, endpoint metrics, cost and failure conventions, and all prespecified secondary rules. The anonymous supplement contains 16 frozen aggregate or row-level sources, their bindings, and a read-only verifier with 70 digest/numerical checks. It independently recomputes the ICC from 1,272 rows, while most other checks verify supplied aggregates rather than re-estimating effects. Additional source excerpts and a standard-library support audit expose the inherited manifest restriction in Appendix B.4. The package does not regenerate endpoints, refit policies, or independently reconstruct the held-out bootstrap intervals.

AI USE STATEMENT

Generative AI tools were used to brainstorm and refine hypotheses and conceptual framing; critique experimental and audit protocols; assist with mathematical and proof drafting; scaffold analysis, verification, and packaging code; support literature search and summarization; interpret frozen results; and edit or translate manuscript text; and create or revise figures and release artifacts. They were not used as evaluators or as sources of experimental labels. The authors checked all AI-assisted mathematical statements, citations, code paths, numerical claims, and prose against primary papers, tests, or frozen project artifacts, and take responsibility for the complete content of the submission.

REFERENCES

Mohsen Bayati, Nima Hamidi, Ramesh Johari, and Khashayar Khosravi. Unreasonable effectiveness of greedy algorithms in multi-armed bandit with many arms. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

Ahsan Bilal, Ahmed Mohsin, Muhammad Umer, Ali Subhan, Hassan Rizwan, Ayesha Mohsin, and Dean Hougen. What if we allocate test-time compute adaptively? *arXiv preprint arXiv:2602.01070*, 2026.

Hao Mark Chen, Guanxi Lu, Yasuyuki Okoshi, Zhiwen Mo, Masato Motomura, and Hongxiang Fan. Rethinking optimal verification granularity for compute-efficient test-time scaling. In *Advances in Neural Information Processing Systems*, 2025.

Adrien Couëtoux, Jean-Baptiste Hoock, Nataliya Sokolovska, Olivier Teytaud, and Nicolas Bonnard. Continuous upper confidence trees. In *Learning and Intelligent Optimization*, pp. 433–445, 2011. doi: 10.1007/978-3-642-25566-3_32.

Florian E. Dorner, Yatong Chen, André F. Cruz, and Fanny Yang. ROC-n-reroll: How verifier imperfection affects test-time scaling. In *International Conference on Learning Representations*, 2026.

Laura Downs, Anthony Francis, Nate Koenig, Brandon Kinman, Ryan Hickman, Krista Reymann, Thomas B. McHugh, and Vincent Vanhoucke. Google scanned objects: A high-quality dataset of 3d scanned household items. In *IEEE International Conference on Robotics and Automation*, 2022.

Pierre Geurts, Damien Ernst, and Louis Wehenkel. Extremely randomized trees. *Machine Learning*, 63:3–42, 2006.

Nanye Ma, Shangyuan Tong, Haolin Jia, Hexiang Hu, Yu-Chuan Su, Mingda Zhang, Xuan Yang, Yandong Li, Tommi Jaakkola, Xuhui Jia, and Saining Xie. Scaling inference time compute for diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2523–2534, 2025.

Martijn R. K. Mes, Warren B. Powell, and Peter I. Frazier. Hierarchical knowledge gradient for sequential sampling. *Journal of Machine Learning Research*, 12(90):2931–2974, 2011.

Gaurav Parmar, Or Patashnik, Daniil Ostashev, Kuan-Chieh Wang, Kfir Aberman, Srinivasa Narasimhan, and Jun-Yan Zhu. Scaling group inference for diverse and high-quality generation. In *International Conference on Learning Representations*, 2026.

Shuhui Qu. Adaptive test-time compute allocation via learned heuristics over categorical structure. *arXiv preprint arXiv:2602.03975*, 2026.

Vinod Raman, Hilal Asi, and Satyen Kale. AdaBoN: Adaptive best-of-N alignment. *arXiv preprint arXiv:2505.12050*, 2025.

Ruchit Rawal, Reza Shirkavand, Sayak Paul, Yuxin Wen, Heng Huang, Yizheng Chen, Tom Goldstein, and Gowthami Somepalli. Flash-BoN: Instant drafts for inference-time scaling in diffusion models. *arXiv preprint arXiv:2607.04461*, 2026.

James E. Smith and Robert L. Winkler. The optimizer's curse: Skepticism and postdecision surprise in decision analysis. *Management Science*, 52(3):311–322, 2006. doi: 10.1287/mnsc.1050.0451.

Charlie Snell, Jaehoon Lee, Kelvin Xu, and Aviral Kumar. Scaling LLM test-time compute optimally can be more effective than scaling model parameters. In *International Conference on Learning Representations*, 2025.

Hado van Hasselt. Double Q-learning. In *Advances in Neural Information Processing Systems*, volume 23, pp. 2613–2621, 2010.

Yizao Wang, Jean-Yves Audibert, and Rémi Munos. Algorithms for infinitely many-armed bandits. In *Advances in Neural Information Processing Systems*, volume 21, 2008.

Jianfeng Xiang, Zelong Lv, Sicheng Xu, Yu Deng, Ruicheng Wang, Bowen Zhang, Dong Chen, Xin Tong, and Jiaolong Yang. Structured 3d latents for scalable and versatile 3d generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2025.

A Additional Protocol and Results

A.1 Complete held-out secondary table

All values below are normalized-utility differences on the same 80 test units. The intervals are marginal 95% unit-cluster bootstrap intervals and are not multiple-comparison adjusted. The common unit set and common bootstrap matrix make a cell-level sign test invalid.

Table 3: All held-out metric–budget cells. Rule denotes current-utility first-expand selection. B1 rows (†) have hindsight-restricted candidate support; see Appendix B.4.

Cell	Q − uniform	Rule − uniform	Q − rule
Chamfer B0	.004281 [.000903, .008504]	.003947 [.000943, .007551]	.000334 [−.002534, .003977]
Chamfer B1†	.004281 [.000697, .009191]	.003280 [.000599, .007699]	.001002 [−.000815, .003064]
Adjacent jump B0	.000314 [−.000079, .000796]	.000147 [−.000206, .000471]	.000167 [−.000098, .000530]
Adjacent jump B1†	.002347 [.000697, .004606]	.001414 [.000147, .003476]	.000933 [.000010, .001964]
Area CV B0	.000727 [−.000059, .001659]	.000530 [.000088, .001159]	.000196 [−.000393, .000943]
Area CV B1†	.004144 [.000697, .009221]	.001502 [.000187, .003888]	.002641 [−.000040, .007041]
Compactness B0	−.001650 [−.004949, .000678]	−.001856 [−.005126, .000364]	.000206 [−.000668, .001198]
Compactness B1†	.004615 [.000726, .008975]	.003083 [.000206, .006392]	.001532 [−.000550, .003928]

A.2 Budget use and raw effect scale

At primary B0 both arms receive 55.007286612 s. Q and uniform credit completed costs of 85.5475% and 85.8222% of nominal budget. Their paired utilization difference is -0.2747 percentage points with 95% CI $[-0.9154, 0.1864]$; mean slack is 7.9499 s and 7.7988 s. Seventy-nine of 80 units complete the same number of actions. Q has two attempted recorded-cost overruns and uniform one. These percentages exclude the cost of the discarded final attempt and uninstrumented overhead; they are not total wall-clock utilization or a no-overshoot guarantee. Raw terminal Chamfer is 0.0485167 for Q and 0.0487159 for uniform, a paired improvement of 0.0001992 $[-0.0000055, 0.0004332]$.

A.3 The second prespecified inexpensive rule

The observed-cost rule chooses, at the first legal expand, the parent with the smallest *already observed* stage-1 plus first-child cost, then returns to uniform branching. It was prespecified alongside the current-utility rule; both are reported, irrespective of outcome. Its primary effect is less conclusive than that of the current-utility rule. No secondary interval is multiplicity-adjusted.

Table 4: Complete observed-cost-rule secondary results. B1† is support-restricted descriptive replay, as in Table 3.

Cell	Cost rule − uniform	Q − cost rule
Silhouette Chamfer B0	0.002857 [−0.000334, 0.006982]	0.001424 [−0.000108, 0.003457]
Silhouette Chamfer B1†	0.000295 [−0.000020, 0.000746]	0.003987 [0.000452, 0.008877]
Adjacent jump B0	−0.000020 [−0.000422, 0.000285]	0.000334 [−0.000049, 0.000825]
Adjacent jump B1†	0.000962 [−0.000353, 0.002985]	0.001385 [0.000491, 0.002425]
Area CV B0	0.000432 [−0.000029, 0.001227]	0.000295 [−0.000815, 0.001404]
Area CV B1†	0.001002 [−0.000363, 0.003407]	0.003142 [0.000451, 0.007591]
Compactness B0	−0.000854 [−0.002327, 0.000256]	−0.000795 [−0.003741, 0.001247]
Compactness B1†	0.000579 [−0.001787, 0.003584]	0.004036 [0.001434, 0.006815]

A.4 Attempt ledger and evidence levels

The method sequence was fixed by development chronology, not selected after the held-out result: one-step EI-per-cost; post-result continuation-only mechanism ablation; finite-horizon Q; exploratory composition–branch factorial replay; then one frozen constrained-branch confirmation. The first three read Train only. Calibration was read once for a precommitted sign gate; test was read once after the gate. The simple current-utility and observed-cost rules were mandatory secondaries in the frozen confirmation protocol.

A.5 Failure and audit semantics

The test schedule contains 1,280 main endpoints and nine replay controls. Eight main rows from two complete structures failed and remained failures; no row was retried or migrated. Collection and numeric-order confirmation auditors passed 9/9 and 10/10 checks. A retained older auditor is 21/22 because it sorts string unit identifiers lexicographically instead of using numeric manifest indices for the frozen bootstrap matrix. The corrected auditor was frozen before test execution and reproduces the runner; the failed artifact remains in the release.

B Implementation and Statistical Contract

B.1 Endpoint metrics and normalization

TRELLIS-image-large generates four structure seeds and four child seeds per structure for each input unit. Stage 2 uses the frozen 25-step, CFG 5.0 sampler. Evaluation uses eight fixed 256×256 RGB views. These metrics are silhouette proxies, not comparisons to a reference 3D mesh, learned human preferences, or verified unseen-surface accuracy.

The generated foreground mask has maximum channel deviation greater than 2 from background RGB $(128, 128, 128)$. The reference mask uses deviation greater than 12 from its median border color. A valid reference has at least 64 foreground pixels. Boundary pixels are obtained by subtracting an eight-neighbor erosion; at most 2,048 are retained by deterministic evenly spaced indices in row-major boundary order. Coordinates are centered at the mean of all foreground pixels and divided by the square root of foreground area. For boundary sets A, B, the distance is

$$d(A,B) = \tfrac{1}{2}\left(|A|^{-1}\sum_{a\in A}\min_{b\in B}\|a-b\|_2 + |B|^{-1}\sum_{b\in B}\min_{a\in A}\|b-a\|_2\right).$$

The primary raw metric is the minimum over eight generated views against the input boundary. It is dimensionless, translation/area-normalized, and best-view-selected; it is not a physical-unit 3D Chamfer distance.

The three secondary raw metrics are: median view compactness $p_v^2/(4\pi a_v)$ (four-edge pixel perimeter p_v, area a_v); area coefficient of variation $\mathrm{std}_v(a_v)/\bar{a}$ with population standard deviation; and maximum cyclic adjacent-view area jump $\max_v |a_{v+1} - a_v|/\bar{a}$. All are lower-is-better under the frozen proxy convention, not universal shape-quality criteria: for example, compactness can favor a rounder silhouette even when the object is not round.

For each metric, successful training values $z_1, \ldots, z_n$ define

$$u(y) = \frac{1 + \sum_{\ell=1}^{n} \mathbf{1}\{z_\ell > y\} + \frac{1}{2}\sum_{\ell=1}^{n} \mathbf{1}\{z_\ell = y\}}{n+1}.$$

A failed endpoint has utility zero and missing raw utility. For held-out evaluation the reference distribution uses all 80 Train units. Each fold-specific predictor uses its own 64-unit training reference distribution, excluding that fold's held-out units. These transforms and the evaluation transform are not refit using calibration or test outcomes.

B.2 Frozen Q model and features

There are 20 forests: four metric models times five unit-held-out folds. Each is an ExtraTreesRegressor (Geurts et al., 2006) with 512 trees, squared-error criterion, maximum depth 8, minimum leaf size 16, `max_features=sqrt`, and `bootstrap=False`. The seed is 2026073013 plus the zero-based metric index. Training uses scikit-learn 1.4.2, four CPU threads, and per-row weights reciprocal to the number of training action–budget rows for that unit. Each held-out prediction averages the five fold models; no new full-Train model is selected. The confirmation adapter reconstructs the prescribed models from Train and checks their tree digests against the saved development models before predicting held-out queries.

The Q input has **76 dimensions**: the original 73-dimensional state/action representation plus three horizon features. The 73-dimensional base contains seven global state statistics (opened parents,

observed children, incumbent maximum, mean, standard deviation, failure fraction, spawn flag); six candidate-status/utility statistics (missing flag, count, maximum, mean, standard deviation, failure fraction); 14 candidate morphology values; eight candidate morphology ranks among opened parents; mean and standard deviation of all 14 morphology values across opened parents (28); and ten reference image descriptors. The 27 entries in base indices 8–34 (zero-based) are exactly zero for spawn in both fitting and inference; index 7 is the missing flag and equals one.

Morphology comprises occupancy count, bounding-box fill, three axis extents, component count, largest voxel-component fraction, mean neighbor degree, fractions with degree at most one and two, boundary-voxel fraction, and three covariance eigenvalue fractions. The eight ranks use occupancy, fill, largest component and mean degree in descending order, and component count, the two low-degree fractions and boundary fraction in ascending order. Reference descriptors comprise foreground fraction, compactness, bounding-box width and height, log aspect ratio, two centroid offsets, component count, largest component fraction, and covariance eigenvalue ratio.

Let $R = \max(0, B - C)$, where C is the recorded cost of completed actions. The three horizon features are

$$\left(R/c_e,\ \max(0, R - c_a)/c_e,\ \left\lfloor \{\max(0, R - c_a) + 10^{-12}\}/c_e \right\rfloor\right).$$

Here $c_e = 5.767616307$ and c_a is 13.751821653 for spawn or 5.767616307 for expand. These *median* feature constants remain frozen even though the final admission rule uses q95. They were not silently replaced during confirmation. Targets are finite-table retrospective continuation maxima, denoted Q_Ω, rather than expected online-optimal action values.

B.3 Controller, tie-breaking and budget credit

The state is a four-vector of observed child counts, with parents opened in fixed seed order and at most four children per parent. The compulsory initial state $(1, 0, 0, 0)$ is charged its recorded cost and contributes its observed utility, but is not treated as a learned decision. All confirmation policies use the same controller: spawn the next parent whenever an admitted spawn exists; otherwise choose an admitted expand. This matches a *statewise controller rule*, not an identical realized action-type schedule, because branch choices affect subsequent costs and states.

Uniform expansion orders candidates by child index then parent index. Constrained Q selects the highest predicted-value expand using the frozen deterministic tie rule: scores within 10^{-15} of the maximum are tied, then ordered by parent index and child index. The current-utility rule selects the parent with the highest already observed child utility at the first legal expand, then uses uniform expansion. The observed-cost rule instead minimizes its observed stage-1 plus first-child cost at that expand. The first expand consumes this intervention even if only one parent is eligible; it is not necessarily the first multi-parent choice.

Budgets are $B_0 = 55.007286612$ and $B_1 = 78.077751840$ seconds. The frozen q95 admission estimates are 19.1848713843 seconds for spawn and 11.227739191750002 for expand. Given a supported candidate, admission tests $C + \widehat{c}_a \le B + 10^{-12}$. If its subsequently revealed recorded next-state cost exceeds $B + 10^{-12}$, replay stops and retains the *pre-action* state. Neither its endpoint nor its cost contributes to the reported completed state. Thus an attempted overrun is observable, and completed utilization can coexist with actual work beyond the budget.

Instrumented child cost includes stage-2 sampling, Gaussian decoding and RGB rendering; spawn additionally includes stage-1 sampling. It does not measure input encoding, model loading, serialization, CPU metric/feature extraction or Q prediction overhead. Failed child rows use the frozen q95 child cost as an imputation, not a measured failure duration. These conventions define a finite-table replay estimand, not a strict end-to-end deployment contract.

B.4 Inherited candidate-support restriction

A post-submission-package source audit found that the query builder retains the legacy manifest enumeration: a candidate is absent if its *recorded future* next-state cost exceeds the largest budget B_1. The q95 admission rule operates on this already-pruned support. This is hindsight information even though the later q95 comparison is online.

A read-only cost-ledger audit reconstructs all 17,654 manifest records without using utility outcomes, fitting a model, or re-estimating any held-out effect. Among every q95-admissible state–action pair in the finite prefix table, no B0 action is removed (0/1,779, zero units). At B1, 13/12,077 are removed, affecting seven units. Restricting to states reachable under the common spawn-first controller, allowing every branch choice, the corresponding counts are 0/348 at B0 and 6/2,829 at B1, the latter across five units. These count *possible support removals*, not the realized Q-policy selection frequency; aggregate released results do not identify their policy-value effect.

The primary B0 comparison is unaffected by this particular support filter. B1 results are preserved numerically but are explicitly *hindsight-support restricted descriptive secondaries*, not confirmation of a fully online policy. We do not repair the harness and repeat test. This restriction is distinct from an earlier development audit of a retrospective strongest-heuristic comparator (62/621 feasibility checks, 61/80 units); the latter is an internal protocol limitation, not evidence that external papers use the same implementation.

B.5 Inference, variance estimators and ranking denominators

The held-out primary is Q minus uniform on silhouette-Chamfer normalized utility at B0. Calibration is a one-shot primary point-estimate sign gate; test is a one-shot evaluation after passage. The reported intervals resample 80 independent test units with replacement, using 5,000 replicates, NumPy `default_rng(2026073014)`, numeric unit-index ordering and linear 2.5th/97.5th percentile quantiles. The same resampling matrix is used across all cells and policy contrasts. These are conditional-on-fixed-policy intervals; they do not incorporate the variability of selecting a method on Train. Only the prespecified primary has confirmatory status, additionally subject to the replay scope stated above. The other seven cells and both simple-rule comparisons have unadjusted marginal intervals. The frozen secondary predictions expected both compactness intervals to include zero: this holds at B0, but not for the recorded B1 result, which also has the support restriction above. The configuration left a named LOUO-baseline comparison as binding-pending; it was not executed as a held-out comparison and is not claimed here. No equivalence or noninferiority margin was registered, and no independent eight-cell sign test is valid.

The empirical within-parent share uses sums of squares centered at observed parent and unit means. It includes 1,272 successful Train rows, 318 complete parents and 80 units. For the separate nested-ANOVA summary, $MS_w = SS_w/(1272 - 318)$ and $MS_b = SS_b/(318 - 80)$, $\widehat{\sigma}^2_w = MS_w$ and $\widehat{\sigma}^2_b = \max(0, (MS_b - MS_w)/4)$. The ICC is $\widehat{\sigma}^2_b/(\widehat{\sigma}^2_b + \widehat{\sigma}^2_w) = 0.959291$. This is a moment summary conditional on observed success, not proof of homoscedastic Gaussian random effects. Empirical sums-of-squares shares and variance-component shares are not interchangeable.

Conditional rank summaries are exploratory decision-set macro averages, not averages over all action rows or independent samples of 80 units per cell. For enumerated sets there are 31,520 sets, 29,068 multi-action sets and 6,639 nondegenerate label sets. For Q-visited states there are 3,408 sets, 2,846 multi-action sets and 504 nondegenerate label sets; 2,342/2,846 (82.3%) have label ranges at or below the frozen numerical tolerance. Undefined rank correlations are excluded, not assigned zero. Macro Spearman/Kendall are 0.3685/0.3472 for enumeration and $-0.0875/-0.0930$ for Q-visited states. These are different state distributions, not an isolated causal effect of an argmax.

Of the 954 aligned constrained-branch sets, 496 are first-position sets and 458 later-position sets. Only 68 and 35, respectively, have nondegenerate retrospective labels. Their Spearman means are 0.5722 and 0.2060; the large denominator of tied sets must not be mistaken for 954 informative rankings. The original cross-parent action-pair share is identically one. A post-hoc replacement based on observed evidence has highest-bin Jaccard overlap 1.0 with the first-branch set; it is the same exploratory evidence, not a second dose-response validation.

Train first-Q-only retention (98.24%), the simple-rule retention (97.9%), and the raw effect scale (0.3837% relative; 13.11% of the mean within-parent candidate SD and 2.42% of the mean within-unit candidate SD) describe development outcomes. The latter mean-of-SD denominators differ from the ANOVA component SDs above. Held-out retention is instead 92.2% and the relative raw effect is 0.4089%; development and test percentages must not be interchanged. The retrospective mean branch-depth range 1.34–1.53 uses its original realized-cost oracle contract, not the q95 confirmation controller, and is not a theoretical depth prediction.

B.6 UNEXECUTED PROTOCOLS AND RELEASE LIMITS

An interruptible peek protocol was frozen (SHA-256 prefix `7dcbd995`) for 64 units outside the sealed split, eight children per parent and steps 5/10/15. Its intended primary was within-parent Spearman for an online decoded/rendered silhouette score, with an early-step threshold of 0.15; nominal cost was $k/25$ sampler cost plus decoding, rendering and Chamfer evaluation. This was a planned cost approximation, not a measurement. Infrastructure and binding checks did not all pass (72/73); the arm was closed without scientific collection. The protocol and its hash supply neither null evidence nor evidence that peeking is ineffective. SDXL and CFG/steps scans likewise contribute no scientific results.

The supplement verifies released evidence rather than regenerating the study. It contains aggregate held-out contrasts, collection/cost records, successful Train metric rows, selected source excerpts and protocol projections. It does not contain all prediction rows, fitted forests, input images, generated assets or per-unit policy-outcome bootstrap inputs. Consequently, the 70 legacy checks are chiefly hash and aggregate checks; only the stated row-level quantities (including the ICC and the added support audit) are independently recomputed. Passing these checks is not full experimental reproduction.

C PROOF DETAILS

C.1 EXCHANGEABLE-SLOT QUOTIENT

Let S be the set of unobserved child labels for parent P, and let $\tau : S \to S$ be any permutation. By joint conditional exchangeability, applying τ to all future observable outcomes, failure indicators and costs leaves their regular conditional distribution given $(P, \mathcal{H}_t)$ invariant. A policy that first requests label j and a policy that first requests $\tau(j)$ can be coupled by permuting all subsequent references to labels with τ. The observed utility multiset and cost path are identical in law. Label-invariant side observations, feasibility and stopping ensure that relabelled actions remain admissible at every step. Since Equation 2 is symmetric in the multiset, expected terminal utility is identical. The quotient contains one expansion action per non-exhausted parent.

C.2 UNSEEN-PARENT NON-RANKING

Let R denote all fresh randomness used to draw a parent and assume $R \perp\!\!\!\perp \mathcal{H}_t$ conditional on input x. Any admissible pre-spawn decision A is measurable with respect to $\sigma(\mathcal{H}_t, X_{\mathrm{sp}})$. Because $X_{\mathrm{sp}} = x_0$ almost surely, A is conditionally independent of R. For any measurable quality set C,

$$\Pr\{\mu(G_1(x, R)) \in C \mid A, \mathcal{H}_t\} = \Pr\{\mu(G_1(x, R)) \in C \mid \mathcal{H}_t\}.$$

Spawn/expand may still depend on x, history, budget, and cost; only the identity of a not-yet-drawn realization is unrankable.