SHARP ACCURACY THRESHOLDS FOR PARALLEL MASKED SAMPLING OF NOISY PAIRS

Anonymous authors
Paper under double-blind review

ABSTRACT

How accurately must conditional heads be learned to preserve two-round generation of independent noisy pairs? We study masked samplers that select a batch, draw independently from the original heads, and permanently commit the values. For n pairs, the critical total population excess-loss budget at fixed depth $K \geq 2$ is $\Theta_{K,\varepsilon}(n^{-1/[2(K-1)]})$ for total variation (TV) accuracy $\varepsilon \in (0,1)$. The heads are coherent: they are conditionals of one model, and the budget equals joint KL. Each pair has a fair representative and a mismatch bit; the model allows every positive exchangeable mismatch law independent of the representatives. A deterministic geometric schedule attains the threshold uniformly. A single model gives the converse against fully informed, randomized, value-adaptive policies. With logarithmically vanishing KL and uniformly vanishing head error, some models require $\Omega(\log n/\log\log n)$ expected rounds; a growing-depth upper matches this order for budgets $(\log n)^{-\zeta}$ with $\zeta > 11/3$. The lower obstruction also survives heterogeneous representative biases. Within the fair exchangeable class, target-aware selection of sampled candidates instead yields vanishing TV in three rounds whenever the KL budget vanishes.

1 INTRODUCTION

Learning-to-sampling guarantees relate prediction error to generated-sample error (Chen et al., 2023). Discrete diffusion uses finite-state corruption kernels (Austin et al., 2021) and can learn reverse dynamics through probability ratios (Lou et al., 2024); masked formulations use conditional-prediction losses (Shi et al., 2024; Sahoo et al., 2024). Their sampling error also depends on parallelism: conditional marginals specify how to predict missing coordinates, but the sampler chooses which coordinates to generate together. We study the accuracy needed for a prescribed number of sequential rounds.

We assume every head is an exact conditional of a common joint law R. Such a model is *coherent*: serial conditional sampling generates R, and small $\mathrm{KL}(P\|R)$ guarantees proximity to the target P. A parallel round replaces a joint conditional by a product of marginals, discarding dependence within the batch. Its generated values then determine the predictions available in later rounds. We ask:

> How accurately must a conditional model be learned to preserve shallow parallel generation using its original heads?

Recent results control prediction and factorization errors and relate schedules to dependence (Lavenant & Zanella, 2025; Chen et al., 2026; Zhao & Cai, 2026). Adaptive ordering can also avoid difficult prediction tasks (Kim et al., 2025). We seek an accuracy budget guaranteeing short execution for *every* model near an easy target. The planner knows P and R, has unlimited offline computation, and may randomize: the lower bound must hold even for its best schedule.

1.1 OUR RESULTS

Our target consists of n independent noisy pairs. Its exact heads generate it in two rounds: reveal one endpoint of every pair, then the other. Nevertheless, coherent models approaching it in joint KL and in every conditional head uniformly can require $\Omega(\log n/\log\log n)$ rounds *in expectation* to

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.

attain any fixed TV error below one. Each round draws independently from the original heads and commits its values; batch choices may adapt to the full history.

We determine the critical accuracy at each fixed depth. Each pair has a fair first endpoint, called its representative, and a bit recording whether its endpoints disagree. Allow any positive joint law of these mismatch bits that is invariant under pair relabeling and independent of the representatives. Let B bound the total population conditional excess loss, equal to $\mathrm{KL}(P\|R)$ under the normalization in Section 2.

Main result (informal; Theorem 2). For fixed $K \geq 2$ and TV accuracy $\varepsilon \in (0,1)$, the largest budget that guarantees an accurate K-round execution for every model in this class satisfies

$$B^*_{n,K}(\varepsilon) = \Theta_{K,\varepsilon}\left(n^{-1/[2(K-1)]}\right). \tag{1}$$

One deterministic geometric schedule attains this order uniformly over the class. Conversely, a single model defeats every allowed policy above the critical scale: if $n^{1/[2(K-1)]}B_n \to \infty$, the minimax TV error tends to one. If the same scaled budget tends to zero, the error tends to zero. Thus the threshold separates uniform success from worst-case failure, with a sharp dimension exponent for every fixed K.

Robustness to informative observations. The near-logarithmic lower bound persists for every positive representative-bias array, with uniform constants even as biases approach zero or one. Finite-alphabet extensions include nonuniform, heterogeneous examples. The two-round target isolates robustness to model error from intrinsic sampling difficulty; Section 8 collects the scope of these guarantees.

2 Model: conditional accuracy and execution depth

All logarithms are natural. Write $[n] = \{1,\dots,n\}$ and $\mathrm{TV}(P,Q) = \frac{1}{2}\sum_x |P(x) - Q(x)|$; KL denotes Kullback–Leibler divergence. A partial context consists of a visible set and its committed values.

An easy target and a coherent oracle class. Fix $\tau = 1/4$ and $v = \tau(1-\tau) = 3/16$. Under P_n, the variables $A_i \sim \mathrm{Ber}(1/2)$ and $E_i \sim \mathrm{Ber}(\tau)$ are mutually independent, and

$$X = (A_1, B_1, \dots, A_n, B_n), \qquad B_i = A_i \oplus E_i. \tag{2}$$

The pairing is known. A model R_n has the same independent fair A variables, independent of a positive exchangeable law R_E on $\{0,1\}^n$. Exchangeability means invariance under mismatch-coordinate permutations. Write $q_{R,i}(\cdot \mid x_S)$ for its original conditional head; positivity makes every such head well-defined.

Terminology. We also call the target P_n the *teacher*. A pair is *fresh*, *open*, or *completed* when zero, one, or both endpoints are committed. Revealing its first endpoint alone is a *birth*; revealing both fresh endpoints in one batch is a *co-sample*. A *pilot* is an initial collection of completed pairs used to inform later predictions. A teacher prefix is drawn from P_n; a generated prefix uses the model heads. Table 1 collects the notation.

The population loss. Let $N = 2n$. For each visible set $S \subset [N]$ and $i \notin S$, define

$$\begin{aligned} e_i(S, x_S) &= \mathrm{KL}\big(P_{n,i}(\cdot \mid x_S)\|q_{R,i}(\cdot \mid x_S)\big), \\ \eta_{\mathrm{train}}(q_R) &= \frac{1}{N}\sum_{k=0}^{N-1} \mathbb{E}_{S\sim\mathrm{Unif}\binom{[N]}{k}} \frac{1}{N-k}\sum_{i\notin S} \mathbb{E}_{X\sim P_n} e_i(S, X_S). \end{aligned} \tag{3}$$

The loss averages over a uniform visible-set size, a uniform set of that size, and a uniform missing coordinate. Every proper partial target context has positive weight. The chain rule along each full singleton order gives

$$D(q_R) := N\eta_{\mathrm{train}}(q_R) = \mathrm{KL}(P_n\|R_n) = \mathrm{KL}(P_E\|R_E). \tag{4}$$

The budget B bounds the *total* loss $D = N\eta_{\text{train}}$. For time-independent heads, D is the population excess of the continuous-time objective with independent mask probability t and weight $1/t$ (Shi et al., 2024); Appendix A proves the identity.

Definition 1 (Irreversible independent-head policy)**.** Starting with no revealed coordinates, at each round a policy chooses a subset of unrevealed coordinates using the complete committed history and private randomness. Conditional on this history and the selected action, the batch has the product law of the original heads q_R at the pre-round context, with fresh sampling randomness. The values are committed and never revised. Every path must reveal all coordinates in at most K rounds. Let Π_K be this class and Q_R^π its full-output law, with private randomness marginalized.

The policy knows both laws and the pairing, with unlimited offline computation. It may choose either orientation, co-sample fresh pairs, and adapt to labels and observed values. Actions precede their batch outcomes; all selected values are drawn at the actual committed context. The resource is the number of physical reveal rounds.

An algorithmic example. Ranking missing positions by current head entropy, choosing a batch, and then drawing its values independently fits this interface. EB-Sampler uses this select-before-sample structure (Ben-Hamu et al., 2025, Fig. 4); the correspondence applies when it samples coherent, unmodified heads. Model error is measured by $\mathrm{KL}(P_n \| R_n)$ and delivered-sample error by $\mathrm{TV}(P_n, Q_R^\pi)$.

The minimax quantity. For $\mathcal{C}_n(B) = \{R_E > 0 : R_E \text{ exchangeable}, \ \mathrm{KL}(P_E \| R_E) \le B\}$, set

$$\mathcal{E}_n(K, B) = \sup_{R_E \in \mathcal{C}_n(B)} \inf_{\pi \in \Pi_K} \mathrm{TV}(P_n, Q_R^\pi), \quad B^*_{n,K}(\varepsilon) = \sup\{B \ge 0 : \mathcal{E}_n(K, B) \le \varepsilon\}. \tag{5}$$

The supremum defining B^* is taken in $[0, +\infty]$. Each oracle has its own fully informed planner. TV concerns the delivered vector, with seed and reveal order marginalized.

3 Sharp budgets for a fixed number of rounds

3.1 Warm-up: a rare component becomes a common drift

Take $0 < w \le 1/8$ and $0 < \Delta \le 1/4$, and consider just one alternative,

$$R_E^{\text{one}} = (1 - w)\,\mathrm{Ber}(\tau)^{\otimes n} + w\,\mathrm{Ber}(\tau + \Delta)^{\otimes n}. \tag{6}$$

Its joint discrepancy is small: $\mathrm{KL}(P_E \| R_E^{\text{one}}) \le -\log(1 - w)$ and $\mathrm{TV}(P_E, R_E^{\text{one}}) \le w$. Now run the natural two-round schedule π_{AB}: reveal every A_i, then every B_i. The first round reveals no information about the mixture component. Every second-round mismatch head consequently has mean $\tau + w\Delta$. Since the execution draws these heads independently,

$$\mathcal{L}_{Q^{\pi_{AB}}_{R^{\text{one}}}}(E) = \mathrm{Ber}(\tau + w\Delta)^{\otimes n}, \qquad \mathrm{TV}(P_n, Q^{\pi_{AB}}_{R^{\text{one}}}) \ge 1 - 2e^{-n(w\Delta)^2/2}. \tag{7}$$

Hoeffding's inequality at count threshold $n\tau + nw\Delta/2$ gives the bound. The parallel batch spreads a rare model perturbation into a common drift, detectable above the $n^{-1/2}$ fluctuation scale. We next extend this example from the two-round schedule to all permitted policies.

3.2 The minimax threshold and its witnesses

Theorem 2 (Fixed-round critical exponent)**.** *For every fixed integer $K \ge 2$ and $\varepsilon \in (0, 1)$, there are positive constants $c_{K,\varepsilon}, C_{K,\varepsilon}$ such that, for all sufficiently large n,*

$$c_{K,\varepsilon} n^{-1/[2(K-1)]} \le B^*_{n,K}(\varepsilon) \le C_{K,\varepsilon} n^{-1/[2(K-1)]}. \tag{8}$$

For any nonnegative budget sequence B_n,

$$n^{1/[2(K-1)]} B_n \longrightarrow 0 \implies \mathcal{E}_n(K, B_n) \longrightarrow 0, \tag{9}$$

$$n^{1/[2(K-1)]} B_n \longrightarrow \infty \implies \mathcal{E}_n(K, B_n) \longrightarrow 1. \tag{10}$$

The upper guarantee is attained by the same deterministic schedule for every oracle in the budget class. The lower guarantee is attained by a single oracle sequence against every allowed policy.

The dimension exponent is sharp for each fixed K and ε. Section 8 discusses constants and growing-depth limits.

A universal schedule. Let $r = K - 1$. Reveal all A_i in the first round. Then reveal the B_i in deterministic batches whose cumulative sizes are

$$c_0 = 0, \qquad c_j = \lfloor n^{j/r} \rfloor \ (1 \le j < r), \qquad c_r = n. \tag{11}$$

Batch j has size $d_j = c_j - c_{j-1}$ and uses the original heads at its generated prefix. The schedule requires only n and K. Table 2 lists the scales; the critical averaged population loss is $B^*/(2n) = \Theta_{K,\varepsilon}(n^{-1-1/[2(K-1)]})$.

A single model that defeats all schedules. The lower construction keeps almost all mass on the target and spreads the remaining mass across a decreasing sequence of mismatch rates. Let $0 < w \le 1/8$, $0 < \lambda \le 1$, $0 < \Delta \le 1/4$, and define

$$b = \frac{w}{r}, \quad \delta_j = \Delta(b\lambda)^j, \quad R_E = (1-w)\,\mathrm{Ber}(\tau)^{\otimes n} + b\sum_{j=0}^{r-1} \mathrm{Ber}(\tau+\delta_j)^{\otimes n}. \tag{12}$$

Components are indexed by $j = 0, \dots, r-1$; δ_r is the terminal drift level used to distinguish the output from the target.

Theorem 3 (Finite all-policy lower bound)**.** *For integers $n, r \ge 1$ and the oracle in (12),*

$$\inf_{\pi \in \Pi_{r+1}} \mathrm{TV}(P_n, Q_R^\pi) \ge [1 - r\lambda]_+ - 2\exp\left[-\frac{n\Delta^2}{2}\left(\frac{w\lambda}{r}\right)^{2r}\right]. \tag{13}$$

In addition, $D(q_R) \le -\log(1-w)$, $\mathrm{TV}(P_n, R_n) \le w$, and the conditional-head TV error is at most Δ uniformly over all coordinates and partial contexts.

All parameters may depend on n. For fixed r, choosing w of order B and decreasing λ sufficiently slowly gives error tending to one when $nB^{2r} \to \infty$. This produces the exponent in Theorem 2. The upper bound covers every positive finite-exchangeable mismatch law.

4 Technical Overview

The two proofs address the same feedback loop: each batch is generated from current predictions, and its values determine the next predictions. The lower proof keeps a detectable drift in this loop; the upper proof controls its accumulated effect.

4.1 Lower Bound: A Stopped Posterior Ladder

A pilot can suppress one alternative. For the one-alternative model in (6), a prefix with c completed pairs and m mismatches has alternative-to-teacher posterior odds

$$O_c(m) = \frac{w}{1-w}\left(\frac{\tau+\Delta}{\tau}\right)^m \left(\frac{1-\tau-\Delta}{1-\tau}\right)^{c-m}, \qquad p_c(m) - \tau = \Delta\frac{O_c(m)}{1+O_c(m)}. \tag{14}$$

The posterior is the model conditioned on committed values. A teacher prefix has expected log likelihood ratio $-c\ \mathrm{kl}(\tau\|\tau+\Delta)$, where kl denotes Bernoulli KL, and can suppress the alternative. Fair singleton observations leave its weight unchanged. A generated pilot instead uses the current predictive mean, so its mismatch rate also reflects the alternative.

Why several scales resist a pilot. In the ladder, every open-pair head uses the posterior-weighted average of the component rates. Independent draws at this intermediate rate can reduce the weight of a high-drift component while leaving lower-drift components plausible (Figure 1). The relevant quantity is therefore the residual *mean* drift. The proof controls it through successive thresholds $\delta_{j+1} = b\lambda\delta_j$, on an event whose failure probability is bounded uniformly over policies. These are lower bounds on the drift; individual posterior trajectories may fluctuate.

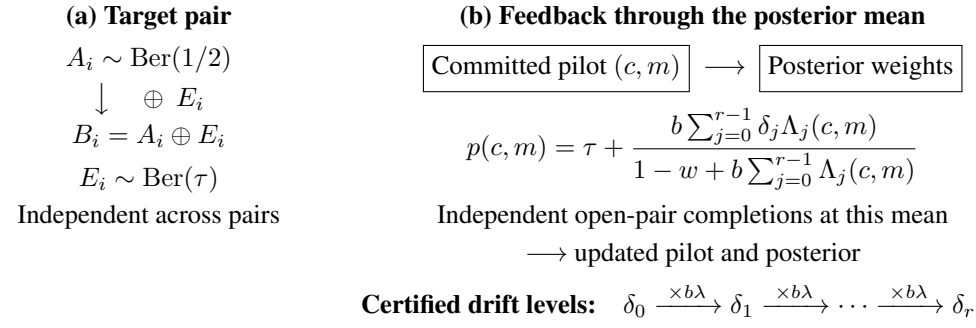


Figure 1: **Pair structure and the posterior ladder.** The target is exactly sampled by revealing all representatives, then their mates. Under the ladder model, each open-pair head averages the component rates, and its generated values feed back into the posterior. On joint certificate success, the residual completion drift is at least δ_r. The arrows between δ_j show the proof thresholds; actual posterior means may fluctuate.

A certificate across scales. For c completed pairs and m mismatches, define the component-to-teacher likelihood ratio and its certificate:

$$\Lambda_j(c,m) = \left(\frac{\tau+\delta_j}{\tau}\right)^m \left(\frac{1-\tau-\delta_j}{1-\tau}\right)^{c-m}, \qquad Z_j = \frac{1-w+b\sum_{\ell=j}^{r-1}\Lambda_\ell}{\Lambda_j}. \tag{15}$$

The numerator retains the teacher and all components at or below level j. Their excess mean is at least $b\delta_j/Z_j$, hence at least δ_{j+1} when $Z_j \le 1/\lambda$. Adding the higher-rate components preserves this bound. Thus one certificate controls the full posterior mean, including its normalization.

Stopping makes the induction valid. For a denominator parameter $\theta_j = \tau + \delta_j$ and a lower parameter $s \le \theta_j$, a completed mismatch with actual mean μ has likelihood-ratio multiplier of expectation

$$\mu\frac{s}{\theta_j} + (1-\mu)\frac{1-s}{1-\theta_j} = 1 + \frac{(\mu-\theta_j)(s-\theta_j)}{\theta_j(1-\theta_j)} \le 1 \quad \text{if } \mu \ge \theta_j. \tag{16}$$

This drift condition is supplied by the preceding certificates. Start each ratio process at the empty context, update through the first failing batch, and freeze thereafter. The decision to update uses only previous outcomes, so the stopped process is a supermartingale with initial weighted expectation at most one. Markov's inequality charges at most λ to each first failure, for total charge at most $r\lambda$.

The induction starts in physical round one: every completed pair is then a co-sample of two fair marginals, with mismatch mean $1/2 \ge \theta_j$. On joint certificate success, every completion has predictable mismatch mean at least $\tau + \delta_r$. Lemma 7 states the exact recursion and first-failure calculation.

A test of the delivered vector. For the final step, we use the same event for every policy:

$$\mathcal{A}_{\mathrm{test}} = \left\{\sum_i E_i \ge n\tau + n\delta_r/2\right\}. \tag{17}$$

Let G denote joint certificate success. A martingale tail bound controls $Q_R^\pi(\mathcal{A}_{\mathrm{test}}^c \cap G)$; teacher Hoeffding controls $P_n(\mathcal{A}_{\mathrm{test}})$. Combining these two tails with the failure charge gives Theorem 3. The event uses only the delivered vector, so the same signed probability gap survives mixing over private seeds. Appendix B supplies the filtrations.

4.2 Upper bound: from target loss to generated samples

First choose the comparison law. The loss controls target contexts, whereas execution produces its own contexts. A full-output comparison bridges these laws. Let $p_c(m)$ be the next-mismatch head after c completed pairs containing m mismatches, and define the teacher-averaged batch discrepancy

$$\mathcal{D}_{n,c,d} = \mathbb{E}_{M\sim\mathrm{Bin}(c,\tau)}\,\mathrm{TV}(\mathrm{Ber}(p_c(M))^{\otimes d}, \mathrm{Ber}(\tau)^{\otimes d}). \tag{18}$$

In comparison law j, the first j mismatch batches use teacher draws and the rest use model heads at the resulting prefix. Adjacent laws share a teacher prefix and the same continuation kernel after the differing batch. TV contraction and the triangle inequality give

$$\mathrm{TV}(P_n, Q_R^{\mathrm{geom}}) \le \sum_{j=1}^{r} \mathcal{D}_{n,c_{j-1},d_j}. \tag{19}$$

Thus it suffices to bound $\mathcal{D}_{n,c,d}$ on teacher prefixes. The implemented schedule continues to use model draws throughout.

Why the loss chain rule alone is too coarse. Write $\mathcal{L}_c = \mathbb{E}_{M\sim\mathrm{Bin}(c,\tau)}\,\mathrm{kl}(\tau\|p_c(M))$. The chain rule gives $\sum_c \mathcal{L}_c \le B$, and Pinsker gives $\mathcal{D}_{n,c,d} \le \sqrt{d\mathcal{L}_c/2}$. Even an estimate $\mathcal{L}_c = O(B/c)$ would yield $\sqrt{Bd/c}$. The sharp exponent requires the stronger term $B\sqrt{d/c}$, which we obtain by separating two parts of the model perturbation.

Separate small energy from small mass. Write the likelihood ratio as

$$f = \frac{dR_E}{dP_E}, \qquad g = \sqrt{f} - 1, \qquad f = 1 + 2g + g^2, \qquad \mathbb{E}_P g^2 \le B. \tag{20}$$

Here $2g$ has small L^2 energy, while g^2 has small total mass and may concentrate on rare configurations.

For a teacher prefix, put $\alpha_c = \mathbb{E}_P[g \mid E_{1:c}]$ and $\beta_c = \mathbb{E}_P[g^2 \mid E_{1:c}]$. Symmetry and the product orthogonal basis give

$$v\mathbb{E}_P(\Delta\alpha_c)^2 \le \frac{B}{n-c}, \qquad \Delta\alpha_c(m) = \alpha_{c+1}(m+1) - \alpha_{c+1}(m). \tag{21}$$

For g^2, normalization by its mass gives an exchangeable law, which decomposes into uniform laws on fixed-total-count *shells*. A prefix count under each shell is hypergeometric. The density-ratio bound of Dümbgen et al. (2021, Theorem 2) controls its predictive shift after a typical teacher prefix. Multiplying by $\mathbb{E}_P g^2 \le B$ yields the loss-linear term. The zero-mass case is $R_E = P_E$.

Use the central window $|M - c\tau| \le t\sqrt{c}$; t measures its width in square-root-count units. Charging exceptional prefixes by their teacher probability gives, for $1 \le c \le n/2$, $1 \le d \le n-c$, and $0 < t \le \sqrt{c}/16$,

$$\mathcal{D}_{n,c,d} \le 16B + 2e^{-2t^2} + 2\sqrt{\frac{2Bd}{n-c}} + 2(t+1)e^{t^2/v}\sqrt{\frac{2d}{vc}}\,B. \tag{22}$$

The last term is the required $B\sqrt{d/c}$ contribution. At the empty prefix, the corresponding bound is

$$\mathcal{D}_{n,0,d} \le \sqrt{\frac{2dB}{n}} + \frac{3}{4\sqrt{2v}}\sqrt{d}\,B. \tag{23}$$

Balance the batches. The loss-linear factors are $\sqrt{c_1}$ and $\sqrt{d_j/c_{j-1}}$. If $c_j \asymp n^{\vartheta_j}$ with $0 = \vartheta_0 < \cdots < \vartheta_r = 1$, their exponents sum to $1/2$. Equal increments $1/r$ minimize the largest exponent, giving (11). For three total rounds, this balances $B\sqrt{c}$ and $B\sqrt{n/c}$ at pilot size $c \asymp \sqrt{n}$. Summing the batch bounds proves the upper threshold; the all-policy lower proves optimality. Appendices C–D give the details.

5 Vanishing error can coexist with growing depth

The finite lower bound permits growing depth: its failure charge vanishes when $r\lambda \to 0$, and its count test separates the laws when $n\delta_r^2 \to \infty$. Set $\ell_n = \log(n+1)$ and, for any fixed $\zeta > 0$, take eventually

$$w_n = \ell_n^{-\zeta}, \quad \Delta_n = \ell_n^{-1}, \quad r_n = \left\lfloor \frac{\ell_n}{4(\zeta+3)\log\ell_n} \right\rfloor, \quad \lambda_n = (r_n\ell_n)^{-1}. \tag{24}$$

Corollary 4 (Failure despite joint and uniform head accuracy). *The models in (12) with (24) satisfy*

$$D(q_{R_n}) \longrightarrow 0, \qquad \mathrm{TV}(P_n, R_n) \longrightarrow 0, \qquad \sup_{S, x_S, i \notin S} \mathrm{TV}(q_{R_n,i}, q_{P_n,i}) \longrightarrow 0. \tag{25}$$

$$\inf_{\pi \in \Pi_{r_n+1}} \mathrm{TV}(P_n, Q^{\pi}_{R_n}) \longrightarrow 1, \qquad \inf_{\pi \in \Pi_{r_n+1}} \mathrm{TV}(R_n, Q^{\pi}_{R_n}) \longrightarrow 1. \tag{26}$$

In the head supremum, both heads are evaluated at the same partial context. Exact target heads still generate P_n in two rounds.

Here $r_n = \Theta(\log n / \log\log n)$. The own-model conclusion follows from $\mathrm{TV}(R_n, Q) \geq \mathrm{TV}(P_n, Q) - w_n$. The budget decays logarithmically: if instead $B_n = n^{-\beta}$ for fixed $\beta > 0$, Theorem 2 gives vanishing error at any fixed depth with $K - 1 > 1/(2\beta)$.

Corollary 5 (Operational tail and expected depth). *Set $K_n = r_n + 1$ and use the models of Corollary 4. Fix $\varepsilon \in (0,1)$. Every almost surely completing policy with the interface of Definition 1, but no hard round cap, and $\mathrm{TV}(P_n, Q^{\pi_n}_{R_n}) \leq \varepsilon$ has completion time T_n satisfying*

$$\mathbb{P}_{\pi_n, R_n}(T_n > K_n) \geq 1 - \varepsilon - o(1), \qquad \mathbb{E}_{\pi_n, R_n} T_n \geq (1 - \varepsilon - o(1)) K_n. \tag{27}$$

Both quantities use the actual execution law, including private randomness.

Completing all remaining coordinates in round K couples a truncated policy to the original whenever it finishes by K; Lemma 8 yields the corollary. For budgets $B_n = \ell_n^{-\zeta}$ with $\zeta > 11/3$, Appendix J also proves a uniform geometric-schedule upper of order $\log n / \log\log n$. The depth order therefore matches in this range.

6 Robustness beyond fair, uninformative endpoints

Fairness has a specific role in the lower proof: singleton likelihoods cancel, so completed mismatch counts determine the posterior. With a non-fair representative, revealing B_i can already provide evidence about the component. The next theorem shows that this information does not remove the growing-depth obstruction, even when the planner may choose either orientation at every pair.

Let $A_i \sim \mathrm{Ber}(a_i)$ independently, where the known array has arbitrary $a_i \in (0,1)$, and keep $B_i = A_i \oplus E_i$. Both target and model use this same representative law, independent of E. Denote the target by $P_{\boldsymbol{a},n}$. In this section, D is defined by (3) with this target, so $D(q_R) = \mathrm{KL}(P_{\boldsymbol{a},n} \| R)$. For $K \geq 2$, use K alternatives rather than $K - 1$, with

$$b = w/K, \qquad \delta_j = \Delta(w\lambda/(9K))^j \quad (0 \leq j \leq K),$$
$$R_E = (1-w)P_E + b \sum_{j=0}^{K-1} \mathrm{Ber}(\tau + \delta_j)^{\otimes n}. \tag{28}$$

Theorem 6 (Uniform heterogeneous-bias lower bound). *For $0 < w \leq 1/8$, $0 < \lambda \leq 1$, and $0 < \Delta \leq 1/4$, the model in (28) satisfies, uniformly over every positive bias array,*

$$\inf_{\pi \in \Pi_K} \mathrm{TV}(P_{\boldsymbol{a},n}, Q^{\pi}_R) \geq [1 - (K-1)\lambda]_+ - 2\exp\left[-\frac{4n\Delta^2}{81}\left(\frac{w\lambda}{9K}\right)^{2K}\right]. \tag{29}$$

Also, $D(q_R) \leq -\log(1-w)$, $\mathrm{TV}(P_{\boldsymbol{a},n}, R) \leq w$, and the worst conditional-head TV error is at most 4Δ. The array may vary with n and approach either endpoint; exact zero or one biases are excluded.

For example, $w = \ell_n^{-1}$, $\Delta = (4\ell_n)^{-1}$, $K = \lfloor \ell_n / (16 \log \ell_n) \rfloor$, and $\lambda = (K\ell_n)^{-1}$ give the analogue of Corollary 4 uniformly over such arrays. Exact target heads retain their two-round sampler.

Why bias requires a different argument. Under mismatch rate s, a singleton B_i has mean $a_i + (1 - 2a_i)s$ and can update the component posterior. Reverse heads must average the nonlinear component conditionals under this full posterior; plugging a posterior mean rate into a reverse head is incorrect. Moreover, reverse sensitivities vanish as a_i approaches zero or one. A teacher conditional potential charges both the first observation and completion of each pair, recovering a uniform drift. Appendix E gives the channel formulas and proof.

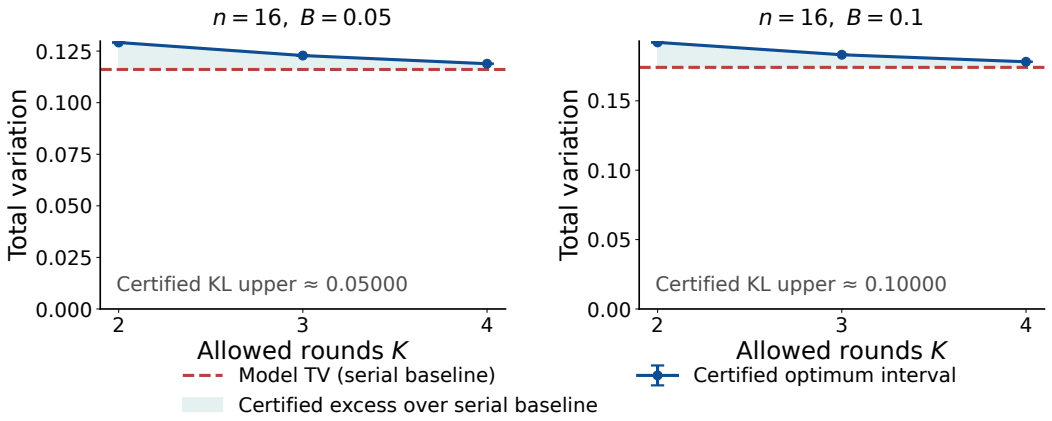


Figure 2: **Finite-depth cost for a fixed coherent oracle.** Each panel fixes one $n = 16$ model. Blue markers show optimal full-output TV within certified intervals narrower than 6.1×10^{-10}. Dashed lines give the serial sampler's model TV; shading shows certified excess over this feasible baseline.

The fixed-depth gap for prescribed biases. The geometric upper extends unchanged to every prescribed positive bias array: after the exact representative round, the mismatch kernels and KL budget are the same as in the fair model. Appendix E.4 gives the reduction. Thus $n^{-1/[2(K-1)]}$ remains a sufficient budget scale, while (29) forces worst-case TV error to one if $n^{1/(2K)} B_n \to \infty$. Closing this gap for a prescribed non-fair law remains open. If the minimax instead ranges over all bias arrays, the fair subclass and the uniform upper already give the exponent in Theorem 2.

Beyond binary alphabets. Appendix F gives lower bounds for fixed-marginal affine interactions with positive conditional score energy in both orientations, including heterogeneous non-exchangeable three-valued pairs. This extension retains uninformative singletons and has no matching upper. The truncation argument also converts both extensions' growing-depth bounds to operational tail and expectation bounds.

7 Certified Finite-Dimensional Experiments

Do short executions incur a detectable cost before the asymptotic regime? For a fixed oracle define $d_K(R) = \inf_{\pi \in \Pi_K} \mathrm{TV}(P_n, Q_R^\pi)$. We compute certificates $L_K \le d_K(R) \le U_K$: a terminal-test Bellman bound covers every policy in Definition 1, and an explicit legal policy supplies the upper bound. Both concern full-output TV with private randomness marginalized.

Setup. We construct coherent mixture oracles for $n \in \{8, 12, 16\}$ pairs and total-KL budgets $B \in \{0.02, 0.05, 0.10\}$. A search over 54 candidates selects one oracle per (n, B) by a floating-point estimate of $d_3(R) - \mathrm{TV}(P_n, R)$. We freeze that oracle and certify depths $K = 2, 3, 4$. All nine models satisfy the actual KL budget and an all-context head-TV bound of $1/4$. Linear programming proposes tests and policies; rational arithmetic with directed rounding certifies them independently of solver tolerances. Appendix G gives full settings and results. Appendix H adds error-amplification, pilot, and small-instance diagnostics. No neural-network training is used.

Execution cost and depth separation. All nine models have $L_3 - \mathrm{TV}(P_n, R) > 0$, between 0.00273 and 0.01474, and satisfy $L_3 > U_4$. This latter inequality certifies a depth separation independently of the serial comparison. For $n = 16$, $B = 0.10$, the model TV is 0.174090, whereas d_2, d_3, d_4 are approximately 0.192094, 0.183191, and 0.178124 (Figure 2). These selected-oracle results establish finite-dimensional existence; Appendix G.1 relates their scope to the asymptotic ladder bound.

Simple schedules and informative endpoints. At three rounds, an optimized fixed batch vector with a hidden uniform pair permutation is within 9.3×10^{-13} of the optimum on all nine models. At

four rounds this family is strictly suboptimal, and equal fixed-label batches beat geometric batches in every comparison. Thus oracle-specific finite-instance rankings differ from the uniform asymptotic guarantee. With biased representatives, $n = 6$, and $a = 1/4$, we certify $L_3 - U_4 > 0.00390$ and head error below $1/4$. Appendix G.6 reports all cases, including the one head-bound exception.

A budget-matched ladder diagnostic. Appendix I compares a single alternative with the genuine two-level ladder at the same actual KL. For $B = 0.05$, $n = 10^8$, and a generated pilot of size $\lfloor \sqrt{n} \rfloor$, a terminal count test gives TV lower witnesses 0.0112 and 0.6469, respectively. These selected-policy diagnostics expose residual ladder drift at larger dimensions; the appendix separately displays the analytic all-policy bound and target-pilot controls.

8 Related work and discussion

Conditional models and generation order. Absorbing-diffusion scores factor through clean-data conditionals (Ou et al., 2025). Connections to order-agnostic autoregression motivate optimized schedules (Hoogeboom et al., 2022) and inference-aligned training weights (Shih et al., 2022). Oracle-based evaluations distinguish model and sampler error (Tang et al., 2026); we ask how much conditional accuracy preserves shallow execution.

Sampling guarantees and schedules. Prediction and within-batch dependence drive adaptive sampling bounds (Lavenant & Zanella, 2025; Ben-Hamu et al., 2025). Other guarantees concern random-order batches (Li & Cai, 2025), dependence-sensitive schedules (Zhao & Cai, 2026; Wainwright, 2026), and information curves (Chen et al., 2026). The latter also proves randomized adaptive TV query lower bounds with an unknown target. Liang et al. (2026) give TV convergence and a constant-step-size Euler lower bound, and matching KL error bounds for the first-hitting sampler. Our threshold instead concerns fully informed execution throughout a coherent KL neighborhood.

The closest execution model. Wen et al. (2026) characterize known-distribution, independent-marginal execution through conditional total correlation; their converses use deterministic policies and KL error. We allow private randomization and value adaptivity, measure marginal output TV, and match the fixed-depth loss threshold near a two-round target.

The common output test preserves the TV converse under private randomization. Appendix M compares assumptions and explains the relation to conditional total correlation.

Why candidate selection changes the depth. The separation in Theorem 12 calibrates the pilot before commitment. In the fair exchangeable class, first reveal all representatives, then independently draw one candidate per remaining coordinate from its original head. A target-aware acceptance rule uses the target-to-model mismatch probability ratio. Conditional on enough acceptances, a uniformly selected pilot has the teacher product law independently of its labels. For $n \geq 2$, $m = \lfloor n/2 \rfloor$, and $B \leq 1/128$, the shortage probability is at most $e^{-n/8}$. The third round completes all remaining coordinates using their original heads, giving

$$\mathrm{TV}(P_n, Q_R^{\mathrm{sel}}) \leq e^{-n/8} + \mathcal{D}_{n,m,n-m}.$$

Our teacher-prefix bound therefore yields vanishing full-output TV whenever $B_n \to 0$, using at most $3n$ proposals and preserving every committed value. Selection prevents the pilot from inheriting the model's prediction bias. This guarantee uses known target probabilities; it does not automatically extend to confidence top-k selection. Appendix K gives the acceptance rule and failure branch.

Scope and limitations. The sharp threshold concerns noisy-pair targets, coherent heads, and the finite-exchangeable mismatch class; finite exchangeability includes laws without a product-mixture representation. The resource is execution depth at a population-loss budget. Head correction, joint sampling, and revision or rejection of generated values enlarge the interface. Appendix K.2 examines nonexchangeable errors, and Proposition 13 gives a two-round guarantee when mismatch total correlation is small. Constants, the critical error curve, and the prescribed-bias exponent remain open; the growing-depth upper currently covers $\zeta > 11/3$. Appendix L treats other output distances, and Appendix M discusses interpretation and extensions.

AI USE STATEMENT

Generative AI tools assisted with literature review, theoretical formulation and proofs, computational checks, and manuscript preparation. The authors take responsibility for the content.

REPRODUCIBILITY STATEMENT

Definition 1 and equations (3)–(5) specify the execution model, loss normalization, and order of quantifiers. Appendices A–F give the core analytic proofs; Appendices J–L prove additional depth, interface, and metric guarantees. The experimental appendices specify the models, parameters, selection rules, policy classes, and evaluated grids, distinguishing rational certificates from floating-point diagnostics. An accompanying anonymous code package contains numerical sources, frozen inputs, and a standard-library verifier that independently recomputes all 27 fair and 15 biased certificates without running the optimizer. The experiments use no external datasets, trained neural-network weights, or Monte Carlo estimates; the analytic theorems do not rely on these computations.

ETHICS STATEMENT

This work studies mathematical limitations of a specified sampling interface. It uses no human-subject data and releases no trained generative model. Its worst-case bounds do not measure the reliability, fairness, or safety of deployed language models.

REFERENCES

Jacob Austin, Daniel D. Johnson, Jonathan Ho, Daniel Tarlow, and Rianne van den Berg. Structured denoising diffusion models in discrete state-spaces. In *Advances in Neural Information Processing Systems*, volume 34, pp. 17981–17993, 2021. URL https://arxiv.org/abs/2107.03006.

Heli Ben-Hamu, Itai Gat, Daniel Severo, Niklas S. Nolte, and Brian Karrer. Accelerated sampling from masked diffusion models via entropy bounded unmasking. In *Advances in Neural Information Processing Systems*, volume 38, pp. 55981–56007, 2025. URL https://arxiv.org/abs/2505.24857.

Sitan Chen, Sinho Chewi, Jerry Li, Yuanzhi Li, Adil Salim, and Anru R. Zhang. Sampling is as easy as learning the score: Theory for diffusion models with minimal data assumptions. In *International Conference on Learning Representations*, 2023. URL https://arxiv.org/abs/2209.11215.

Sitan Chen, Kevin Cong, and Jerry Li. Optimal inference schedules for masked diffusion models. In *Proceedings of the Thirty Ninth Conference on Learning Theory*, volume 336 of *Proceedings of Machine Learning Research*, pp. 1279–1311, 2026. URL https://proceedings.mlr.press/v336/chen26e.html.

Persi Diaconis and David Freedman. Finite exchangeable sequences. *The Annals of Probability*, 8(4), 1980. doi: 10.1214/aop/1176994663. URL https://doi.org/10.1214/aop/1176994663.

Lutz Dümbgen, Richard Samworth, and Jon Wellner. Bounding distributional errors via density ratios. *Bernoulli*, 27(2):818–852, 2021. doi: 10.3150/20-BEJ1256. URL https://arxiv.org/abs/1905.03009v6.

Guhao Feng, Yihan Geng, Jian Guan, Wei Wu, Liwei Wang, and Di He. Theoretical benefit and limitation of diffusion language model. In *Advances in Neural Information Processing Systems*, volume 38, 2025. doi: 10.52202/085713-0825. URL https://doi.org/10.52202/085713-0825.

Jiatao Gu, James Bradbury, Caiming Xiong, Victor O. K. Li, and Richard Socher. Non-autoregressive neural machine translation. In *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=B1l8BtlCb.

Emiel Hoogeboom, Alexey A. Gritsenko, Jasmijn Bastings, Ben Poole, Rianne van den Berg, and Tim Salimans. Autoregressive diffusion models. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=Lm8T39vLDTE.

Jaeyeon Kim, Kulin Shah, Vasilis Kontonis, Sham M. Kakade, and Sitan Chen. Train for the worst, plan for the best: Understanding token ordering in masked diffusions. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 30749–30768, 2025. URL https://proceedings.mlr.press/v267/kim25ah.html.

Hugo Lavenant and Giacomo Zanella. Error bounds and optimal schedules for masked diffusions with factorized approximations. *arXiv preprint arXiv:2510.25544*, 2025. URL https://arxiv.org/abs/2510.25544v2.

Gen Li and Changxiao Cai. Breaking AR's sampling bottleneck: Provable acceleration via diffusion language models. In *Advances in Neural Information Processing Systems*, volume 38, pp. 11700–11725, 2025. doi: 10.52202/085713-0392. URL https://doi.org/10.52202/085713-0392.

Yuchen Liang, Zhiheng Tan, Ness Shroff, and Yingbin Liang. Sharp convergence rates for masked diffusion models. *arXiv preprint arXiv:2602.22505*, 2026. URL https://arxiv.org/abs/2602.22505v1.

Anji Liu, Oliver Broadrick, Mathias Niepert, and Guy Van den Broeck. Discrete copula diffusion. In *International Conference on Learning Representations*, 2025. URL https://arxiv.org/abs/2410.01949.

Aaron Lou, Chenlin Meng, and Stefano Ermon. Discrete diffusion modeling by estimating the ratios of the data distribution. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 32819–32848, 2024. URL https://proceedings.mlr.press/v235/lou24a.html.

Jingyang Ou, Shen Nie, Kaiwen Xue, Fengqi Zhu, Jiacheng Sun, Zhenguo Li, and Chongxuan Li. Your absorbing discrete diffusion secretly models the conditional distributions of clean data. In *International Conference on Learning Representations*, 2025. URL https://arxiv.org/abs/2406.03736.

Subham Sekhar Sahoo, Marianne Arriola, Yair Schiff, Aaron Gokaslan, Edgar Marroquin, Justin T. Chiu, Alexander Rush, and Volodymyr Kuleshov. Simple and effective masked diffusion language models. In *Advances in Neural Information Processing Systems*, volume 37, pp. 130136–130184, 2024. URL https://arxiv.org/abs/2406.07524.

Jiaxin Shi, Kehang Han, Zhe Wang, Arnaud Doucet, and Michalis K. Titsias. Simplified and generalized masked diffusion for discrete data. In *Advances in Neural Information Processing Systems*, volume 37, 2024. URL https://arxiv.org/abs/2406.04329.

Andy Shih, Dorsa Sadigh, and Stefano Ermon. Training and inference on any-order autoregressive models the right way. In *Advances in Neural Information Processing Systems*, volume 35, pp. 2762–2775, 2022. URL https://arxiv.org/abs/2205.13554.

Luhan Tang, Longxuan Yu, Shaorong Zhang, and Greg Ver Steeg. Is your diffusion sampler actually correct? a sampler-centric evaluation of discrete diffusion language models. In *International Conference on Machine Learning*, 2026. URL https://icml.cc/virtual/2026/poster/61598.

Benigno Uria, Iain Murray, and Hugo Larochelle. A deep and tractable density estimator. In *Proceedings of the 31st International Conference on Machine Learning*, volume 32 of *Proceedings of Machine Learning Research*, pp. 467–475, 2014. URL https://proceedings.mlr.press/v32/uria14.html.

Martin J. Wainwright. The data geometry of masking diffusion: Certified-optimal schedules via unmasking growth complexity. *arXiv preprint arXiv:2608.13520*, 2026. URL https://arxiv.org/abs/2608.13520v1.

Chuling Wen, Weijie Liang, and Jian Lu. Conditional total correlation and the serial depth of adaptive parallel sampling. *arXiv preprint arXiv:2608.25505*, 2026. URL https://arxiv.org/abs/2608.25505v1.

Yunxiao Zhao and Changxiao Cai. Adaptation to intrinsic dependence in diffusion language models. *arXiv preprint arXiv:2602.20126*, 2026. URL https://arxiv.org/abs/2602.20126v1.

A Loss normalization and elementary comparisons

Write $\mathrm{kl}(u\|v) = \mathrm{KL}(\mathrm{Ber}(u)\|\,\mathrm{Ber}(v))$. Expectations with subscript P refer to the target; unqualified expectations in an execution proof refer to the actual policy law. We first identify the loss with a population masked-prediction objective, then use coherence to reduce it to joint KL.

Independent masking. Let S_t retain each coordinate independently with probability $1-t$, independently of $X \sim P$. For arbitrary positive heads q that depend on the visible context but not additionally on t, define e_i and $\eta_{\mathrm{train}}(q)$ by (3). The population excess of the continuous-time masked cross-entropy objective with retention schedule $\alpha_t = 1-t$ (Shi et al., 2024, Eq. (5)) is

$$\int_0^1 \frac{1}{t}\mathbb{E}_{S_t,X\sim P}\sum_{i\notin S_t} e_i(S_t, X_{S_t})\,dt = N\eta_{\mathrm{train}}(q). \tag{30}$$

Indeed, for a particular visible set S of size $k < N$ and a missing coordinate i, its integrated coefficient is

$$\int_0^1 t^{N-k-1}(1-t)^k\,dt = \frac{k!(N-k-1)!}{N!} = \frac{1}{\binom{N}{k}(N-k)}.$$

The integral is over $0 < t < 1$. Subtracting the loss of the true conditional head converts each conditional cross-entropy to e_i, so (30) identifies the population excess loss. The identity holds for arbitrary positive heads, and the integral is finite when there are finitely many such heads.

Coherent heads. Order-agnostic training averages autoregressive objectives over reveal orders (Uria et al., 2014; Hoogeboom et al., 2022); absorbing diffusion is also connected to clean-data conditionals (Ou et al., 2025). We use the standard chain rule under the additional assumption that all heads are conditionals of one joint law. The identity below fixes the normalization of our accuracy budget. For a positive joint law R and a permutation σ of $[N]$, the chain rule gives

$$\mathrm{KL}(P\|R) = \sum_{k=0}^{N-1}\mathbb{E}_{X\sim P}\,\mathrm{KL}\big(P_{\sigma(k+1)}(\cdot\mid X_{\sigma(1:k)})\|R_{\sigma(k+1)}(\cdot\mid X_{\sigma(1:k)})\big). \tag{31}$$

A uniform permutation has a uniform size-k prefix set, followed by a uniform coordinate outside that set. Averaging (31) proves $N\eta_{\mathrm{train}} = \mathrm{KL}(P\|R)$ on any finite alphabet. In the binary pair model, $(A,B)\leftrightarrow(A,E)$ is a bijection and the common representative law is independent of E, so its factors cancel and give (4). This cancellation does not require fair representatives.

Elementary comparisons. We will use the elementary bounds

$$\mathrm{TV}(\mathrm{Ber}(p)^{\otimes d},\mathrm{Ber}(\tau)^{\otimes d}) \le \sqrt{\frac{d}{2}\,\mathrm{kl}(p\|\tau)} \le \sqrt{\frac{d}{2v}}|p-\tau|. \tag{32}$$

The second inequality follows from $\mathrm{KL}\le\chi^2$ for Bernoulli laws. If $f = dR/dP$, then Jensen's inequality gives $-2\log\mathbb{E}_P\sqrt{f} \le \mathbb{E}_P[-\log f]$ and $-\log x \ge 1-x$ gives

$$H_{\mathrm{sq}} := \mathbb{E}_P(\sqrt{f}-1)^2 = 2 - 2\mathbb{E}_P\sqrt{f} \le \mathrm{KL}(P\|R). \tag{33}$$

The notation H_{sq} is kept distinct from η_{train}.

B Proof of the finite fair-pair lower bound

B.1 Literal conditional heads and the full posterior

Use the parameters of Theorem 3 and write $\theta_j = \tau + \delta_j$. A singleton observation in any pair is fair under every mixture component. The likelihood ratio of a partial context to the target therefore depends only on the completed mismatch count (c,m) and is $\Lambda_j(c,m)$ from (15). The heads average the component predictions using the model posterior at the committed context. The posterior mismatch mean for an open pair is

$$p(c,m) - \tau = \frac{b\sum_{j=0}^{r-1}\delta_j\Lambda_j(c,m)}{1-w+b\sum_{j=0}^{r-1}\Lambda_j(c,m)}. \tag{34}$$

Table 1: Notation used in the main results, proofs, and numerical certificates.

Symbols	Meaning
n, $N = 2n$	Number of pairs and number of coordinates in the pair model.
K, $r = K - 1$, B	Round cap, fair-ladder component count, and joint-KL budget.
b, δ_j, v	Component weight, component excess mismatch rate, and target mismatch variance $\tau(1-\tau)$.
a_i, $v_{A,i}$	Representative bias and variance $a_i(1-a_i)$ in the biased extension.
α_c, β_c	Conditional projections of $g = \sqrt{dR_E/dP_E} - 1$ and g^2 after observing c mismatch bits under the target.
Λ_j, Z_j	Component likelihood ratio and posterior-drift certificate.
$\mathcal{M}_{j,s,i}$	Stopped likelihood-ratio process through round i.
$\mathcal{A}_{\text{test}}$, T_π	Terminal output test event and policy completion time.
$\mathcal{D}_{n,c,d}$	Expected TV error of a size-d batch after a length-c target prefix.
U_s, $\mathcal{R}_s$	Uniform law on the count-s shell and its prefix density ratio relative to the target.
L_K, U_K	Numerical lower and upper certificates for the optimal K-round TV error.

This is the same mismatch mean in either orientation. It is unaffected by labels, the values of singly observed endpoints, or their number. For a fresh pair, the execution samples its two endpoints independently from their fair marginal heads, so a co-sample has mismatch mean $1/2$. This distinction between the model posterior and the actual sampling rule drives the lower bound.

Let Z_j be as in (15). In the submixture containing the target and components $j, \dots, r-1$, the normalized excess mean is at least

$$\frac{b\delta_j\Lambda_j}{1-w+b\sum_{\ell=j}^{r-1}\Lambda_\ell} = \frac{b\delta_j}{Z_j} \geq b\lambda\delta_j = \delta_{j+1} \quad \text{on } \{Z_j \leq 1/\lambda\}. \tag{35}$$

All omitted components have excess at least $\delta_j \geq \delta_{j+1}$. Adding them cannot lower the full mean below $\tau+\delta_{j+1}$. This proves the certificate implication with the entire posterior denominator present.

B.2 A PREDICTABLE STOPPED PROCESS FOR EACH DESTINATION

The certificate implication controls the drift while earlier certificates hold. The next lemma stops the ratio updates after the first earlier certificate fails and bounds the probability charged to each possible first failure.

For a rate s, define $\Lambda^{(s)}(c,m) = (s/\tau)^m((1-s)/(1-\tau))^{c-m}$, so $\Lambda_j = \Lambda^{(\theta_j)}$ and $\Lambda^{(\tau)} = 1$. For $s \in \{\tau, \theta_j, \dots, \theta_{r-1}\}$, a completed mismatch Y multiplies $\Lambda^{(s)}/\Lambda_j$ by

$$M_{j,s}(Y) = \left(\frac{s}{\theta_j}\right)^Y \left(\frac{1-s}{1-\theta_j}\right)^{1-Y}. \tag{36}$$

Its conditional expectation is at most one whenever the actual mean of Y is at least θ_j, by (16). Birth-only observations have multiplier one.

Number rounds $1, \dots, r+1$ and let G_j denote $Z_j \leq 1/\lambda$ at the end of round $j+1$, for $0 \leq j < r$. Let $\mathcal{H}_i$ contain the committed history and private choices exposed through round i, and let C_{i+1} be the set of pairs completed in the selected action of round $i+1$. Write Y_k for the mismatch of a completed pair k.

Lemma 7 (Stopped ratios and first-failure accounting)**.** *Fix a destination $0 \leq j < r$ and a rate $s \in \{\tau, \theta_j, \dots, \theta_{r-1}\}$. For $0 \leq i \leq j$, define the active indicator and the nonnegative process*

$$I_i^{\text{act}} = \mathbf{1}\{G_0 \cap \cdots \cap G_{i-1}\}, \qquad \mathcal{M}_{j,s,0} = 1,$$

$$\mathcal{M}_{j,s,i+1} = \mathcal{M}_{j,s,i}\left(1 - I_i^{\text{act}} + I_i^{\text{act}} \prod_{k \in C_{i+1}} M_{j,s}(Y_k)\right), \tag{37}$$

where an empty intersection has indicator one and an empty product equals one. Then $(\mathcal{M}_{j,s,i})_{i=0}^{j+1}$ *is a supermartingale with respect to* $(\mathcal{H}_i)_{i=0}^{j+1}$*. The weighted process*

$$\widetilde{Z}_{j,i} = (1-w)\mathcal{M}_{j,\tau,i} + b\sum_{\ell=j}^{r-1}\mathcal{M}_{j,\theta_\ell,i}$$

therefore satisfies

$$\mathbb{E}\widetilde{Z}_{j,j+1} \le Z_j(\varnothing) = 1 - w + b(r-j) = 1 - bj \le 1. \tag{38}$$

Moreover,

$$\mathbb{P}(G_0 \cap \cdots \cap G_{j-1} \cap G_j^c) \le \lambda, \qquad \mathbb{P}\left(\bigcap_{j=0}^{r-1} G_j\right) \ge [1 - r\lambda]_+. \tag{39}$$

Proof. The indicator I_i^{act} is $\mathcal{H}_i$-measurable, so the decision to update precedes the fresh outcomes of round $i+1$. In the first round there are no open pairs, and every completion is a co-sample of mean $1/2 \ge \theta_j$. In an active later round $i+1$, the preceding certificate G_{i-1} and (35) give open-completion mean at least $\theta_i \ge \theta_j$. Co-samples again have mean $1/2$. Conditional on the complete pre-round history and selected action, fresh coordinate draws are independent. Grouping co-sampled coordinates by pair, the expected batch factor is a product of factors each at most one by (36). Averaging over a randomized action preserves this inequality. Inactive values remain unchanged, proving the supermartingale assertion and (38).

On the first-failure event $G_0 \cap \cdots \cap G_{j-1} \cap G_j^c$, every update through round $j+1$ is active. Hence $\mathcal{M}_{j,s,j+1}$ equals the actual ratio $\Lambda^{(s)}/\Lambda_j$ and $\widetilde{Z}_{j,j+1} = Z_j > 1/\lambda$. Markov's inequality gives the first bound in (39); the disjointness of the first-failure events gives the second. Thus (37) includes the batch at which the first failure occurs, then freezes every later value. The supermartingale bound holds under the actual execution law, before restricting to a first-failure event. Empty actions have identity factors, and paths that finish early can be padded with identity stages for analysis. □

B.3 Filtration and a policy-independent terminal event

Expose the exactly n mismatch completions in physical completion order, with a fixed label tie-break inside each batch. Before each completion, expose the selected action, private choices used to select it, and intervening birth-only outcomes. For a fresh co-sample, expose one fair endpoint as auxiliary side information before its mismatch. Its mismatch remains fair; an open completion has the frozen round-entry mean in (34). Independence of fresh draws within the physical batch makes these conditional laws valid after earlier blocks are exposed.

More precisely, let $\mathcal{F}_\ell^-$ be the information immediately before mismatch Y_ℓ, and let $\mathcal{F}_\ell^+$ additionally contain its outcome. Then $\mathcal{F}_{\ell-1}^+ \subseteq \mathcal{F}_\ell^- \subseteq \mathcal{F}_\ell^+$. Set $\mu_\ell = \mathbb{E}[Y_\ell \mid \mathcal{F}_\ell^-]$. Conditional Hoeffding's lemma and a tower step through the intervening information yield, for every $s \in \mathbb{R}$ and $t \ge 0$,

$$\mathbb{E}\exp\left(s\sum_{\ell=1}^{n}(Y_\ell - \mu_\ell)\right) \le e^{ns^2/8}, \qquad \mathbb{P}\left(\sum_{\ell=1}^{n}(Y_\ell - \mu_\ell) \le -t\right) \le e^{-2t^2/n}. \tag{40}$$

This is an unconditional bound under the original execution law, for every policy. On the event in (39), every completion has predictable mean at least $\tau + \delta_r$, including completions in the last round. The hard cap requires that all n pairs finish.

For the terminal event $\mathcal{A}_{\mathrm{test}}$ in (17), Hoeffding's inequality under the target gives $P_n(\mathcal{A}_{\mathrm{test}}) \le e^{-n\delta_r^2/2}$. On the good event, $\mathcal{A}_{\mathrm{test}}^c$ implies a deviation below $-n\delta_r/2$ in (40). Therefore

$$Q_R^\pi(\mathcal{A}_{\mathrm{test}}) - P_n(\mathcal{A}_{\mathrm{test}}) \ge [1 - r\lambda]_+ - 2e^{-n\delta_r^2/2}. \tag{41}$$

Substituting $\delta_r = \Delta(w\lambda/r)^r$ proves (13). The event $\mathcal{A}_{\mathrm{test}}$ depends only on the final output, so the same bound applies to the marginal output law of a privately randomized policy.

Finally, $R_n \ge (1-w)P_n$ implies $D(q_R) \le -\log(1-w)$, and the mixture form gives $\mathrm{TV}(P_n, R_n) \le w$. Singleton heads agree exactly with the target, and every open mismatch head lies between τ and $\tau + \Delta$, proving the uniform head bound. This completes Theorem 3.

B.4 From a hard cap to a tail and an expectation bound

The following reduction applies to any finite product space with the execution interface of Definition 1. All probabilities and expectations of the completion time are under the actual execution law, including the policy's private randomness.

Lemma 8 (Truncation of a variable-length policy)**.** *Fix a target law P and a positive model law R on a finite product space. A variable-length policy obeys the independent original-head and irrevocable-commitment rules of Definition 1, with no deterministic cap, and completes all coordinates almost surely. Let T_π be the first round by whose end all coordinates are committed, and let Q_R^π be its output law. For an integer $K \geq 1$, define*

$$d_K(R) = \inf_{\rho \in \Pi_K} \mathrm{TV}(P, Q_R^\rho).$$

If $\mathrm{TV}(P, Q_R^\pi) \leq \varepsilon$, then

$$\mathbb{P}_{\pi,R}(T_\pi > K) \geq [d_K(R) - \varepsilon]_+, \qquad \mathbb{E}_{\pi,R} T_\pi \geq K[d_K(R) - \varepsilon]_+. \tag{42}$$

The expectation may be infinite.

Proof. Construct $\pi^{\leq K} \in \Pi_K$ by following π for the first $K-1$ rounds, stopping if all coordinates have been committed, and otherwise selecting every remaining coordinate in round K. This last batch is sampled independently from the original heads at its pre-round context.

Couple the policies using the same private choices and coordinate sampling randomness through round $K-1$. At round K, use the same fresh coordinate randomness wherever both policies sample a coordinate. On $\{T_\pi = K\}$, the original policy must select all remaining coordinates before observing that round's samples. Its conditional batch law is therefore exactly the truncation's batch law. The outputs also agree on $\{T_\pi < K\}$, so the coupling inequality and the triangle inequality give

$$\begin{aligned} d_K(R) &\leq \mathrm{TV}(P, Q_R^{\pi^{\leq K}}) \\ &\leq \mathrm{TV}(P, Q_R^\pi) + \mathrm{TV}(Q_R^\pi, Q_R^{\pi^{\leq K}}) \leq \varepsilon + \mathbb{P}_{\pi,R}(T_\pi > K). \end{aligned}$$

Taking the positive part proves the tail bound. The pointwise inequality $T_\pi \geq K\mathbf{1}\{T_\pi > K\}$ proves the expectation bound. The coupling changes neither the model nor any head; it only changes the last selected action. □

Application to growing depth. For the single model sequence of Corollary 4, put $K_n = r_n + 1$. Equation (26) gives $d_{K_n}(R_n) \geq 1 - o(1)$, uniformly over the policies in Π_{K_n}. Applying Lemma 8 with a fixed $\varepsilon \in (0,1)$ proves Corollary 5 and $\mathbb{E}_{\pi_n,R_n} T_n = \Omega_\varepsilon(\log n / \log\log n)$. The same argument applies to the biased and finite-alphabet growing-depth constructions under their stated uniform-constant assumptions.

C Uniform upper bound over the finite-exchangeable class

C.1 Mean shift and symmetric projection

Let $f = dR_E/dP_E$, $g = \sqrt{f} - 1$, and $H_{\mathrm{sq}} = \mathbb{E}_P g^2 \leq B$. Write $S = \sum_i E_i$, and put $V = S/n - \tau$. Exchangeability gives $p_0 = \mathbb{E}_R S/n$. Since $\mathbb{E}_P V = 0$, $\mathbb{E}_P V^2 = v/n$, and $|V| \leq 3/4$,

$$|p_0 - \tau| = |2\mathbb{E}_P(Vg) + \mathbb{E}_P(Vg^2)| \leq 2\sqrt{vB/n} + \tfrac{3}{4}B. \tag{43}$$

Together with (32), this proves (23).

The products of $\xi_i = (E_i - \tau)/\sqrt{v}$ form an orthonormal basis under P_E. Since g is symmetric, there are coefficients γ_j such that

$$g = \sum_{A \subseteq [n]} \gamma_{|A|} \prod_{i \in A} \xi_i, \quad \mathrm{Var}_P(g) = \sum_{j=1}^{n} \binom{n}{j} \gamma_j^2, \quad v\mathbb{E}_P(\Delta\alpha_c)^2 = \sum_{j=1}^{c+1} \binom{c}{j-1} \gamma_j^2. \tag{44}$$

Exactly $(n-c)\binom{c}{j-1}$ of the j-subsets have one element outside $[c]$. Hence $\binom{n}{j} \geq (n-c)\binom{c}{j-1}$, proving (21). The iid target law supplies the orthonormal basis; exchangeability of R_E makes the coefficients depend only on subset size.

C.2 A SHELL ESTIMATE FROM A CLASSICAL DENSITY-RATIO BOUND

For an integer $s \in \{0, \dots, n\}$, let U_s be uniform on mismatch vectors with total count s. For a particular length-c prefix with m ones, define

$$\mathcal{R}_s(c, m) = \frac{U_s(E_{1:c} = e_{1:c})}{P_E(E_{1:c} = e_{1:c})} = \frac{\mathrm{Hyp}(n, s, c)(m)}{\mathrm{Bin}(c, \tau)(m)}, \qquad z_s = \frac{s - m}{n - c}. \tag{45}$$

The count multiplicity cancels in the ratio. Incompatible shells have $\mathcal{R}_s = 0$ and contribute zero in products below.

Lemma 9 (Shell smoothing; consequence of Dümbgen et al. (2021))**.** *For integers n, c, m with $1 \le c \le n/2$, $0 \le m \le c$, and $t > 0$, if $|m - c\tau| \le t\sqrt{c}$ and $t/\sqrt{c} \le 1/16$, then*

$$\sup_s \mathcal{R}_s(c, m)|z_s - \tau| \le J_{c,t} := \frac{2(t+1)e^{t^2/v}}{\sqrt{c}}. \tag{46}$$

Proof. Put $x = m/c$ and $\theta = s/n$. In our notation, Dümbgen et al. (2021, Theorem 2) states that, for integers $n \ge 2$, $1 \le c \le n$, $c - 1 \le n/2$, and $s \in \{0, \dots, n\}$,

$$\sup_{k:\, \mathrm{Bin}(c,s/n)(k) > 0} \frac{\mathrm{Hyp}(n, s, c)(k)}{\mathrm{Bin}(c, s/n)(k)} \le (1 - 1/n)^{-(c-1)} \le [1 - (c-1)/n]^{-1}.$$

At $s = 0$ and $s = n$, both laws are the same point mass, so their density ratio on its support is one. The assumption $c \le n/2$ meets the theorem's conditions and bounds the last display by two. The central-window assumption puts $x \in [3/16, 5/16]$, so endpoint shells are incompatible with this prefix and contribute zero to (46). For the remaining compatible shells, Bernoulli likelihood ratios, $\mathrm{kl}(x\|\tau) \le (x - \tau)^2/v$, and Pinsker give

$$\mathcal{R}_s(c, m) \le 2e^{c\,\mathrm{kl}(x\|\tau) - c\,\mathrm{kl}(x\|\theta)} \le 2e^{t^2/v} e^{-2c(\theta - x)^2}.$$

Since $n\theta = cx + (n-c)z_s$, we have $|z_s - \tau| \le 2|\theta - x| + t/\sqrt{c}$. Now use $\sup_{u \ge 0} 2ue^{-2cu^2} \le 1/\sqrt{c}$ and $e^{-2cu^2} \le 1$. □

C.3 THE TARGET-PREFIX BATCH BOUND

For a prefix of count m, let $\alpha_c(m) = \mathbb{E}_P[g \mid E_{1:c} = e_{1:c}]$ and $\beta_c(m) = \mathbb{E}_P[g^2 \mid E_{1:c} = e_{1:c}]$, and set

$$h_c(m) = \mathbb{E}_P[f \mid E_{1:c} = e_{1:c}] = 1 + 2\alpha_c(m) + \beta_c(m). \tag{47}$$

For $u \in \{\alpha, \beta\}$, write $\Delta u_c(m) = u_{c+1}(m+1) - u_{c+1}(m)$. The next head satisfies $p_c(m) = \tau h_{c+1}(m+1)/h_c(m)$. Conditional centering under the product target gives

$$p_c(m) - \tau = \frac{2v\Delta\alpha_c(m) + v\Delta\beta_c(m)}{h_c(m)}. \tag{48}$$

For $M \sim \mathrm{Bin}(c, \tau)$, Jensen and Markov imply $P\{\alpha_c(M) < -1/4\} \le 16B$; on the complementary event $h_c \ge 1/2$.

If $H_{\mathrm{sq}} > 0$, the law with density g^2/H_{sq} relative to P_E is exchangeable. Let ν_s be its total-count probabilities. Then

$$v\Delta\beta_c(m) = H_{\mathrm{sq}} \sum_s \nu_s \mathcal{R}_s(c, m)(z_s - \tau). \tag{49}$$

If $H_{\mathrm{sq}} = 0$, both sides are zero without defining ν. Lemma 9 bounds the absolute value by $BJ_{c,t}$ on $|M - c\tau| \le t\sqrt{c}$. Consequently, on the intersection of the central-window and good-denominator events,

$$|p_c(M) - \tau| \le 4v|\Delta\alpha_c(M)| + 2BJ_{c,t}. \tag{50}$$

The probability of their complement is at most $16B + 2e^{-2t^2}$. Bound TV there by one, apply (32) on the intersection, and use Cauchy–Schwarz with (21). This yields

$$\mathcal{D}_{n,c,d} \le 16B + 2e^{-2t^2} + 2\sqrt{2Bd/(n-c)} + \sqrt{2d/v}\, BJ_{c,t}, \tag{51}$$

which is (22). The expectation is under the full target-prefix law, with the exceptional events charged through their probabilities. The bound holds for every next-batch size $1 \le d \le n - c$.

C.4 FULL-OUTPUT HYBRIDS AND FIXED-DEPTH LIMITS

Fix deterministic groups with sizes $d_j = c_j - c_{j-1}$. Generate representatives using their exact common law. For $0 \le j \le r$, define a full-output law H_j in which the first j mismatch groups use target draws and all later groups use q_R at that hybrid's actual prefix. Thus $H_0 = Q_R^{\mathrm{geom}}$ and $H_r = P_n$.

For the pair H_{j-1}, H_j, the first $j-1$ groups have the same target law. Conditional on their prefix, only the next product kernel differs; after it, both laws apply the same continuation kernel as a function of the prefix. Conditional TV and contraction imply $\mathrm{TV}(H_{j-1}, H_j) \le \mathcal{D}_{n,c_{j-1},d_j}$. Summing proves (19). These full-output hybrids explain why a bound averaged over target prefixes controls the error of the actual policy, whose prefixes are generated by the model heads.

For the schedule (11) with fixed $r \ge 2$, every c_{j-1}, $j \ge 2$, eventually meets the conditions of Lemma 9 for fixed $t > 0$. Moreover,

$$\sqrt{c_1} = n^{1/(2r)}(1 + o(1)), \qquad \sqrt{d_j/c_{j-1}} \le n^{1/(2r)}(1 + o(1)), \qquad d_j/(n - c_{j-1}) \le 1. \tag{52}$$

All square-root-in-B terms tend to zero if $B \to 0$. For a budget sequence with $n^{1/(2r)}B_n \to u \in [0, \infty)$, the finite bounds therefore give

$$\limsup_{n\to\infty} \sup_{R_E \in \mathcal{C}_n(B_n)} \mathrm{TV}(P_n, Q_R^{\mathrm{geom}}) \le 2(r-1)e^{-2t^2} + C_r(t)u, \tag{53}$$

where a valid coefficient is

$$C_r(t) = \frac{3}{4\sqrt{2v}} + 2(r-1)(t+1)e^{t^2/v}\sqrt{2/v}. \tag{54}$$

For $u = 0$, send $t \to \infty$ after taking the limsup. The policy does not depend on t. For fixed accuracy, choose t with $2(r-1)e^{-2t^2} < \varepsilon/4$ and take $c_{r+1,\varepsilon} = \varepsilon/(4C_r(t))$. The limiting bound has strict slack below ε, giving an eventual uniform guarantee. For $r = 1$, (23) with $d = n$ proves both conclusions directly. All constants here concern fixed r.

D COMBINING THE BOUNDS AND CHOOSING GROWING PARAMETERS

Fix $r = K - 1$ and take $\Delta = 1/4$. For a positive budget B_n, set

$$w_n = \min\{1/16, 1 - e^{-B_n}\}, \qquad u_n = nw_n^{2r}. \tag{55}$$

Then $D(q_R) \le B_n$ and $w_n \ge \min\{B_n, 1\}/16$. If $n^{1/(2r)}B_n \to \infty$, then $u_n \to \infty$. Choose eventually $\lambda_n = u_n^{-1/(4r)}$. The failure charge $r\lambda_n$ vanishes, while the exponent in (13) is $\sqrt{u_n}/(32r^{2r}) \to \infty$. This proves (10). Choosing λ_n fixes the component locations and thus a single model in each dimension.

For the upper bound on the critical budget, put $g_0 = 1 - \varepsilon$ and choose

$$\lambda = \frac{g_0}{2r}, \qquad C_{r+1,\varepsilon} = \frac{r}{\lambda}\left[64\log(8/g_0)\right]^{1/(2r)}. \tag{56}$$

At $B_n = C_{r+1,\varepsilon}n^{-1/(2r)}$, eventually $w_n = 1 - e^{-B_n}$ and $nw_n^{2r} \ge C_{r+1,\varepsilon}^{2r}/2$. The exponent is at least $\log(8/g_0)$, so

$$\mathcal{E}_n(r+1, B_n) \ge 1 - g_0/2 - g_0/4 = \varepsilon + g_0/4 > \varepsilon.$$

Budget monotonicity gives $B^*_{n,r+1}(\varepsilon) \le B_n$, without requiring attainment or continuity. Appendix C gives the other inequality and (9). This proves Theorem 2.

For Corollary 4, use (24), whose parameters are legal for all sufficiently large n. Directly,

$$\begin{aligned} r_n\lambda_n &= 1/\ell_n, \\ \log\frac{r_n}{w_n\lambda_n} &= 2\log r_n + (\zeta+1)\log\ell_n = (\zeta+3)\log\ell_n - 2\log\log\ell_n + O_\zeta(1), \\ \log\left[n\Delta_n^2(w_n\lambda_n/r_n)^{2r_n}\right] &= \tfrac{1}{2}\log n + o(\log n) \longrightarrow \infty. \end{aligned} \tag{57}$$

Theorem 3 therefore gives TV tending to one against every policy under the stated cap. Its KL, joint TV, and uniform-head bounds tend to zero. Finally, for every π, $\mathrm{TV}(R_n, Q_R^\pi) \ge \mathrm{TV}(P_n, Q_R^\pi) - w_n$. Taking an infimum proves the own-model assertion. The finite lower inequality retains its explicit dependence on the depth, allowing the choice $K = K_n$ here.

Table 2: Fixed-round critical scales in the fair finite-exchangeable class. Cumulative mismatch counts omit the initial representative round and integer rounding. Constants may depend on K and the desired TV accuracy.

Total rounds K	Cumulative mismatch counts	Critical total loss B^*
2	n	$\Theta(n^{-1/2})$
3	$n^{1/2}, n$	$\Theta(n^{-1/4})$
4	$n^{1/3}, n^{2/3}, n$	$\Theta(n^{-1/6})$
Fixed K	$n^{1/(K-1)}, \dots, n$	$\Theta(n^{-1/[2(K-1)]})$

E Proof for arbitrary heterogeneous representative biases

Work with the model of Theorem 6, with the population loss evaluated under $P_{\boldsymbol{a},n}$, and put $v_{A,i} = a_i(1-a_i)$ and $\theta_j = \tau + \delta_j$. All component parameters lie in $[1/4, 1/2]$. We first prove the lower bound, accounting for the information in B-only observations, then show that the geometric upper is independent of the representative biases.

E.1 Complete-context likelihood and nonlinear reverse heads

For a mismatch parameter s, define

$$r_{s,i} = a_i + (1-2a_i)s, \qquad f_{s,i}(1) = \frac{(1-a_i)s}{r_{s,i}}, \qquad f_{s,i}(0) = \frac{a_i s}{1-r_{s,i}}. \tag{58}$$

Here $r_{s,i}$ is the B_i marginal, and $f_{s,i}(x)$ is the mismatch probability when completing A_i given $B_i = x$. At any partial context h, its full component likelihood relative to the teacher is

$$\Lambda_s(h) = \prod_{i:\,\text{completed}} \frac{\mathrm{Ber}(s)(E_i)}{\mathrm{Ber}(\tau)(E_i)} \prod_{i:\,B\text{-only}} \frac{\mathrm{Ber}(r_{s,i})(B_i)}{\mathrm{Ber}(r_{\tau,i})(B_i)}. \tag{59}$$

An A-only pair contributes no factor. Before normalization, the teacher has weight $1-w$ and alternative j has weight $b\Lambda_{\theta_j}$. In physical round t, write p_t for the model posterior mean at the round-entry context and abbreviate it by p below. This posterior is held fixed throughout the batch, including when we expose its independent blocks sequentially for analysis. In general, $f_{s,i}(x)$ is nonlinear in s, so the reverse-completion head must average the component conditionals: it is $\mathbb{E}_{\mathrm{post}} f_{s,i}(x)$, not $f_{p,i}(x)$.

For $1 \le j < K$, set

$$Z_j(h) = \frac{1 - w + b\sum_{\ell=j}^{K-1} \Lambda_{\theta_\ell}(h)}{\Lambda_{\theta_j}(h)}. \tag{60}$$

The same submixture calculation as (35) gives

$$Z_j \le 1/\lambda \quad \Longrightarrow \quad p - \tau \ge b\lambda\delta_j = 9\delta_{j+1}. \tag{61}$$

Since $r_{s,i}, 1-r_{s,i} \in [1/4, 3/4]$, differentiation yields

$$\frac{16}{9} v_{A,i} \le \partial_s f_{s,i}(x) \le 16 v_{A,i}. \tag{62}$$

On the event in (61),

$$\mathbb{E}_{\mathrm{post}} f_{s,i}(x) - f_{\tau,i}(x) \ge \tfrac{16}{9} v_{A,i}(p-\tau) \ge 16 v_{A,i}\delta_{j+1} \ge f_{\theta_{j+1},i}(x) - f_{\tau,i}(x).$$

The pair-specific variance cancels in this comparison, giving the same factor $1/9$ for every positive bias, uniformly as $\min_i v_{A,i}$ approaches zero.

A B completion given A has mismatch mean p. A B-only birth has mean $r_{p,i}$, ordered in the direction from $r_{\tau,i}$ toward larger s; that direction reverses when $a_i > 1/2$, which is allowed. An A birth has common mean a_i. Finally, independently co-sampling a fresh pair gives

$$\mu_{\mathrm{co},i} = a_i + (1-2a_i) r_{p,i}, \qquad \mu_{\mathrm{co},i} - p = v_{A,i}[1 - 4(p-\tau)] \ge 0. \tag{63}$$

Thus co-sampling dominates the current posterior mean throughout the allowed parameter range, even when $\Delta > \min_i v_{A,i}$.

E.2 Warm start and stopping

At the empty context,

$$p_0 - \tau = b \sum_{j=0}^{K-1} \delta_j \geq b\Delta \geq \delta_1. \tag{64}$$

There are no open pairs yet. Every first-round B birth is therefore beyond component θ_j in its ordered direction for each $j \geq 1$, and every co-sample has mismatch mean at least $p_0 \geq \theta_j$. The highest component provides a warm start rather than serving as the first certificate denominator.

For Bernoulli channel parameters v_s, v_j and actual mean μ, the expected likelihood multiplier is

$$1 + \frac{(\mu - v_j)(v_s - v_j)}{v_j(1 - v_j)}. \tag{65}$$

It is at most one when the actual mean is beyond v_j in the direction away from v_s. To apply this identity to $\Lambda_s/\Lambda_{\theta_j}$, the physical factors are: one for an A birth; the $r_{s,i}$ Bernoulli factor for a B birth; the s mismatch factor for a B completion; the $f_{s,i}(x)$ mismatch factor for an A completion; and the s mismatch factor for a fresh co-sampled pair. At a reverse completion, the conditional likelihood factor replaces the earlier B-only marginal factor with the completed-pair likelihood.

Let G_j be $Z_j \leq 1/\lambda$ at the end of physical round j, for $1 \leq j < K$. For each destination j, update the ratios through rounds $t \leq j$ while all previous certificates pass; include the first failing batch and then freeze. The first round is valid by (64). In an active later round t, (61) and (62) put every completion or birth beyond component θ_t, and hence beyond θ_j. Equation (63) handles co-samples. Grouping each fresh pair as one block, fresh block factors are independent conditional on the pre-round history and action. Equations (65) and the stopping rule give

$$\mathbb{E}\widetilde{Z}_{j,j} \leq Z_j(\varnothing) = 1 - bj \leq 1, \qquad \mathbb{P}\left(\bigcap_{j=1}^{K-1} G_j\right) \geq [1 - (K-1)\lambda]_+. \tag{66}$$

The proof is Markov's inequality on disjoint first failures exactly as in Appendix B, and includes arbitrary privately randomized actions.

E.3 An orientation-independent potential

A B-first mismatch completion need not itself have mean at least τ. We instead use the following increments for newly touched pairs:

Physical observation	Increment V
Fresh A_i only	0
Fresh $B_i = x$ only	$f_{\tau,i}(x) - \tau$
Complete B_i given A_i	$E_i - \tau$
Complete A_i given $B_i = x$	$E_i - f_{\tau,i}(x)$
Fresh co-sampled pair	$E_i - \tau$

For each pair these increments telescope to $E_i - \tau$, irrespective of orientation. Their conditional range widths are at most one. We expose a co-sample as one whole block so that its mean averages over both endpoints.

The birth drift of a B-first pair is

$$\mathbb{E}[f_{\tau,i}(B_i) - \tau \mid \text{pre-block information}] = \chi_i(p - \tau), \qquad \chi_i = \frac{3(1 - 4v_{A,i})}{3 + 4v_{A,i}} \geq 0. \tag{67}$$

This follows by subtracting the zero teacher expectation, using $f_{\tau,i}(1) - f_{\tau,i}(0) = v(1 - 2a_i)/(r_{\tau,i}(1 - r_{\tau,i}))$ and $r_{\tau,i}(1 - r_{\tau,i}) = 3/16 + v_{A,i}/4$.

On the common good event in (66), the posterior excess at every birth is at least δ_K: this follows from (64) in round one and the preceding certificate thereafter. A B-first pair therefore contributes

at least $\chi_i \delta_K$ at birth and at least $(16/9) v_{A,i} \delta_K$ at completion. We bound each contribution at its own predictable context and combine them using

$$\chi_i + \frac{16}{9} v_{A,i} - \frac{4}{9} = \frac{(1 - 4v_{A,i})(15 - 16v_{A,i})}{9(3 + 4v_{A,i})} \geq 0. \tag{68}$$

An A-first completion contributes at least δ_K, and a co-sampled pair does the same by (63). Thus the sum of all predictable increments is at least $(4/9) n \delta_K$ on the good event.

To justify concentration, expose blocks in a fixed order within each selected batch. Let $\mathcal{F}_\ell^-$ contain all earlier observations, the private action choice, and the identity of the next block; let $\mathcal{F}_\ell^+$ also contain that block's complete outcome. Then $\mathcal{F}_{\ell-1}^+ \subseteq \mathcal{F}_\ell^- \subseteq \mathcal{F}_\ell^+$. The frozen heads and fresh independence preserve each calculated mean $\mu_\ell = \mathbb{E}[V_\ell \mid \mathcal{F}_\ell^-]$ after earlier blocks of the same batch. Pad by zero increments to $2n$ blocks. Conditional Hoeffding and a tower step give

$$\mathbb{P}\left\{ \sum_{\ell=1}^{2n} (V_\ell - \mu_\ell) \leq -t \right\} \leq e^{-t^2/n} \quad (t \geq 0), \qquad \sum_\ell V_\ell = \sum_i E_i - n\tau. \tag{69}$$

The event $\mathcal{A}_{\text{test}} = \{\sum_i E_i \geq n\tau + (2/9) n \delta_K\}$ has target probability at most $e^{-(8/81) n \delta_K^2}$. On the good event, its complement forces a downward deviation of at least $(2/9) n \delta_K$ in (69). Hence

$$Q_R^\pi(\mathcal{A}_{\text{test}}) - P_{\boldsymbol{a},n}(\mathcal{A}_{\text{test}}) \geq [1 - (K-1)\lambda]_+ - 2e^{-(4/81) n \delta_K^2},$$

which proves (29). All constants are uniform over $a_i \in (0,1)$.

The KL and joint-TV bounds follow from the mixture form. Fresh A heads are exact; fresh B and open-B errors are at most Δ; (62) bounds open-A errors by $16 v_{A,i} \Delta \leq 4\Delta$. This completes Theorem 6. For its growing-depth example, $\log[n\Delta^2 (w\lambda/(9K))^{2K}] = \frac{1}{2} \log n + o(\log n)$ and $(K-1)\lambda \leq 1/\ell_n$. The own-model conclusion again follows by subtracting w from the target-TV bound.

E.4 The geometric upper is independent of the bias array

Fix any $\boldsymbol{a} \in (0,1)^n$, and let R_E be any positive finite-exchangeable law with $\mathrm{KL}(P_E \| R_E) \leq B$. Both joint laws use the same product representative law $P_{\boldsymbol{a},A} = \bigotimes_i \mathrm{Ber}(a_i)$, independently of the mismatches. Run the schedule (11), first revealing all representatives. This first round draws A exactly from $P_{\boldsymbol{a},A}$. Given A and any completed mismatch prefix, an original B_i head draws $E_i = A_i \oplus B_i$ from the corresponding conditional of R_E. That conditional does not depend on the values of A.

Consequently, if $Q_{E,R}^{\text{geom}}$ denotes the mismatch output law of the geometric schedule, it is the same law as in the fair case, and the bijection $(A,B) \leftrightarrow (A,E)$ gives

$$\mathrm{TV}(P_{\boldsymbol{a},n}, Q_R^{\text{geom}}) = \mathrm{TV}(P_{\boldsymbol{a},A} \otimes P_E, P_{\boldsymbol{a},A} \otimes Q_{E,R}^{\text{geom}}) = \mathrm{TV}(P_E, Q_{E,R}^{\text{geom}}). \tag{70}$$

The loss identity also gives $D(q_R) = \mathrm{KL}(P_E \| R_E)$. Every finite upper estimate in Appendix C therefore holds with the same constants, uniformly over bias arrays. In particular, the sufficient critical scale remains $n^{-1/[2(K-1)]}$. Taking a supremum over all positive bias arrays cannot worsen this upper, while the fair array supplies the matching lower. The optimal exponent for a prescribed non-fair array remains open.

F A finite-alphabet structural extension

We next extend the argument to finite alphabets with nonuniform marginals. Here singleton marginals are component-invariant, whereas Appendix E allows informative singletons. For finite alphabets at pair i, take positive marginals μ_i, ν_i and an interaction matrix H_i with

$$\sum_y \nu_i(y) H_i(x,y) = 0, \qquad \sum_x \mu_i(x) H_i(x,y) = 0. \tag{71}$$

Fix $\theta_* < 0$ and assume $1 + \theta H_i(x, y) > 0$ for every $\theta \in [\theta_*, 0]$. Define

$$P_{i,\theta}(x, y) = \mu_i(x)\nu_i(y)(1 + \theta H_i(x, y)), \quad P_\theta = \bigotimes_i P_{i,\theta}, \quad P = P_{\theta_*}. \tag{72}$$

In this extension, the population loss uses the target $P = P_{\theta_*}$ in (3); thus $D(q_R) = \mathrm{KL}(P\|R)$ by Appendix A. Let

$$\phi_i(x, y) = \frac{H_i(x, y)}{1 + \theta_* H_i(x, y)},$$
$$J_i^A(x) = \sum_y \nu_i(y) \frac{H_i(x, y)^2}{1 + \theta_* H_i(x, y)}, \qquad J_i^B(y) = \sum_x \mu_i(x) \frac{H_i(x, y)^2}{1 + \theta_* H_i(x, y)}. \tag{73}$$

Assume all row and column energies are at least $\kappa > 0$, and each ϕ_i has range width at most $C < \infty$. Different pairs may have different matrices and marginals.

Theorem 10 (Fixed-marginal interactions)**.** *For $n, r \geq 1$, $0 < w \leq 1/8$, $0 < \lambda \leq 1$, and $0 < \Delta \leq -\theta_*$, set $b = w/r$, $\delta_j = \Delta(b\lambda)^j$, $\theta_j = \theta_* + \delta_j$, and $R = (1 - w)P + b\sum_{j=0}^{r-1} P_{\theta_j}$. Under Definition 1, with the corresponding finite alphabets,*

$$\inf_{\pi \in \Pi_{r+1}} \mathrm{TV}(P, Q_R^\pi) \geq [1 - r\lambda]_+ - 2\exp\left[-\frac{n\kappa^2\Delta^2}{2C^2}\left(\frac{w\lambda}{r}\right)^{2r}\right]. \tag{74}$$

Also $D(q_R) = \mathrm{KL}(P\|R) \leq -\log(1 - w)$ and $\mathrm{TV}(P, R) \leq w$.

Proof. Singleton likelihood factors equal one. At a completed pair, the component likelihood factor is $(1 + \theta_j H_i)/(1 + \theta_* H_i)$. The model posterior mean $\bar{\theta}$ at the physical round's entry defines each open head by the same affine formula as (72); it remains fixed throughout the batch. Fresh co-sampling produces the product law $P_{i,0}$. Define Z_j from the context likelihoods as in (15). The same mixture argument gives $Z_j \leq 1/\lambda \Rightarrow \bar{\theta} \geq \theta_{j+1}$ at each certificate context.

For a row with $h(y) = H_i(x, y)$ and parameters $s \leq t \leq u$, row centering yields

$$\mathbb{E}_{Y \sim \nu_i(1+uh)} \frac{1 + sh(Y)}{1 + th(Y)} = 1 + (u - t)(s - t)\sum_y \nu_i(y) \frac{h(y)^2}{1 + th(y)} \leq 1. \tag{75}$$

The same holds in each column. This orders likelihood ratios without ordering the observed symbols. The first-round completion parameter is 0, which dominates every θ_j. An active later round has parameter at least the appropriate next level. The stopped-ratio construction of Appendix B therefore gives a good-history event of probability at least $[1 - r\lambda]_+$. First failures are included before freezing, and all conditioning includes the chosen private action.

The score has teacher conditional mean zero in either orientation, and

$$\mathbb{E}_{P_{i,u}}[\phi_i \mid A_i = x] = (u - \theta_*)J_i^A(x), \qquad \mathbb{E}_{P_{i,u}}[\phi_i \mid B_i = y] = (u - \theta_*)J_i^B(y). \tag{76}$$

To check this, expand $1 + uH_i = (1 + \theta_* H_i) + (u - \theta_*)H_i$ and use (71). On good history, each completed pair thus has predictable score mean at least $\kappa\delta_r$. For co-samples, expose one endpoint as auxiliary information before the score; the other is still sampled from its independent marginal, corresponding to $u = 0$.

Expose completed scores in physical completion order. Use pre-score sigma-fields that include the selected action, intervening singleton outcomes, and the auxiliary co-sample endpoint, followed by post-score sigma-fields. They are nested as in Appendix B. Each centered score has conditional range width at most C. For $t \geq 0$, the unconditional lower-tail bound is $\exp(-2t^2/(nC^2))$. Under the teacher, the n scores are independent, centered, and have the same range bound. The common terminal event $\{\sum_i \phi_i(A_i, B_i) \geq n\kappa\delta_r/2\}$ therefore proves (74). It depends on neither orientation choices nor hidden seeds. Domination and the chain rule prove the loss and joint-TV statements. □

At every context the head-TV error is at most $M\Delta$, where

$$M = \frac{1}{2}\max\left\{\max_{i,x}\sum_y \nu_i(y)|H_i(x, y)|, \max_{i,y}\sum_x \mu_i(x)|H_i(x, y)|\right\}. \tag{77}$$

Indeed singleton heads agree, and the open heads are affine in a posterior parameter between θ_* and $\theta_* + \Delta$. To take dimension-dependent limits, fix $\theta_* < 0$ and require $\kappa \geq \kappa_0 > 0$, $C \leq C_0 < \infty$, and $M \leq M_0 < \infty$ uniformly in n. For fixed depth, use any fixed $0 < \Delta_0 \leq -\theta_*$ in place of $1/4$ in Appendix D; this changes only constants and gives the same supercritical lower exponent. For growing depth, take $\Delta_n = 1/\log(n+1)$, which is eventually admissible, and use the remaining parameters in (24). The finite bound then gives vanishing loss and head error with a near-logarithmic depth obstruction. No matching upper over this enlarged class is established.

Explicit nonuniform three-valued example. Take

$$\mu = (1/2, 1/3, 1/6), \quad \nu = (1/2, 1/4, 1/4), \quad H = \begin{pmatrix} 3/5 & -2/5 & -4/5 \\ -21/40 & 7/20 & 7/10 \\ -3/4 & 1/2 & 1 \end{pmatrix}, \quad \theta_* = -1/2. \tag{78}$$

Both weighted marginal means are zero, and teacher density factors lie in $[1/2, 7/5]$. The row energies are $17/42, 14504/43329, 26/33$ and the column energies are $3096/7777, 17/99, 74/91$. Thus $\kappa = 17/99$; the score ranges from $-4/7$ to 2, giving $C = 18/7$; and $M = 2/5$. Each row and column has three distinct interaction values. For distinct component parameters, the head likelihood ratio is strictly monotone in H. Thus each completion head has three distinct likelihood-ratio values, which no deterministic binary statistic can preserve. This constructed family gives an explicit three-valued instance of the structural assumptions.

Nonexchangeable interaction strengths. The structural conditions also allow dependence to vary across pairs, without changing what singleton observations reveal. Keep μ, ν, H, θ_* from (78), but choose any deterministic strengths

$$\gamma_i \in [1/2, 1], \qquad H_i = \gamma_i H, \qquad P_{i,\theta}(x, y) = \mu(x)\nu(y)(1 + \theta\gamma_i H(x, y)). \tag{79}$$

The weighted centering identities remain valid, so the singleton laws are still μ and ν in every component. The density factors lie in $[1/2, 7/5]$ throughout $\theta \in [-1/2, 0]$. Theorem 10 applies uniformly over these strength arrays with

$$\kappa = \frac{232}{5621}, \qquad C = \frac{18}{7}, \qquad M \leq \frac{2}{5}. \tag{80}$$

For the energy bound, each summand is nondecreasing in γ because

$$\frac{d}{d\gamma} \frac{\gamma^2 h^2}{1 - \gamma h/2} = \frac{\gamma h^2 (2 - \gamma h/2)}{(1 - \gamma h/2)^2} \geq 0.$$

At $\gamma = 1/2$, the smallest of the three row and three column energies is the second-column energy $232/5621$. Moreover, $\gamma h/(1 - \gamma h/2) \in [-4/7, 2]$, and scaling by $\gamma \leq 1$ cannot increase (77).

This family is nonexchangeable whenever the strengths are not all equal. Indeed, direct averaging under the teacher gives the observable distinction

$$\mathbb{E}_P[H(A_i, B_i)] = -\frac{\gamma_i}{2} \sum_{x,y} \mu(x)\nu(y)H(x, y)^2 = -\frac{247}{1280}\gamma_i. \tag{81}$$

For example, $\gamma_i = 1/2 + (i-1)/(2(n-1))$ for $n \geq 2$ gives different dependence strengths at every pair. The coherent mixture R is likewise nonexchangeable: writing Θ for its component parameter, $\mathbb{E}_R H(A_i, B_i) = \gamma_i(247/640)\mathbb{E}\Theta$, where $\mathbb{E}\Theta < 0$ because the teacher component has positive mass. The target still has independent pairs and is sampled in two rounds by revealing all A_i, then all B_i. This instance of Theorem 10 shows how component-invariant singleton information, uniformly positive conditional score energy, and a common ordering of component parameters support the lower bound without exchangeability. A matching upper for this nonexchangeable class remains open.

G FINITE-DIMENSIONAL EVALUATION AND CERTIFICATION

We evaluate the original-head sampling interface on finite distributions for which the complete output law can be computed. For a fixed oracle R, write

$$d_K(R) = \inf_{\pi \in \Pi_K} \mathrm{TV}(P, Q_R^\pi).$$

The policy class includes full-history decisions and private randomization; every selected coordinate is sampled independently from its conditional head at the pre-round context and committed permanently. We compute rigorous certificates $L_K \le d_K(R) \le U_K$ using directed rational arithmetic. In particular, $L_K - \mathrm{TV}(P, R) > 0$ certifies additional execution error relative to the oracle's joint law, and $L_K - U_{K+1} > 0$ certifies a depth separation for the *same* oracle. The serial reference $\mathrm{TV}(P, R)$ is feasible at unrestricted depth, and a different execution may improve upon it.

G.1 WITNESS SELECTION AND BUDGET CALIBRATION

In the fair experiment, A_i are independent fair bits, $B_i = A_i \oplus E_i$, and the teacher mismatch law is $P_E = \mathrm{Ber}(\tau)^{\otimes n}$ with $\tau = 1/4$. Candidate oracle mismatch laws are positive mixtures $R_E = \sum_\ell w_\ell \mathrm{Ber}(p_\ell)^{\otimes n}$. We use $n \in \{8, 12, 16\}$ and joint-KL budgets $B \in \{0.02, 0.05, 0.1\}$. At $n = 8$, the candidate grid contains the following eight families:

- Three single-alternative mixtures with weights $(1 - w, w)$ and rates $(\tau, \tau + \delta)$, where $\delta \in \{1/16, 1/8, 1/4\}$.
- A two-level ladder with weights $(1 - w, w/2, w/2)$ and rates $(\tau, \tau + 1/4, \tau + w/40)$, with $w \le 1/8$.
- Four relaxed two-level mixtures with the same weights and rates $(\tau, \tau + 1/4, \tau + \delta)$, where $\delta \in \{1/64, 1/32, 1/16, 1/8\}$ and $w \le 1/2$.

The single-alternative families also have $w \le 1/2$. At each larger dimension, we retain the three single-alternative families, the ladder, and the relaxed spacing with the largest three-round excess at $n = 8$ for that budget. Its weight is recalibrated at the new dimension. This gives $24 + 15 + 15 = 54$ candidate configurations.

The total alternative weight w is calibrated numerically toward the actual joint-KL target $B(1 - 10^{-6})$, then rounded down to $2^{-24}\mathbb{Z}$. If a family's weight cap prevents reaching the target, we retain and report its smaller actual KL, so that budget comparisons can account for the remaining slack. All subsequent probability calculations use the resulting exact rational parameters. A candidate is certified feasible only when a rigorous upper enclosure of $\mathrm{KL}(P\|R)$ is at most B. Since every component rate lies in $[1/4, 1/2]$, all fair-model conditional-head errors are at most $1/4$; we additionally check every distinct completion context directly.

For each (n, B) we select the largest numerical value of $d_3(R) - \mathrm{TV}(P, R)$ in the candidate grid, then keep that oracle fixed for $K = 2, 3, 4$. This finite-grid search supplies existence witnesses: certification of any feasible selected oracle gives a valid finite-dimensional lower bound. The outer supremum over R remains unoptimized.

Table 3 specifies every selected oracle exactly. The single and relaxed families belong to the general exchangeable class in Theorem 2. Their weights can exceed the restriction imposed on the explicit ladder in Theorem 3, so these computations evaluate execution costs in the larger exchangeable class.

The scale of the explicit ladder guarantee. For comparison, in Theorem 3 take $K = 3$, $r = 2$, $w = \lambda = 1/8$, and $\Delta = 1/4$. Then $\delta_r = 2^{-16}$ and the displayed lower bound is $3/4 - 2\exp(-n/2^{33})$. It guarantees TV at least $1/2$ for $n \ge \lceil 2^{33} \log 8 \rceil = 17{,}862{,}266{,}831$ pairs. This is a sufficient threshold for one unoptimized parameter choice. The certificates below establish smaller finite-dimensional costs for selected models in the same exchangeable class. Their scope is finite-instance execution error; the explicit ladder's constants and asymptotic exponent are analyzed in the theorem. Appendix H separately examines error amplification and pilot behavior.

Table 3: Exact selected mismatch mixtures. Write $w = M/2^{24}$. A single alternative has weights $(1-w, w)$ at $(\tau, \tau+\delta)$; a relaxed mixture has weights $(1-w, w/2, w/2)$ at $(\tau, 1/2, \tau+\delta)$. Here n is the number of pairs, so the output dimension is $2n$.

n	B	M	δ	Family
8	0.02	1822566	$1/4$	Single
8	0.05	3269660	$1/4$	Single
8	0.10	5095329	$1/4$	Single
12	0.02	2617140	$1/64$	Relaxed
12	0.05	4860030	$1/64$	Relaxed
12	0.10	7769258	$1/64$	Relaxed
16	0.02	2785943	$1/8$	Single
16	0.05	4778537	$1/8$	Single
16	0.10	7166016	$1/8$	Single

G.2 Exact state reduction and the all-policy optimization

For fair representatives, let $s = (c, m, o)$ record the numbers of completed pairs, completed mismatches, and pairs with exactly one visible endpoint. Singleton values and orientations do not inform the mismatch component. The conditional mismatch probability when completing an open pair is

$$p(c,m) = \frac{\sum_\ell w_\ell p_\ell^{m+1}(1-p_\ell)^{c-m}}{\sum_\ell w_\ell p_\ell^{m}(1-p_\ell)^{c-m}}.$$

An aggregate action (u, h, z) completes u open pairs, samples both coordinates of h fresh pairs, and reveals one coordinate of z fresh pairs. Its entries are nonnegative integers satisfying $u \le o$ and $h + z \le n - c - o$. Conditional on the *entire labeled history*, the next state satisfies

$$c' = c + u + h, \qquad o' = o - u + z, \qquad m' = m + X + Y,$$
$$X \sim \mathrm{Bin}(u, p(c,m)), \qquad Y \sim \mathrm{Bin}(h, 1/2),$$

with independent X, Y. The last round forces $(u, h, z) = (o, n - c - o, 0)$. These transition laws hold for every labeled history, regardless of its generating law.

Nonnegative state–action occupation measures, with the usual initial-mass and flow-conservation constraints, form a finite linear program. Its objective is half the ℓ_1 distance between the terminal mismatch-count law and $\mathrm{Bin}(n, \tau)$, linearized with absolute-value variables. The LP evaluates TV after marginalizing the policy's private randomness, as required by the delivered-output objective. Every labeled policy induces admissible aggregate flows. Conversely, an aggregate policy can be lifted by uniformly selecting eligible pair labels and fresh singleton orientations. The lifted law is invariant under pair permutations and simultaneous flips of both endpoints within each pair. These symmetries are transitive on outputs with a fixed mismatch count, so its count TV equals its full $2n$-coordinate TV. Thus the reduced LP has optimum $d_K(R)$. For an arbitrary nonsymmetric policy, count TV provides only a lower bound on full-output TV.

G.3 Directed rational certificates

Let C denote the terminal statistic: the mismatch count in the fair experiment, or the four-type histogram in the biased experiment below. Write p_{C} for its target law. The floating-point LP supplies a terminal test and a candidate policy. For the lower certificate, clip the test to $[0, 1]$ and round it down to a dyadic function φ on the grid $S^{-1} = 2^{-40}$. This bounded test is valid independently of the numerical dual's accuracy. Starting from terminal integer values $S\varphi$, compute

$$V_t(s) = \min_{a \in \mathcal{A}_t(s)} \left\lfloor \sum_{s'} \mathsf{P}(s' \mid s, a) V_{t-1}(s') \right\rfloor, \qquad L_K = \max\{0, V_K(s_0)/S - \mathbb{E}_{p_{\mathsf{C}}}\varphi\},$$

where the transition kernel P and teacher expectation are exact rationals. The final-action restriction is included in $\mathcal{A}_t$. Induction bounds the downward Bellman error by t/S. Since the minimum

includes every legal action and the aggregate kernel holds after every labeled history, L_K lower-bounds full-output TV for every policy in Π_K, including history-dependent randomized policies: $\varphi \circ \mathsf{C}$ is a valid bounded test of the full output, without any symmetry assumption on that policy.

For the upper certificate, interpret the recovered nonnegative action weights as exact decimal rationals and normalize them at each state. Unspecified states receive an empty action before the last round and forced completion in the last round. Forward propagation rounds each edge contribution downward on the grid 2^{-80}. Its terminal subprobability q_{C}^- is componentwise below the true terminal-statistic law q_{C} of this explicitly defined policy. With $e = 1 - \sum_z q_{\mathsf{C}}^-(z)$, the triangle inequality gives

$$U_K = \min\left\{1, \frac{\sum_z |p_{\mathsf{C}}(z) - q_{\mathsf{C}}^-(z)| + e}{2}\right\} \geq \mathrm{TV}(p_{\mathsf{C}}, q_{\mathsf{C}}).$$

For the symmetric lift Q, $\mathrm{TV}(p_{\mathsf{C}}, q_{\mathsf{C}}) = \mathrm{TV}(P, Q) \geq d_K(R)$, making this a full-output upper certificate. Exact rational endpoints are authoritative; displayed decimal intervals enclose them by directed rounding. Their widths describe arithmetic precision.

Model TV is computed from exact count masses. To certify KL, range-reduce each positive rational likelihood ratio as $x = 2^k y$, $1 \leq y < 2$, and use

$$\log y = 2\sum_{j=0}^{47} \frac{z^{2j+1}}{2j+1} + \rho, \qquad z = \frac{y-1}{y+1}, \qquad 0 \leq \rho \leq \frac{2z^{97}}{97(1-z^2)}.$$

Termwise directed rounding on a 2^{-120} grid, the analogous bounds for $\log 2$, and exact averaging under P produce a rigorous KL interval. Negative k reverses the endpoints when multiplying the $\log 2$ interval.

G.4 Finite-depth results

Table 4 reports all selected oracles and depths. Every three-round excess is positive, and each oracle satisfies $L_3 > U_4$. The certificate widths before decimal display rounding are below 6.1×10^{-10} over all 27 cases; for $K = 3$ alone they are below 9.3×10^{-13}. The selected oracles differ across dimensions and budgets, and the larger dimensions use a reduced candidate grid, so the comparisons describe individual instances rather than a dimension exponent. At $n = 12$, all three four-round upper certificates are below the model TV: a legal execution can improve on the serial law R.

Table 4: All nine fair oracles. Model TV and d_K are approximate displays; both endpoints of each optimum enclosure round to the displayed value. KL^+ is a certified upper bound rounded upward. The last column is a downward-rounded bound on $G_3 = L_3 - \mathrm{TV}(P_n, R)$, so it need not equal the difference of the rounded columns. Parameters are in Table 3.

n	B	KL^+	$\mathrm{TV}(P_n, R)$	$d_2(R)$	$d_3(R)$	$d_4(R)$	$G_3 \geq$
8	.02	.01999997	.05801152	.06857509	.06259081	.05936764	.00457929
8	.05	.04999994	.10407192	.12351996	.11316915	.10733143	.00909722
8	.10	.09999988	.16218221	.19201755	.17691617	.16803951	.01473395
12	.02	.01999998	.05322553	.06478036	.05596105	.05226142	.00273551
12	.05	.04999995	.09883984	.12048458	.10444571	.09770445	.00560587
12	.10	.09999990	.15800565	.19138687	.16710498	.15686703	.00909932
16	.02	.01999998	.06768111	.07539204	.07160591	.06921703	.00392480
16	.05	.04999995	.11608876	.12916895	.12284334	.11887869	.00675457
16	.10	.09999989	.17408967	.19209434	.18319063	.17812359	.00910096

G.5 Fixed schedules and hidden label randomization

The schedule comparison reuses the nine selected oracles at $K = 3, 4$. After sampling all A coordinates in the first round, a fixed-label schedule samples predetermined B groups of sizes $\boldsymbol{b} = (b_1, \ldots, b_r)$, where $r = K - 1$ and $\sum_j b_j = n$. We compare cumulative sizes $\lfloor n^{j/r} \rfloor$ and $\lfloor jn/r \rfloor$, and exhaust all ordered nonnegative compositions to find the best fixed-label schedule for

each oracle. Empty groups permit fewer effective rounds. There are 984 such schedule evaluations across the 18 oracle–depth settings.

Let M_j be the mismatch count in group j. The generated joint law of $(M_1, \dots, M_r)$ is a product of conditional binomial laws with rates $p(\sum_{i<j} b_i, \sum_{i<j} M_i)$; the teacher law is the product of $\mathrm{Bin}(b_j, \tau)$ laws. Both distributions are uniform on the same fibers of this vector, so *vector* TV equals full-output TV. Replacing the vector by its sum would generally underestimate the error of a fixed-label schedule.

Separately, we evaluate the policy that first privately permutes pair labels uniformly, then uses a fixed batch-size vector. After marginalizing this permutation, the mismatch law is exchangeable and its full-output TV equals total-count TV. We optimize this value separately over batch-size vectors. Private permutation changes the policy itself. The optimized family contains one batch-size vector per oracle; it excludes random mixtures of distinct vectors and other non-adaptive policies. Gaps from $d_K(R)$ therefore measure this family's suboptimality. These oracle-specific comparisons are separate from the uniform minimax guarantee.

Table 5 gives every comparison; Figure 3 displays $n = 16$. At $K = 3$, the best privately permuted fixed batch vector lies within 9.3×10^{-13} of $d_3(R)$ for all nine models. For example, at $n = 16$ and $B = 0.10$, the two mismatch batches each have size eight, following the representative round. At $K = 4$, the certified lower bounds on this family's suboptimality lie between 0.000683 and 0.002700. Equal fixed-label batches improve on geometric fixed-label batches in all nine four-round comparisons, but only five of the nine three-round comparisons. Theorem 2 instead compares schedules through their class-uniform asymptotic guarantees.

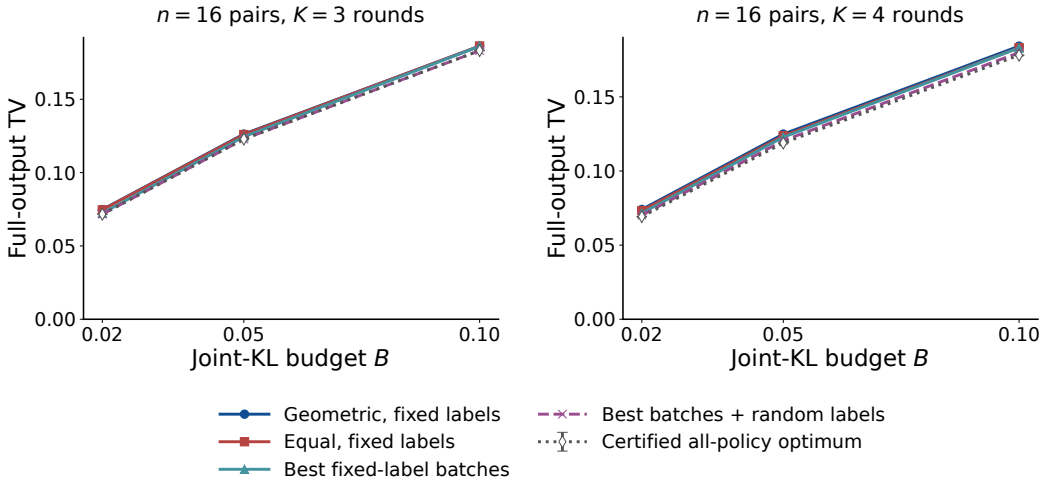


Figure 3: **Simple schedules versus the all-policy optimum.** For each $n = 16$ oracle, the same model is used at three and four rounds. Fixed-label TV is computed from the vector of batch counts; hidden-permutation TV is computed after marginalizing the private permutation. The latter is a different policy. Best batch sizes are oracle-specific, whereas geometric and equal schedules do not use the oracle parameters. Optimum enclosures are narrower than the markers.

G.6 Biased representatives

For the bias comparison, $A_i \sim \mathrm{Ber}(a)$ independently of E, with $a \in \{1/2, 2/5, 1/4\}$. We freeze

$$R_E = \frac{11681887}{16777216}\,\mathrm{Ber}(1/4)^{\otimes n} + \frac{5095329}{16777216}\,\mathrm{Ber}(1/2)^{\otimes n}$$

across biases, dimensions, and depths. We use $n = 4$ with $K = 2, 3, 4$ and $n = 6$ with $K = 3, 4$, for 15 cases. The $n = 6$ extension was subject to a prespecified assembly-resource gate, which passed. At fixed n, joint model KL and TV are independent of a: the common representative-bit

Table 5: Complete schedule comparison, in full-output TV (rounded displays). "Best fixed" and "Best perm." optimize one A-first batch vector, with fixed labels or a hidden uniform pair permutation, respectively. Both optimize within these specified families. The final column displays the all-policy certified optimum, whose unrounded enclosure is given by the construction above.

n	B	K	Geometric	Equal	Best fixed	Best perm.	$d_K(R)$
8	.02	3	.06889748	.06913058	.06342333	.06259081	.06259081
8	.02	4	.06749225	.06612404	.06318673	.06036439	.05936764
8	.05	3	.12110126	.12115944	.11471879	.11316915	.11316915
8	.05	4	.11775944	.11600686	.11216048	.10917188	.10733143
8	.10	3	.18449655	.18499496	.18005284	.17691617	.17691617
8	.10	4	.17938648	.17728969	.17506672	.17073859	.16803951
12	.02	3	.06383769	.06214984	.06067055	.05596105	.05596105
12	.02	4	.06263196	.06172682	.05820118	.05294498	.05226142
12	.05	3	.11500552	.11223945	.10904045	.10444571	.10444571
12	.05	4	.11262931	.11101040	.10566566	.09894718	.09770445
12	.10	3	.18009425	.17536863	.17321404	.16710498	.16710498
12	.10	4	.17469055	.17301394	.16694662	.15850294	.15686703
16	.02	3	.07477652	.07458138	.07234096	.07160591	.07160591
16	.02	4	.07408637	.07323261	.07152549	.07019252	.06921703
16	.05	3	.12627548	.12592019	.12442824	.12284334	.12284334
16	.05	4	.12489343	.12376135	.12279122	.12045984	.11887869
16	.10	3	.18625445	.18630924	.18613168	.18319063	.18319063
16	.10	4	.18422415	.18315500	.18299912	.17992215	.17812359

factor cancels under the invertible map $(A, E) \mapsto (A, B)$. Changing n changes the actual KL; the budget slack relative to the original 0.1 budget is reported below.

The controlled state is the histogram of nine pair types, $\varnothing, A0, A1, B0, B1, 00, 01, 10, 11$. For component rate p, put $q_p = a + (1-2a)p$. Their likelihoods are respectively

$$1, \quad 1-a, \quad a, \quad 1-q_p, \quad q_p, \quad (1-a)(1-p), \quad (1-a)p, \quad ap, \quad a(1-p).$$

The component posterior is proportional to its prior weight times the product of these likelihoods over the current histogram. In particular, B-only observations must enter this posterior. Within component p, the completion heads include

$$\Pr(B=1 \mid A=0,p) = p, \qquad \Pr(B=1 \mid A=1,p) = 1-p,$$
$$\Pr(A=1 \mid B=0,p) = \frac{ap}{1-q_p}, \qquad \Pr(A=1 \mid B=1,p) = \frac{a(1-p)}{q_p}.$$

The oracle averages these quantities against the *current* component posterior. In particular, each nonlinear reverse-head expression is averaged as a whole. Actions include both singleton orientations, completion of each singleton type, and independent co-sampling of both coordinates of fresh pairs. Exact multinomial convolution gives the histogram transition kernel.

Terminal tests use the full four-type histogram of completed pairs. Pair permutations are transitive on each such histogram, so uniformly lifted histogram policies again have full-output TV equal to histogram TV. The all-action Bellman and directed forward certificates apply unchanged. Bias may amplify reverse-head errors: we enumerate all available heads at every partial histogram and report the maximum. The comparison thus measures sensitivity to bias at a fixed mismatch mixture; its realized head-error budget can vary with the bias.

Table 6 reports the full grid. Joint KL is below 0.05330874 at $n = 4$ and 0.07756038 at $n = 6$; model TV is approximately 0.12931201 and 0.14784871, respectively, for every listed bias. All 15 optimum enclosures have width below 1.8×10^{-12}. Every tested adjacent depth has a strict certified separation, including all six three-to-four-round comparisons. Several feasible policies satisfy $U_K < \mathrm{TV}(P, R)$, as shown in Figure 4. The robust feature across these instances is depth separation; the sign of the excess over the serial baseline varies.

We retain the one head-bound exception: at $n = 6$, $a = 2/5$, the exact maximum equals $434801408/1735642263 > 1/4$. At $a = 1/4$ the measured maximum is below $1/4$ at both di-

mensions, and the six-pair three-to-four-round gap is greater than 0.00390. This supplies a non-fair finite example within the quarter-head constraint. The finite grid leaves both uniformity over all biases and the fixed-bias asymptotic exponent unresolved.

Table 6: Fixed-mismatch-mixture bias study. All TV values except the last column are rounded displays; the last column is a downward-rounded certificate for $d_3 - d_4$. A dash marks an untested depth. †: the worst head error exceeds $1/4$; the comparison fixes the mismatch law rather than a common realized head-error budget.

n	a	Max. head TV	$d_2(R)$	$d_3(R)$	$d_4(R)$	$d_3 - d_4 \geq$
4	$1/2$	.19431312	.13252227	.12982138	.12560647	.00421490
4	$2/5$	.21198934	.13252227	.12831854	.12471677	.00360176
4	$1/4$	.20577191	.13225312	.12592810	.12305997	.00286813
6	$1/2$	.23328603	—	.14594037	.14022087	.00571950
6	$2/5$	.25051326†	—	.14478692	.13964850	.00513842
6	$1/4$	.23725145	—	.14232287	.13842113	.00390174

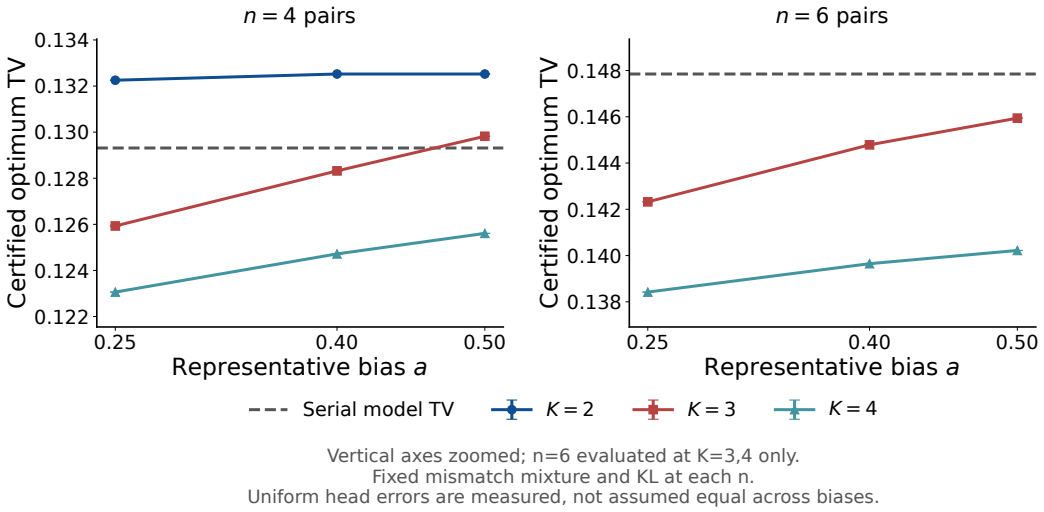


Figure 4: **Depth separation with informative endpoints.** The mismatch mixture is fixed across biases, dimensions, and depths. At fixed n, joint KL and model TV are unchanged, but conditional heads and optimal execution error change with a. Vertical axes are zoomed; arithmetic enclosures are narrower than the markers. Only $K = 3, 4$ are evaluated at $n = 6$. The $n = 6, a = 2/5$ setting exceeds the quarter-head constraint (Table 6); no point is excluded for this reason. The dashed model-TV line is the feasible serial baseline; several parallel executions improve upon it.

G.7 Implementation, validation, and limitations

We use Python 3.13.13, NumPy 2.2.6, and SciPy 1.17.1 (HiGHS), with standard-library integer and rational arithmetic for certificates. Runs use one CPU thread and an eight-GiB address-space limit, without a GPU. Fair LPs allow 250,000 state–action variables and 120 seconds. Two four-round timeouts were recovered with HiGHS interior-point solves, without changing the selected models or ranking. The schedule enumeration allows 600 seconds; biased runs allow 1,200 seconds overall and 90 seconds per LP.

The $n = 4$ bias grid is unconditional. The $n = 6$ extension requires a $(a, K) = (2/5, 4)$ kernel-graph build below 120,000 variables, three million transition entries, and 120 seconds, independently of measured error gaps. It passes with 111,615 variables and 1,069,677 entries. All 27 fair and 15 biased cases are certified; unsuccessful solver attempts are retained in the records.

Independent checks. Direct enumeration of labeled outputs checks reduced kernels, TV calculations, exact-target controls, and reverse heads at small dimensions. Fair-bias cases agree with the separate three-count implementation. The biased LP and certificate share the reduced kernel; direct enumeration independently checks that reduction. These checks are not proof-assistant verification.

H Model error, pilot feedback, and small-instance policy optima

This appendix reports three complementary diagnostics: a dimension sweep for a prescribed two-round execution, a three-round self-generated-pilot study, and small-instance policy optimization. All three use fair representatives and $\tau = 1/4$. The calculations evaluate finite distributions directly in double precision. We retain every parameter setting, including the ladder's smaller errors and instances where the best pilot agrees with the numerical all-policy optimum.

Scope of the diagnostics. E1 evaluates a prescribed execution, E2 searches a finite pilot grid, and E3 numerically optimizes all policies for two fixed small-instance oracles. Numerical residuals provide implementation checks. Appendix G supplies rigorous arithmetic enclosures for a separately selected set of oracles. The single and ladder families here have different actual KL budgets, so their comparison describes mechanisms rather than matched-budget hardness.

H.1 E1: Decreasing model error and increasing execution error

For $n = 2^j$, $j = 8, \dots, 20$, define

$$s_n = (n/256)^{-1/8}, \qquad w_n = s_n/8, \qquad \Delta_n = s_n/4,$$
$$R_E = (1 - w_n)\,\mathrm{Ber}(\tau)^{\otimes n} + w_n\,\mathrm{Ber}(\tau + \Delta_n)^{\otimes n}.$$

The oracle and teacher have the same independent fair representative bits. The prescribed execution samples all representatives in round one and all partners in round two. Frozen heads make its mismatch law $\mathrm{Ber}(\tau + w_n\Delta_n)^{\otimes n}$: each head uses the mixture's mean rate. Consequently, with $\mathrm{Bin}(n, p)$ denoting the binomial count law,

$$\mathrm{TV}(P, R) = w_n\,\mathrm{TV}\big(\mathrm{Bin}(n, \tau), \mathrm{Bin}(n, \tau + \Delta_n)\big),$$
$$\mathrm{TV}(P, Q_R^{\mathrm{two}}) = \mathrm{TV}\big(\mathrm{Bin}(n, \tau), \mathrm{Bin}(n, \tau + w_n\Delta_n)\big).$$

These are full-output identities. We also evaluate the actual joint $\mathrm{KL}(P\|R)$ by summing over all $n+1$ mismatch counts. The inequalities $\mathrm{KL}(P\|R) \le -\log(1 - w_n)$ and uniform head-TV error at most Δ_n provide reference bounds alongside these measured joint quantities. The worst head error is bounded above by Δ_n and can be smaller.

Table 7 and Figure 5 show model TV decreasing from 0.124997 to 0.044194 while this execution's TV increases from 0.428496 to 0.999996, illustrating amplification under the prescribed execution. Binomial TV is evaluated using the monotone-likelihood-ratio crossing and binomial CDFs; KL uses the full finite sum with no intentional tail truncation. The maximum teacher mass residual is 8.89×10^{-16}. Comparisons against direct PMF sums and small-count rational-mass calculations give discrepancies below 1.10×10^{-15} and 8.68×10^{-17}, respectively.

H.2 E2: A self-generated, privately randomized pilot

Here $n \in \{64, 256, 1024, 4096\}$, $w = 1/8$, $\Delta = 1/4$, and $\lambda = 1/5$. Let $\delta = \Delta(w/2)\lambda = 1/320$. We compare

$$R_E^{\mathrm{single}} = (1 - w)\,\mathrm{Ber}(\tau)^{\otimes n} + w\,\mathrm{Ber}(\tau + \Delta)^{\otimes n},$$
$$R_E^{\mathrm{ladder}} = (1 - w)\,\mathrm{Ber}(\tau)^{\otimes n} + \tfrac{w}{2}\,\mathrm{Ber}(\tau + \Delta)^{\otimes n} + \tfrac{w}{2}\,\mathrm{Ber}(\tau + \delta)^{\otimes n},$$
$$R_E^{\mathrm{matched}} = (1 - w)\,\mathrm{Ber}(\tau)^{\otimes n} + w\,\mathrm{Ber}\big(\tau + (\Delta + \delta)/2\big)^{\otimes n}.$$

The last family matches the ladder's *initial predictive drift*, $w(\Delta + \delta)/2 = 81/5120$. The original single alternative has initial drift $w\Delta = 1/32$. Actual KL and model TV remain family-dependent.

Table 7: All 13 settings of E1. KL is in nats. Output TV refers only to the all-representatives/all-partners execution. Values are rounded to six decimal places; w_n and Δ_n are defined by the displayed formulas.

n	w_n	Δ_n	Actual KL	Model TV	Output TV
2^{8}	0.125000	0.250000	0.133519	0.124997	0.428496
2^{9}	0.114626	0.229251	0.121745	0.114626	0.500786
2^{10}	0.105112	0.210224	0.111057	0.105112	0.579329
2^{11}	0.096388	0.192776	0.101355	0.096388	0.662711
2^{12}	0.088388	0.176777	0.092541	0.088388	0.747051
2^{13}	0.081052	0.162105	0.084526	0.081052	0.826688
2^{14}	0.074325	0.148651	0.077233	0.074325	0.895051
2^{15}	0.068157	0.136313	0.070591	0.068157	0.946426
2^{16}	0.062500	0.125000	0.064539	0.062500	0.978417
2^{17}	0.057313	0.114626	0.059021	0.057313	0.993754
2^{18}	0.052556	0.105112	0.053987	0.052556	0.998863
2^{19}	0.048194	0.096388	0.049394	0.048194	0.999892
2^{20}	0.044194	0.088388	0.045200	0.044194	0.999996

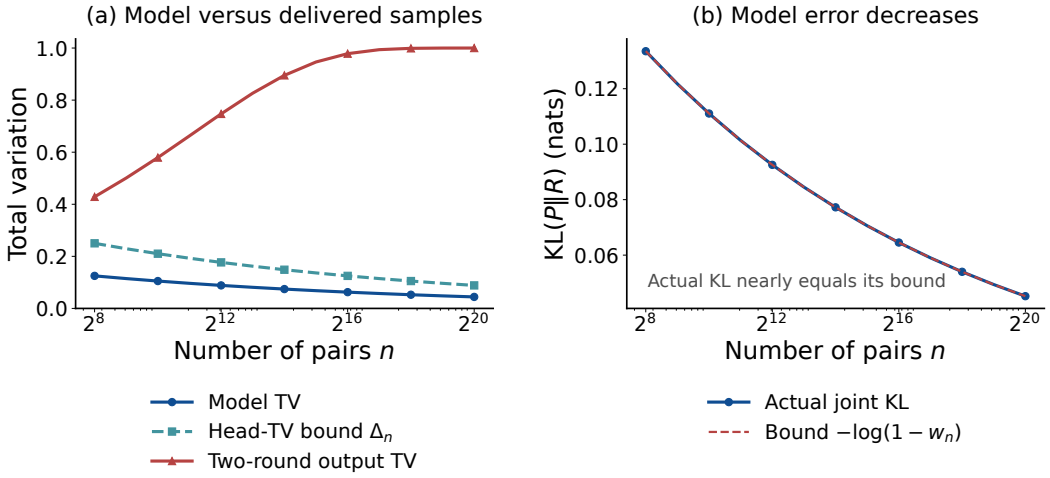


Figure 5: E1: the coherent model approaches the teacher while a particular two-round execution becomes inaccurate. Left: model TV, the uniform head-TV bound, and output TV. Right: actual joint KL and its contamination bound. The 13 dimensions and their numerical values are given in Table 7. All output errors concern the prescribed execution.

The policy first samples every representative. It then chooses a uniformly random subset of c partner coordinates, samples those coordinates in round two, and completes the remainder in round three. The subset is private randomness and is marginalized out of the delivered output. For a mixture with weights w_ℓ and rates p_ℓ, put $p_0 = \sum_\ell w_\ell p_\ell$. The pilot mismatch count is

$$M \sim \mathrm{Bin}(c, p_0), \qquad p_c(m) = \frac{\sum_\ell w_\ell p_\ell^{m+1}(1-p_\ell)^{c-m}}{\sum_\ell w_\ell p_\ell^{m}(1-p_\ell)^{c-m}}.$$

Thus the pilot law uses the model's initial prediction p_0 in place of the teacher rate τ. Conditional on $M = m$, the remaining mismatch count is $\mathrm{Bin}(n-c, p_c(m))$. If $b_{s,p}$ denotes a binomial PMF, extended by zero outside $\{0, \dots, s\}$, the delivered count law is

$$q_c(t) = \sum_{m=0}^{c} b_{c,p_0}(m) b_{n-c,p_c(m)}(t-m).$$

Private uniform label selection makes the output exchangeable, so the TV between q_c and $\mathrm{Bin}(n, \tau)$ is full-output TV. For a deterministic labeled pilot, full-output TV generally requires the joint law of

Table 8: E2: actual model discrepancies, both endpoint schedules, and the best pilot in the evaluated grid. "Matched" means initial-drift matched to the ladder; actual KL is reported separately. The endpoint column reports $c = 0$ and $c = n$, which have identical output laws.

n	Family	Actual KL	Model TV	Endpoint TV	Best c	Best TV
64	Single	0.119400	0.120695	0.221893	30	0.128686
64	Ladder	0.056019	0.060526	0.115565	30	0.064222
64	Matched	0.054024	0.090915	0.115565	43	0.094514
256	Single	0.133519	0.124997	0.428496	56	0.128170
256	Ladder	0.064560	0.062499	0.227397	56	0.064217
256	Matched	0.121937	0.121517	0.227397	93	0.127599
1024	Single	0.133531	0.125000	0.742672	82	0.103774
1024	Ladder	0.064659	0.062509	0.437099	82	0.053613
1024	Matched	0.133526	0.124999	0.437099	181	0.125392
4096	Single	0.133531	0.125000	0.976509	93	0.083909
4096	Ladder	0.065051	0.063684	0.752873	93	0.047521
4096	Matched	0.133531	0.125000	0.752873	424	0.106021

pilot and remaining counts. We also compute $\mathbb{E}[p_c(M) - \tau]$ under the actual pilot law. For $c = n$ this is a posterior diagnostic only: no remaining coordinates use that posterior head. Both endpoints $c = 0$ and $c = n$ have output law $\mathrm{Bin}(n, p_0)$.

Complete pilot grid. For each dimension and family we use the deduplicated union of $\{0, n\}$, 12 rounded geometrically spaced values from 1 to $n - 1$, and rounded values of $n^{1/4}, n^{1/2}, n^{3/4}, n/2$. The realized sets are

$$\begin{aligned}
\mathcal{C}_{64} &= \{0, 1, 2, 3, 5, 7, 8, 10, 14, 20, 23, 30, 32, 43, 63, 64\},\\
\mathcal{C}_{256} &= \{0, 1, 2, 3, 4, 5, 8, 12, 16, 21, 34, 56, 64, 93, 128, 154, 255, 256\},\\
\mathcal{C}_{1024} &= \{0, 1, 2, 4, 6, 7, 12, 23, 32, 44, 82, 155,\\
&\qquad 181, 290, 512, 545, 1023, 1024\},\\
\mathcal{C}_{4096} &= \{0, 1, 2, 5, 8, 10, 21, 44, 64, 93, 199, 424,\\
&\qquad 512, 903, 1923, 2048, 4095, 4096\}.
\end{aligned}$$

This gives $3(16 + 18 + 18 + 18) = 210$ configurations. Figure 7 contains every positive-c value; the $c = 0$ values excluded by its logarithmic axis appear in Table 8. The reported best c minimizes TV over exactly this finite grid. Other integer pilot sizes and other three-round policies lie outside the E2 search.

The pilot substantially reduces output error in these settings. At $n = 4096$, for example, the best evaluated TVs are 0.083909 for the single alternative, 0.047521 for the ladder, and 0.106021 for the initial-drift-matched single alternative. The ladder also has smaller actual KL and model TV. These comparisons illustrate posterior feedback while showing that matching the initial drift leaves substantial differences in both model accuracy and execution error. Output TV may be below model TV because parallel execution of conditional heads can change the oracle's joint law.

The count calculations retain the full finite sums without intentional tail pruning; tiny probabilities may still underflow in double precision. The maximum output-mass residual over the 210 configurations is 1.56×10^{-15}. An independent rational calculation of all labeled outputs and all pilot subsets for $n \le 4$ checks 42 settings, with maximum TV discrepancy 1.25×10^{-16}. These checks support the implementation and symmetry reduction. The larger-dimensional values retain the double-precision status stated at the start of this appendix.

H.3 E3: ALL-POLICY OPTIMIZATION FOR TWO FIXED SMALL-INSTANCE ORACLES

We next hold the single and ladder mixtures of E2 fixed across round caps and evaluate $n \in \{2, 4, 6, 8\}$ and $K \in \{2, 3, 4\}$. This produces 24 occupation-LP problems. The controlled-state reduction and its relation to full-output TV are described in Appendix G; the present values use floating-point optimization.

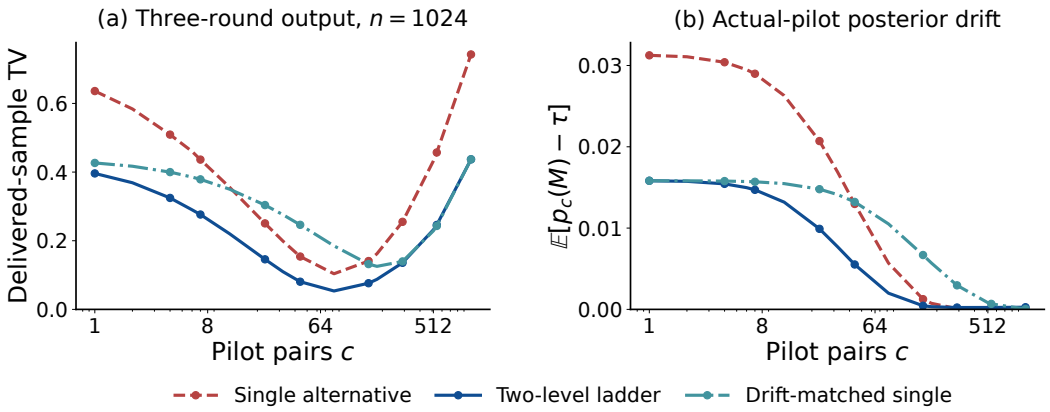


Figure 6: E2 at $n = 1024$: output TV (left) and expected posterior drift under the self-generated pilot (right). Curves connect evaluated positive pilot sizes. Increasing the pilot changes both the pilot's own accumulated error and the heads used by the remaining batch. At $c = n$, the posterior drift describes the final posterior, after all bits have been generated.

Table 9: All 24 E3 numerical policy optima, compressed into one row per dimension and fixed oracle. The same R is used across the three depth columns. Actual KL and model TV differ between the single alternative and the ladder. The final three columns report floating-point LP optima. Appendix G certifies a separately selected set of oracles.

n	Family	Actual KL	Model TV	$K = 2$	$K = 3$	$K = 4$
2	Single	0.005273	0.039062	0.045898	0.039062	0.039062
2	Ladder	0.001436	0.019824	0.023480	0.019824	0.019824
4	Single	0.011509	0.053223	0.053683	0.052340	0.050735
4	Ladder	0.003386	0.026942	0.026959	0.026254	0.025484
6	Single	0.018178	0.060852	0.072374	0.061006	0.057245
6	Ladder	0.005726	0.030737	0.037114	0.031045	0.028950
8	Single	0.024961	0.066751	0.079001	0.072153	0.068458
8	Ladder	0.008319	0.033863	0.039771	0.036282	0.034420

The comparison schedules sample one representative from every pair in round one, then complete uniformly chosen open pairs. Equivalently, they use an all-A first round and a privately randomized pair permutation. With $r = K - 1$, their cumulative completion counts are $\lceil nj/r \rceil$ for equal batches and $\lfloor n^{j/r} \rfloor$ for the geometric schedule, ending at n. The best completion grid enumerates every nondecreasing integer prefix $0 \le c_1 \le \cdots \le c_{r-1} \le n$, followed by $c_r = n$. Thus "fixed" refers to the batch-size vector while pair labels remain privately randomized. For $K = 3$, this grid contains every integer pilot size. Appendix G separately evaluates fixed-label schedules. The equal-batch convention here uses ceilings, whereas that fixed-label comparison uses floors.

In Figure 8, the all-policy optimum and best random-label pilot agree numerically at three rounds: their maximum discrepancy is 1.11×10^{-16} across the eight tested instances. The ladder again has smaller numerical output error than the single alternative throughout Table 9, together with smaller actual KL. E3 therefore motivates the separately budget-calibrated and rationally certified comparisons of Appendix G. The observed agreement concerns these instances; optimality of such pilots in general remains unresolved.

At four rounds, the broader policy class does improve on this restricted schedule family. For $n = 8$, the single-alternative LP value is approximately 0.0684583, versus 0.0695835 for the best randomized-label batch vector; the ladder values are 0.0344204 and 0.0350223, respectively. These small numerical gaps do not isolate the value of adaptation or private randomness.

Across the 24 LPs, the largest flow residual is 8.33×10^{-17}, the largest absolute primal–dual objective gap is 4.17×10^{-17}, and the largest dual-slack violation is 5.56×10^{-15}. An independently

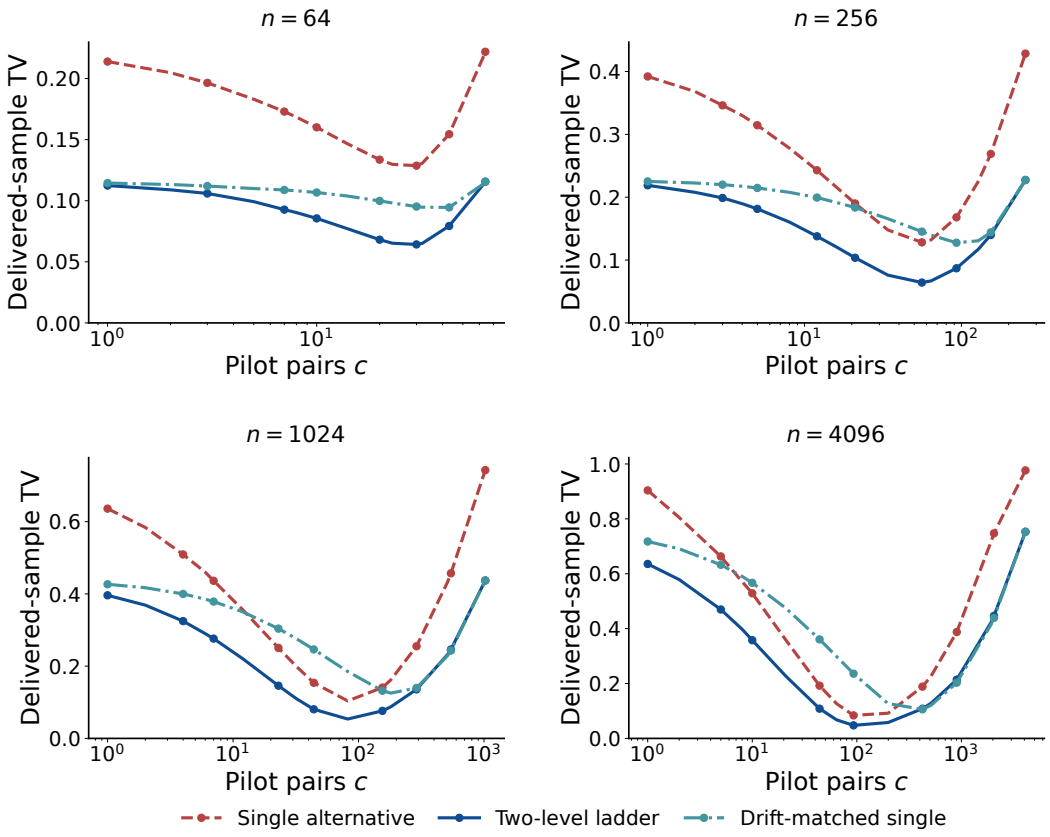


Figure 7: All positive-pilot configurations of E2 across the four dimensions and three oracle families: 198 evaluated values in total. Table 8 supplies the 12 additional $c = 0$ values and the endpoint identities. Lines connect the evaluated grid points. Actual KL differs across families.

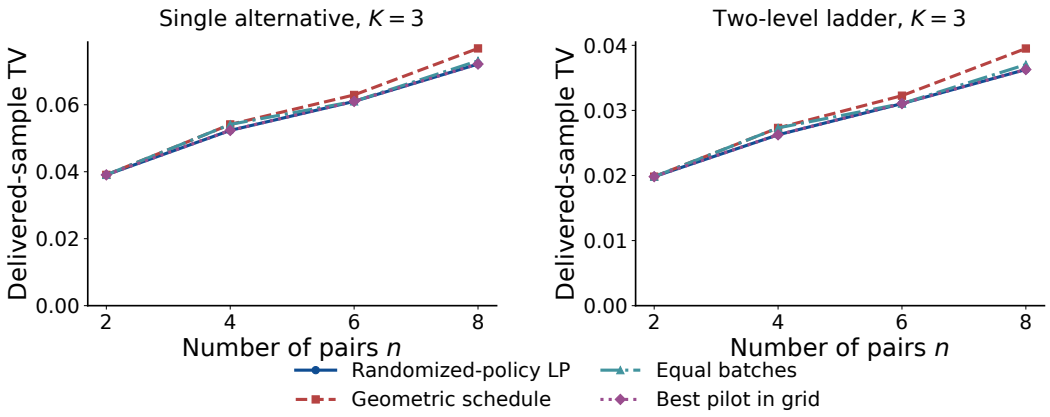


Figure 8: E3 at $K = 3$. The numerical all-policy optimum and the best hidden-uniform-label pilot agree to floating-point accuracy in all eight oracle–dimension settings, so their curves overlap. Equal and geometric schedules also randomize pair labels; no fixed-label TV is shown here. Table 9 reports every depth, including the unplotted $K = 2, 4$ cases.

recomputed forward law differs from the LP terminal law by at most 2.78×10^{-16} per count. Literal-coordinate LP checks at $n = 2$ agree within 1.12×10^{-16}. A separate one-pair control also checks that private mixtures are evaluated at the level of output laws, as required by the sampling objective. These residuals are the implementation diagnostics described at the start of the appendix.

The software versions are those in Appendix G. This study uses one CPU thread, an eight-GiB address-space cap, and a 45-minute overall limit; each E3 LP allows 120 seconds and 250,000 state–action variables. All listed settings completed using direct finite-law calculations on the CPU.

I Budget-matched pilots at larger dimensions

The small-instance certificates in Appendix G optimize policies for a fixed oracle, while the earlier pilot study uses models with different actual KL. Here we isolate the ladder mechanism by matching actual joint KL and increasing the dimension. We evaluate selected three-round policies and count tests, retaining a separate curve for the analytic all-policy bound.

I.1 Models, calibration, and policies

We use $\tau = 1/4$ and compare the single alternative in (6), with $\Delta = 1/4$, against the genuine two-level ladder

$$R_E^{\mathrm{lad}} = (1-w)P_E + \frac{w}{2}\,\mathrm{Ber}(1/2)^{\otimes n} + \frac{w}{2}\,\mathrm{Ber}(1/4 + w/24)^{\otimes n}. \tag{82}$$

This is (12) with $r = 2$, $b = w/2$, and $\lambda = 1/3$; its lower component changes with w during calibration. Both families obey $0 < w \le 1/8$. For each $n \in \{10^4, 10^5, 10^6, 10^7, 10^8, 10^9\}$ and $B \in \{0.02, 0.05, 0.10\}$, we numerically solve $\mathrm{KL}(P_E \| R_E) = B$ separately for each family. The ladder at $n = 10^4$, $B = 0.10$ is infeasible under the weight cap: its KL at $w = 1/8$ is 0.0694020. We report this case as missing rather than comparing unmatched budgets. The remaining 35 calibrations have absolute numerical residual below 1.9×10^{-14}.

The missing case can also be excluded analytically, without assuming KL is monotone in w. Put $b = w/2$, $\gamma = b/(1-b)$, and $\delta = w/24$. Dropping the high-rate component gives $R_E \ge (1-b)[(1-\gamma)P_E + \gamma\,\mathrm{Ber}(\tau+\delta)^{\otimes n}]$. The bound $\mathrm{KL} \le \chi^2$ then yields

$$\mathrm{KL}(P_E \| R_E) \le -\log(1-b) + \frac{\gamma^2}{1-\gamma}\left[(1+\delta^2/v)^n - 1\right].$$

The right side increases with w and is below 0.081 at $w = 1/8$, $n = 10^4$.

For each model, let $p_c(m)$ be its posterior predictive mismatch rate and $q_0 = p_0(0)$. The generated-pilot policy reveals all representatives in round one, chooses a hidden uniform set of c pair labels and completes them in round two, then completes the remaining pairs in round three. Thus

$$M \sim \mathrm{Bin}(c, q_0), \qquad H \mid M \sim \mathrm{Bin}(n-c, p_c(M)), \qquad S = M + H. \tag{83}$$

We also evaluate a *teacher-pilot control*, replacing the law of M by $\mathrm{Bin}(c, \tau)$ while retaining the same completion heads. This comparison isolates the information carried by the pilot. The original-head interface does not supply teacher draws; Appendix K gives a target-aware candidate-selection implementation for a half-size pilot.

Each model uses 12 geometrically spaced integer pilot sizes from 1 to $\lfloor n/2 \rfloor$, together with $c = \lfloor \sqrt{n} \rfloor$, for 13 distinct sizes. Evaluating both pilot laws gives 910 cases. No training, Monte Carlo sampling, or GPU computation is used.

I.2 Count witnesses and numerical evaluation

For each case, evaluate the signed terminal-event discrepancy

$$W(c) = \max_{k \in \mathcal{K}} \left[\mathbb{E}_M \mathbb{P}\{\mathrm{Bin}(n-c, p_c(M)) \geq k - M\} - \mathbb{P}\{\mathrm{Bin}(n, \tau) \geq k\} \right]_+ . \tag{84}$$

The threshold set $\mathcal{K}$ contains integer thresholds at target mean plus z target standard deviations for $z \in \{0, 0.5, 1, 1.5, 2, 3, 4, 5, 6\}$, and at the midpoint between the target and execution means. Maximizing over this finite set remains a TV lower witness for the specified output law. It neither computes full TV nor optimizes over policies; a small witness alone does not certify accurate sampling.

The KL computation sums over a target binomial count, with a retained window chosen at tail tolerance $10^{-16}/(n+1)$. The omitted absolute integral is bounded using $\max\{-\log(1-w), n \log 2\}$. Pilot expectations use tail tolerance 10^{-12}. Bernstein windows have radius $\sqrt{2Np(1-p)\log(2/\xi)} + 2\log(2/\xi)/3$ for a binomial law with parameters (N, p) and tolerance ξ; CDF and survival-function evaluations record their actual omitted mass. Retained PMFs are normalized to this retained mass to correct accumulated floating-point mass error.

Posterior weights and expected drift are summed over every retained pilot count. For long supports, the conditional tail in (84) is bracketed by endpoint sums over at most 2,048 ordered bins. Both $p_c(m)$ and the conditional tail probability are nondecreasing in m, so this grouping has explicit lower and upper sums in exact arithmetic. The figures use the lower sum. The largest recorded grouping gap is 4.95×10^{-4}; at $c = \lfloor \sqrt{n} \rfloor$, the retained supports are small enough to use ungrouped sums. Maximum KL-tail and pilot-mass omissions are below 2.5×10^{-18} and 5.2×10^{-14}, respectively. Full-support recomputation checks five representative grouped or ungrouped cases, including the largest grouping gap. These truncation and grouping controls leave floating-point roundoff uncertified; this study is distinct from the rational certificates in Appendix G.

I.3 Results and posterior mechanism

At actual KL $B = 0.05$, the ladder's generated-pilot witness rises from 0.06048 at $n = 10^6$ to 0.18420, 0.64692, and 0.98634 at $n = 10^7, 10^8, 10^9$, respectively. The single-alternative witnesses remain near 0.01123 over these dimensions. Matching KL therefore reveals a separation obscured by the earlier unmatched-budget comparison. It concerns this pilot schedule, while the analytic all-policy curve remains substantially smaller at the displayed scales. For example, at $n = 10^9$, $B = 0.10$, calibrated weight $w \approx 0.095163$ gives a finite all-policy bound of approximately 0.05650 by direct substitution in Theorem 3.

Figure 10 shows why pilot information and pilot error must be considered together. Under a generated pilot, the high-drift component loses posterior mass, but the lower-drift component can subsequently gain it. The expected predictive drift can increase again as c grows. This behavior is consistent with the lower-bound proof: its recursively shrinking thresholds are certified lower bounds, not a deterministic contraction rule for the actual posterior mean. For the single alternative, a larger pilot suppresses the remaining predictive drift, yet a sufficiently large generated pilot already incurs substantial count error. Teacher pilots avoid that committed pilot error, but require an enlarged interface.

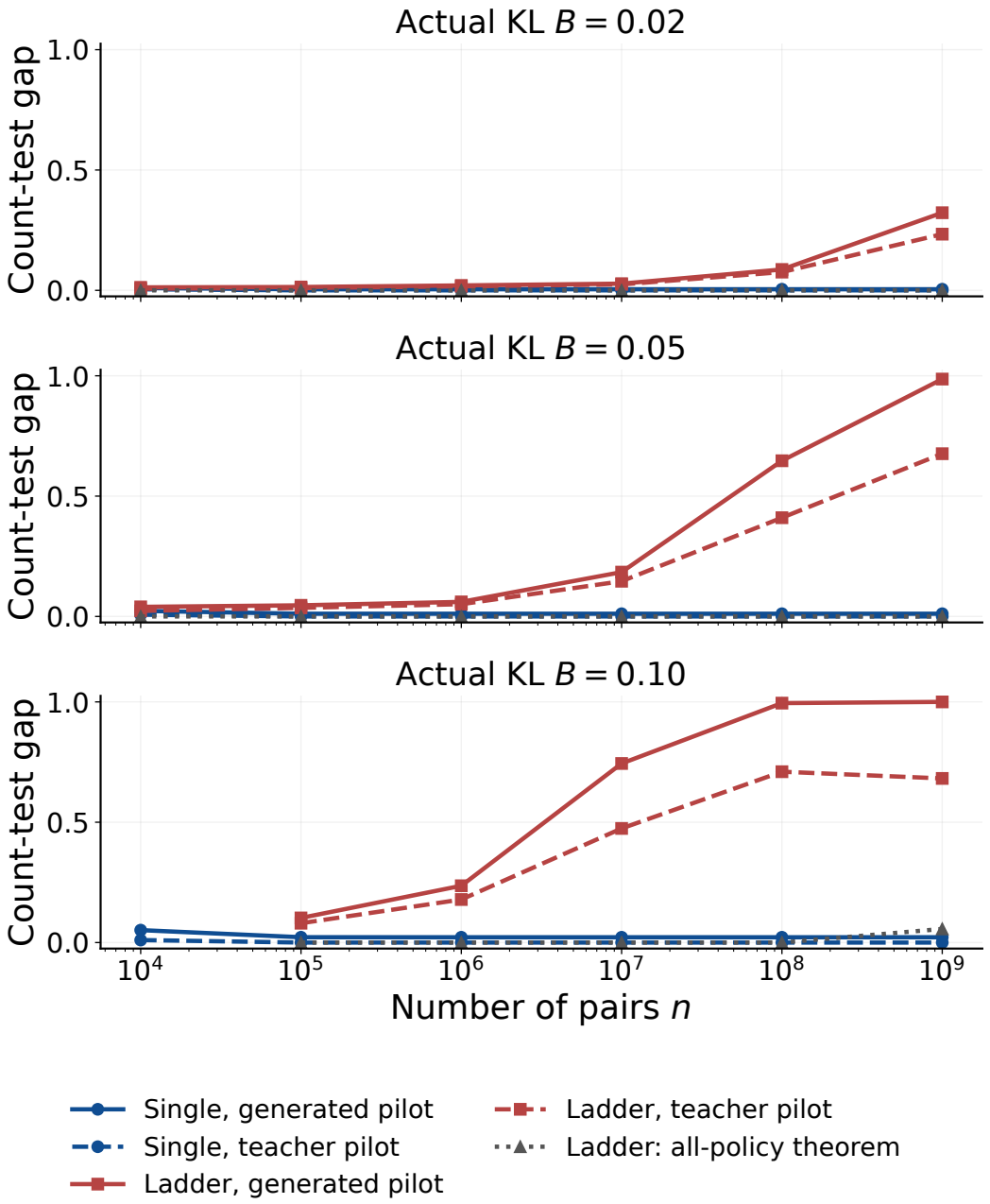


Figure 9: **Matched actual KL at a geometric-size pilot.** Solid curves use generated pilots and dashed curves use teacher pilots, always with $c = \lfloor \sqrt{n} \rfloor$. Values are numerical count-test lower witnesses for the prescribed output law. The gray curve separately evaluates Theorem 3 for the two-level ladder and covers every policy with three rounds. The missing red point at $n = 10^4$, $B = 0.10$ is infeasible under $w \leq 1/8$.

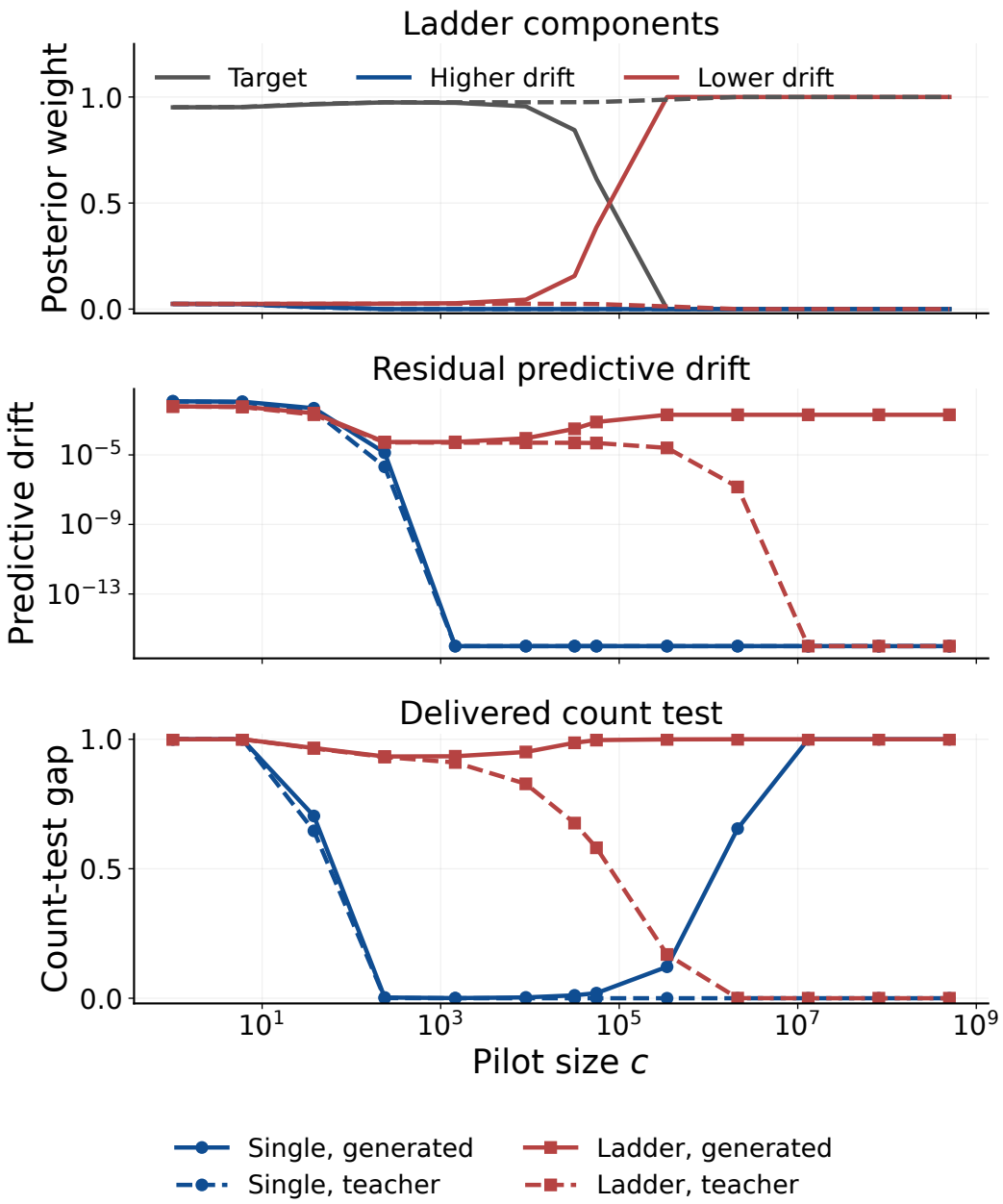


Figure 10: **Posterior feedback at matched KL.** At $n = 10^9$, $B = 0.05$, solid lines use generated pilots and dashed lines teacher pilots. Top: expected posterior component weights for the ladder. Middle: expected residual mismatch drift (values below 10^{-16} are clipped for display). Bottom: a terminal-count lower witness, including errors already committed in the pilot. Lines connect the evaluated sizes, rather than individual sample trajectories.

I.4 A TARGET-AWARE HALF-PILOT CONTROL

To give an upper as well as a lower diagnostic, take $c = \lfloor n/2 \rfloor$ and the candidate-selection policy of Theorem 12. For the present models, $q_0 \in [\tau, \tau + 1/32]$ and $\alpha_{\mathrm{acc}} \geq 23/24$. Its failure probability is

$$p_{\mathrm{fail}} = \mathbb{P}\{\mathrm{Bin}(n, \alpha_{\mathrm{acc}}) < c\} \leq e^{-n/8}.$$

Conditional on success, its retained pilot has the teacher law. Since $\tau \leq p_c(m) \leq 1/2$, Pinsker and the Bernoulli KL bound give

$$\mathrm{TV}(P_n, Q_R^{\mathrm{sel}}) \leq \min\left\{1,\ p_{\mathrm{fail}} + \sqrt{\frac{n-c}{2v}}\,\mathbb{E}_{M\sim\mathrm{Bin}(c,\tau)}[p_c(M) - \tau]\right\}. \tag{85}$$

This bound compares the augmented laws containing the hidden pilot labels and then marginalizes those labels. In evaluating it, omitted pilot probability is charged by one outside the retained sum. Unlike a small count-test witness, a small value of this analytic upper expression controls full-output TV. Its numerical evaluation is still floating-point, not an interval certificate.

For the ladder at $B = 0.05$, this evaluated upper is 0.01782 at $n = 10^6$ and 1.41×10^{-7} at $n = 10^7$. At larger dimensions the displayed values reach the charged truncation mass, around 5×10^{-14}; this plateau should not be interpreted as certified numerical accuracy at that scale.

All calibrated model parameters, pilot sizes, selected test thresholds, component-weight expectations, omission and grouping diagnostics, and the one infeasible calibration are included in the accompanying numerical supplement. The count formulas use supports of order $\sqrt{n \log n}$ instead of enumerating length-n samples, making the large dimensions computationally inexpensive.

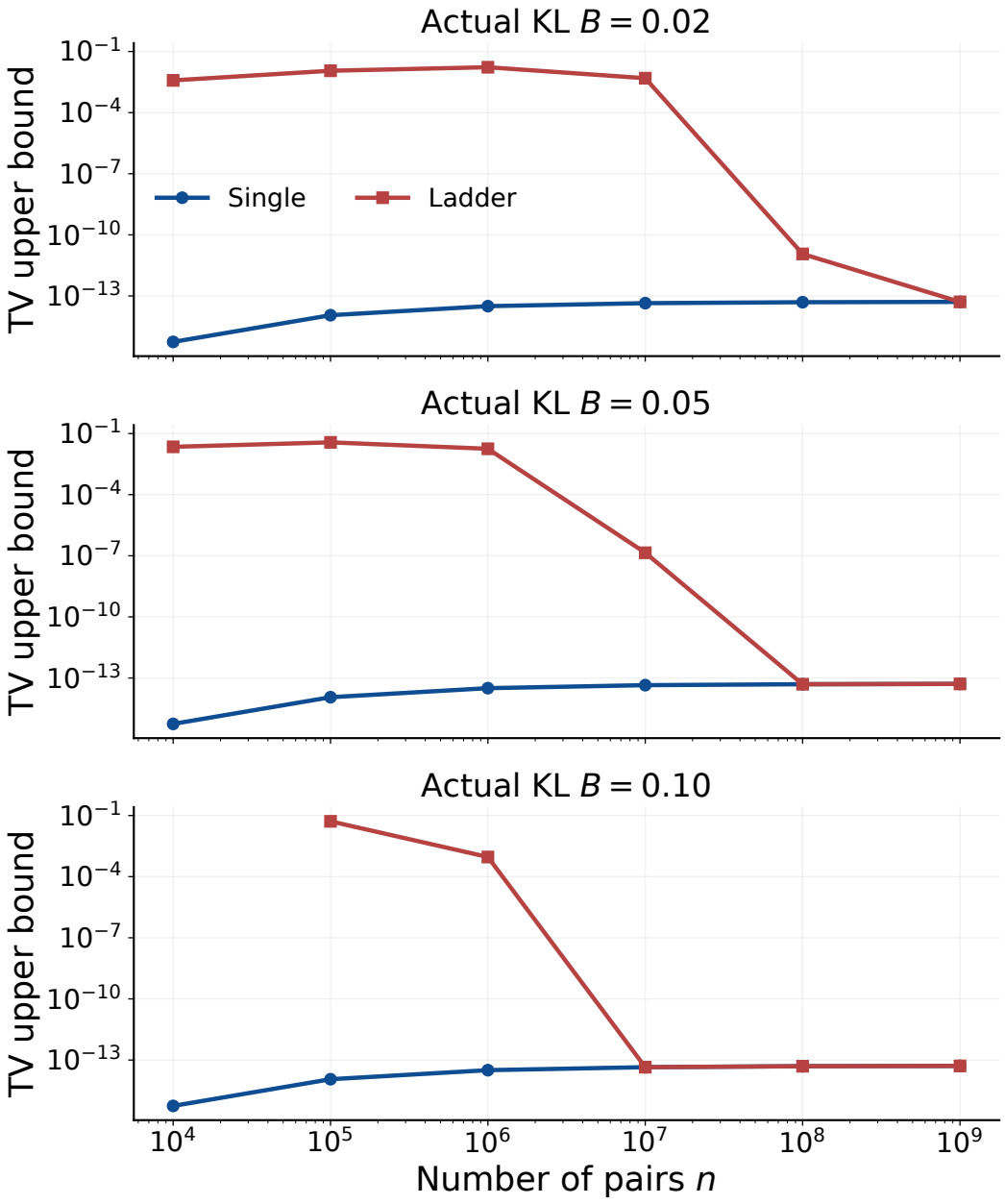


Figure 11: **Target-aware selection control.** Floating-point evaluation of the full-output upper bound (85), using a teacher-distributed retained pilot of size $\lfloor n/2 \rfloor$ and charging candidate shortages. This policy inspects proposed values before committing them and therefore lies outside Definition 1.

J A MATCHING GROWING-DEPTH ORDER FOR POLYLOGARITHMIC BUDGETS

The finite batch estimates also yield a growing-depth upper bound when the KL budget decays sufficiently quickly as a power of $\log n$. We track the number of batches explicitly throughout; the fixed-depth limit in (53) is not used.

For a fixed accuracy $\varepsilon \in (0,1)$, define the minimax sufficient depth by

$$K_n^{\min}(B,\varepsilon) = \min\{K \in \mathbb{N} : K \geq 1,\ \mathcal{E}_n(K,B) \leq \varepsilon\},$$

with value $+\infty$ if the set is empty. This definition retains the quantifier order in (5): each model has its own fully informed policy.

Corollary 11 (Growing-depth order for polylogarithmic budgets)**.** *Let $\ell_n = \log(n+1)$ and $B_n = \ell_n^{-\zeta}$ for a fixed $\zeta > 11/3$. For any fixed*

$$C > \frac{1}{2(\zeta - 11/3)}, \qquad r_n = \left\lceil \frac{C\ell_n}{\log \ell_n} \right\rceil,$$

the geometric schedule (11), preceded by its representative round, satisfies

$$\sup_{R_E \in \mathcal{C}_n(B_n)} \mathrm{TV}(P_n, Q_R^{\mathrm{geom}}) \longrightarrow 0. \tag{86}$$

Consequently, for every fixed $\varepsilon \in (0,1)$,

$$K_n^{\min}(B_n,\varepsilon) = \Theta_{\zeta,\varepsilon}\left(\frac{\log n}{\log\log n}\right). \tag{87}$$

Proof. **A finite bound with explicit batch dependence.** Consider first any $r \geq 2$ for which $q = n^{1/r} \geq 2$. For the geometric schedule, $c_j = \lfloor q^j \rfloor$ for $1 \leq j < r$, $c_r = n$, and $d_j = c_j - c_{j-1}$. For every $2 \leq j \leq r$,

$$c_1 \leq q, \qquad c_{j-1} \geq \tfrac{1}{2}q^{j-1}, \qquad \frac{d_j}{c_{j-1}} \leq 2q, \qquad c_{j-1} \leq \frac{n}{q} \leq \frac{n}{2}, \qquad \frac{d_j}{n - c_{j-1}} \leq 1.$$

All batches are nonempty. Thus, for any $0 < t \leq \sqrt{c_1}/16$, (22) applies to every noninitial mismatch batch. Using (23) for the first batch and summing through the full-output hybrid (19) gives, uniformly over $R_E \in \mathcal{C}_n(B)$,

$$\begin{aligned}\mathrm{TV}(P_n, Q_R^{\mathrm{geom}}) &\leq 16(r-1)B + 2(r-1)e^{-2t^2} + (2r-1)\sqrt{2B} \\ &\quad + B\sqrt{q}\left[\frac{3}{4\sqrt{2v}} + \frac{4(r-1)(t+1)e^{t^2/v}}{\sqrt{v}}\right], \qquad v = \frac{3}{16}.\end{aligned} \tag{88}$$

This inequality is finite-dimensional and allows r and t to depend on n.

Choosing the growing schedule. Take $r = r_n$, $B = B_n$, and, for all sufficiently large n,

$$t_n^2 = \tfrac{1}{2}\log r_n + \log\log r_n.$$

Integer rounding gives

$$q_n = n^{1/r_n} = \ell_n^{1/C}(1 + o(1)), \qquad c_1 \asymp \ell_n^{1/C}, \qquad t_n^2 = O(\log \ell_n).$$

Hence $q_n \geq 2$ and $t_n \leq \sqrt{c_1}/16$ eventually, so all domain conditions used in (88) hold. Its exceptional-prefix term satisfies

$$2(r_n - 1)e^{-2t_n^2} \leq \frac{2}{(\log r_n)^2} \longrightarrow 0.$$

Also $r_n B_n \to 0$ and $r_n\sqrt{B_n} \to 0$, since $\zeta > 11/3 > 2$. For the largest loss-linear term, $1/v = 16/3$ gives

$$\begin{aligned}B_n\sqrt{q_n}\, r_n(t_n+1)e^{t_n^2/v} &= O\Big(B_n\sqrt{q_n}\, r_n^{11/3}(\log r_n)^{35/6}\Big) \\ &= O\Big(\ell_n^{11/3+1/(2C)-\zeta}(\log \ell_n)^{13/6}\Big) \longrightarrow 0.\end{aligned}$$

The exponent of ℓ_n is strictly negative by the choice of C. The remaining term $B_n\sqrt{q_n}$ also vanishes. This proves (86) and the upper bound in (87).

A lower witness within the same budget. Put

$$\rho_n = \left\lfloor \frac{\ell_n}{4(\zeta+3)\log \ell_n} \right\rfloor, \qquad w_n = \frac{B_n}{2}, \qquad \Delta_n = \ell_n^{-1}, \qquad \lambda_n = (\rho_n \ell_n)^{-1}.$$

Use the ladder (12) with $r = \rho_n$. Its parameters satisfy the assumptions of Theorem 3 eventually, and its loss obeys

$$D(q_{R_n}) \le -\log(1-w_n) \le \frac{w_n}{1-w_n} \le 2w_n = B_n.$$

The last inequality holds once $w_n \le 1/2$. Moreover,

$$\begin{aligned} \rho_n \lambda_n &= \ell_n^{-1}, \\ \log \frac{\rho_n}{w_n \lambda_n} &= 2\log \rho_n + (\zeta+1)\log \ell_n + \log 2 \\ &= (\zeta+3)\log \ell_n - 2\log\log \ell_n + O_\zeta(1), \\ \log\left[n\Delta_n^2 \left(\frac{w_n\lambda_n}{\rho_n} \right)^{2\rho_n} \right] &= \tfrac{1}{2}\log n + o(\log n) \longrightarrow \infty. \end{aligned}$$

The finite all-policy bound therefore yields $\mathcal{E}_n(\rho_n+1, B_n) \to 1$. For every fixed $\varepsilon < 1$, monotonicity in the round cap implies $K_n^{\min}(B_n, \varepsilon) > \rho_n + 1$ eventually, proving the lower bound in (87). □

The result matches the order of growth, while leaving its leading constant undetermined. The restriction $\zeta > 11/3$ comes from the accumulated central-window factor $r(t+1)e^{t^2/v}$ in the present upper estimate; it is a limitation of this proof, not an established boundary for sampling. For $0 < \zeta \le 11/3$, the growing-depth lower bound remains available, but this argument supplies no matching upper bound.

K Execution interfaces and structural scope

The order of selection and sampling matters for the lower bound. We first give a precise separation when a policy can inspect candidate values before committing a batch. We then isolate the role of exchangeability and give a sufficient guarantee in terms of mismatch dependence.

K.1 Three rounds with target-aware candidate selection

Enlarge Definition 1 as follows. In a round, a policy may independently propose one value at each unrevealed coordinate from its original head, inspect the proposals, and choose which proposed values to commit, using additional private randomness. Uncommitted proposals are discarded; committed values remain unchanged. Future heads use only the committed context. The policy retains the original assumption of complete knowledge of P and R.

Theorem 12 (Target-aware selection)**.** *Let $n \ge 2$, $m = \lfloor n/2 \rfloor$, and let $R_E \in \mathcal{C}_n(B)$ with $B \le 1/128$. There is a three-round policy in the enlarged interface whose output Q_R^{sel} satisfies*

$$\mathrm{TV}(P_n, Q_R^{\mathrm{sel}}) \le e^{-n/8} + \mathcal{D}_{n,m,n-m}. \tag{89}$$

Consequently, for $0 < t \le \sqrt{m}/16$,

$$\begin{aligned} \mathrm{TV}(P_n, Q_R^{\mathrm{sel}}) \le{}& e^{-n/8} + 16B + 2e^{-2t^2} + 2\sqrt{2B} \\ &+ \frac{4(t+1)}{\sqrt{v}} e^{t^2/v} B, \qquad v = 3/16. \end{aligned} \tag{90}$$

For every sequence $B_n \to 0$, these policies satisfy

$$\sup_{R_E \in \mathcal{C}_n(B_n)} \mathrm{TV}(P_n, Q_R^{\mathrm{sel}}) \longrightarrow 0. \tag{91}$$

Every execution finishes in three rounds and uses at most $3n$ coordinate proposals.

Proof. Write $q = R_E(E_i = 1)$, which is independent of i by exchangeability, and put

$$\alpha_{\mathrm{acc}} = \min\left\{\frac{q}{\tau}, \frac{1-q}{1-\tau}\right\}, \qquad a(e) = \alpha_{\mathrm{acc}} \frac{\mathrm{Ber}(\tau)(e)}{\mathrm{Ber}(q)(e)} \quad (e \in \{0,1\}). \tag{92}$$

Positivity gives $q \in (0,1)$, and $a(e) \in [0,1]$. Marginal contraction and Pinsker imply $|q - \tau| \le \sqrt{B/2} \le 1/16$. Hence $q \in [3/16, 5/16]$ and $\alpha_{\mathrm{acc}} \ge 3/4$.

In round one, reveal all A_i from their original heads. They have exactly their target law. In round two, independently propose every B_i and form the candidate mismatch $Y_i = A_i \oplus B_i$. Conditional on A, the Y_i are independent $\mathrm{Ber}(q)$ variables, and their law does not depend on A. Mark candidate i as accepted with probability $a(Y_i)$, independently across positions. If at least m candidates are accepted, choose a uniformly random m-subset of the accepted positions and commit those proposals. Otherwise commit no B in this round. In round three, generate every remaining B independently from the original heads at the committed context. This also specifies the failure branch and guarantees completion on every path.

Let C_i be the acceptance indicator and $G = \{\sum_i C_i \ge m\}$. For each $e \in \{0,1\}$,

$$\mathbb{P}(C_i = 1, Y_i = e) = \alpha_{\mathrm{acc}} \,\mathrm{Ber}(\tau)(e). \tag{93}$$

Thus the C_i are independent $\mathrm{Ber}(\alpha_{\mathrm{acc}})$ variables, and Hoeffding gives $\mathbb{P}(G^c) \le e^{-n/8}$. Conditional on the entire acceptance mask, the values at its accepted positions are independent $\mathrm{Ber}(\tau)$ variables. The mask is independent of A. Consequently, conditional on G, the selected set I is uniform among m-subsets, and $(E_i)_{i \in I}$ has the target product law independently of I and A. This statement concerns the retained values; acceptance depends on the raw candidate values through (92).

To compare full output laws on G, write $Q_R^{\mathrm{sel},G}$ for the output law conditional on G and augment an independent target sample with a uniform m-subset I. The target and execution now have the same joint law of (A, I, E_I). Given this common prefix, the target draws the remaining mismatches independently with rate τ, while execution uses the product of the original conditional heads. Exchangeability makes the conditional discrepancy depend only on the pilot count $\sum_{i \in I} E_i$. Averaging therefore gives $\mathrm{TV}(P_n, Q_R^{\mathrm{sel},G}) \le \mathcal{D}_{n,m,n-m}$. Charging G^c by its probability proves (89).

Apply (22) with $c = m$ and $d = n - m$. Since $d = n - c$ and $d/c \le 2$, it gives (90). For (91), fix $t > 0$ and take the limit superior as $n \to \infty$. The bound becomes $2e^{-2t^2}$, uniformly over the class. Letting $t \to \infty$ completes the proof. The proposal counts are n in each of the first two rounds and at most n in the last. □

For any fixed error tolerance, (90) also gives a positive sufficient budget independent of n, for all sufficiently large n. On the same vanishing-loss models as Corollary 5, target-aware selection therefore achieves vanishing error in three fixed rounds, while Definition 1 requires a growing expected depth to reach a fixed error below one. The acceptance probabilities use the known target rate to calibrate the retained pilot. This establishes an interface separation; it does not give the same guarantee for a prescribed confidence top-k rule or for a sampler without target knowledge.

K.2 A FIXED-LABEL OBSTRUCTION BEYOND EXCHANGEABILITY

Exchangeability allows the geometric schedule to choose labels without consulting R. To see why this feature needs an assumption, fix that schedule, put $r = K - 1$, and let S be its last mismatch batch, of size $n - c_{r-1}$. For $0 < w < 1$ and $0 < \Delta \le 1/4$, consider

$$R_E = (1-w)P_E + w\left[\mathrm{Ber}(\tau)^{\otimes S^c} \otimes \mathrm{Ber}(\tau + \Delta)^{\otimes S}\right]. \tag{94}$$

Keep the exact independent representatives. The two components agree on all mismatch coordinates generated before the last batch, so their posterior weights remain $(1-w, w)$. The last batch has independent mismatch rate $\tau + w\Delta$, giving

$$\mathrm{TV}(P_n, Q_R^{\mathrm{geom}}) \ge 1 - 2\exp[-|S|(w\Delta)^2/2], \qquad \mathrm{KL}(P_E \| R_E) \le -\log(1-w). \tag{95}$$

The count test proving (7), restricted to S, proves this bound. For fixed K, $|S| \asymp n$; with fixed Δ, the schedule fails when $nw^2 \to \infty$.

This example concerns the fixed-label schedule. A planner that knows R can place its pilot inside S, and a private random relabeling changes which informative coordinates appear early. Thus (95) does not refute the minimax exponent for arbitrary positive R_E. That larger-class question remains open here. Averaging coordinate permutations of R_E produces an exchangeable joint law, but its conditional heads generally differ from the original heads used with a random schedule: taking a conditional and averaging joint laws need not commute. A randomized-schedule upper therefore requires its own argument.

K.3 A DEPENDENCE-SENSITIVE TWO-ROUND GUARANTEE

Proposition 13 (Mismatch dependence)**.** *Allow any positive mismatch law R_E, with the same exact independent representatives as the target, and suppose* $\mathrm{KL}(P_E\|R_E) \le B$. *Define*

$$\mathrm{TC}(R_E) = \mathrm{KL}\left(R_E \middle\| \bigotimes_{i=1}^{n} (R_E)_i\right). \tag{96}$$

The two-round schedule that generates all A and then all B satisfies

$$\mathrm{TV}(P_n, Q_R^{AB}) \le \min\left\{1, \sqrt{B/2} + \sqrt{\mathrm{TC}(R_E)/2}\right\}. \tag{97}$$

If R_E is a product law, this schedule generates R_n exactly.

Proof. In (A, E) coordinates the output is $P_A \otimes \bigotimes_i (R_E)_i$. The triangle inequality through $P_A \otimes R_E$ and Pinsker applied to the two divergences in the statement prove (97). The product case follows directly from the output formula. □

This sufficient bound applies without exchangeability and isolates dependence among mismatches, excluding the target's inherent within-pair dependence. It gives a regime in which two-round execution is accurate; it does not determine a sharp threshold as a joint function of B and total correlation.

L OTHER ERROR METRICS AND REMAINING SCOPE

L.1 HELLINGER DISTANCE AND FORWARD KL

For probability laws P, Q on a common finite space, define their affinity and normalized squared Hellinger distance by

$$\mathrm{Aff}(P, Q) = \sum_x \sqrt{P(x)Q(x)}, \qquad h^2(P, Q) = 1 - \mathrm{Aff}(P, Q). \tag{98}$$

The elementary inequalities

$$h^2(P, Q) \le \mathrm{TV}(P, Q) \le \sqrt{1 - \mathrm{Aff}(P, Q)^2} = \sqrt{h^2(P, Q)(2 - h^2(P, Q))} \tag{99}$$

transfer both the vanishing-error and the error-to-one conclusions of Theorem 2 to h^2. In particular, if $B^{*,h^2}_{n,K}(\varepsilon)$ is defined by replacing TV with h^2 in (5), then

$$B^*_{n,K}(\varepsilon) \le B^{*,h^2}_{n,K}(\varepsilon) \le B^*_{n,K}\big(\sqrt{\varepsilon(2-\varepsilon)}\big). \tag{100}$$

Thus the same sharp dimension exponent holds for each fixed K and $\varepsilon \in (0, 1)$.

Jensen's inequality also gives

$$\mathrm{KL}(P\|Q) \ge -2\log \mathrm{Aff}(P, Q) \ge -\log\big(1 - \mathrm{TV}(P, Q)^2\big). \tag{101}$$

Consequently, the single-oracle supercritical lower bound forces forward KL to diverge for every allowed policy. A TV upper alone does not give a forward-KL upper, because KL is sensitive to events assigned very small output probability.

For two rounds, the upper can be proved directly. Let $q = R_E(E_i = 1)$ for an exchangeable model and $B \le 1/128$. Put $\overline{E} = n^{-1}\sum_i E_i$ and $g = \sqrt{dR_E/dP_E} - 1$. Using (20), exchangeability, and $\mathrm{Var}_P(\overline{E}) = v/n$ gives

$$\begin{aligned} |q - \tau| &= \left|2\mathbb{E}_P[(\overline{E} - \tau)g] + \mathbb{E}_P[(\overline{E} - \tau)g^2]\right| \\ &\le 2\sqrt{vB/n} + \tfrac{3}{4}B. \end{aligned} \tag{102}$$

The two-round output has independent mismatches of rate q. As $q \in [3/16, 5/16]$, the Bernoulli bound $\mathrm{kl}(\tau\|q) \le (\tau - q)^2/[q(1-q)]$ yields

$$\mathrm{KL}(P_n\|Q_R^{AB}) = n\,\mathrm{kl}(\tau\|q) \le 10(B + nB^2). \tag{103}$$

Together with (101) and Theorem 2, this proves the same sharp exponent $1/2$ for the two-round forward-KL budget at every fixed positive KL tolerance. For general $K > 2$, a matching forward-KL upper remains open here.

L.2 Coherence and more complex targets

For arbitrary positive, possibly incoherent heads, the population loss in (3) is still well-defined. Coherence supplies a common joint law and the identity $D(q_R) = \mathrm{KL}(P_n\|R_n)$ used in our upper bounds. Without it, that identification and the likelihood-ratio decomposition require a replacement argument. Our coherent lower witnesses also belong to any enlarged class allowing incoherent heads, but the present upper theorem does not cover that enlarged class.

The two-round target fixes the baseline depth in the sharp result. For targets whose exact heads require more than two rounds, neither an exponent obtained by shifting K nor a general reduction to the noisy-pair analysis is established here. Within-block dependence and the information carried by partial block observations would have to be controlled separately.

M Interpretation, related models, and parameter dependence

M.1 Which question does the accuracy threshold answer?

The planner knows both the target and the conditional model. The budget controls population prediction error, and the resource is the number of synchronous rounds needed to turn the prescribed heads into an accurate sample. Thus the result addresses robustness of a sampling interface to model error. It imposes no restrictions on the computation used to design a schedule, and gives no lower bound on the number of training examples or the optimization cost of fitting a model.

The quantifier order in (5) allows each model its own planner. A lower witness must defeat all such planners for the same model, while the upper theorem provides one schedule that works uniformly. The target is deliberately simple: its exact heads admit two-round sampling. This isolates the cost of executing inaccurate heads from the target's intrinsic dependence structure.

M.2 Comparison with nearby sampling theories

The conditional-total-correlation identity of Wen et al. (2026) can be applied to execution with exact heads of the model R and explains the dependence discarded within a batch. Our budget, however, is $\mathrm{KL}(P\|R)$ and our output is compared with P. Moreover, mixing policies changes the marginal output distribution: a lower bound on each deterministic policy's KL need not survive as a lower bound on the mixture's TV. Thus their identity does not directly yield our all-policy converse or the sharp neighborhood threshold. Proposition 13 records a simple sufficient guarantee obtained when mismatch total correlation itself is small.

Related motivations also arise outside scheduling theory. Gu et al. (2018) identify the difficulty of independently predicting output positions in multimodal translation distributions and introduce latent fertilities. This connects to the dependence lost when a mixture is replaced by products of its marginals. Liu et al. (2025) use a separate copula model to restore dependencies within a diffusion step, augmenting the original-head interface studied here. Feng et al. (2025) distinguish perplexity and sequence-error guarantees for diffusion language models; their metric-dependent conclusions motivate separating our full-output TV guarantee from token-level performance.

Table 10: The roles of the target, the error measure, and the prediction budget in three nearby formulations. The policy classes are described at the level used by the corresponding converse results.

	Chen et al. (2026)	Wen et al. (2026)	This work
Information	Unknown target accessed through queries	Known distribution with exact conditional marginals	Known target and coherent model
Policy	Randomized adaptive querying	Deterministic, value-adaptive batch selection	Randomized, value-adaptive batch selection
Error and resource	TV and inference-query complexity	Forward KL and target-averaged serial depth	Output TV and hard-cap or execution-averaged depth
Role of accuracy	Oracle/query formulation	Within-batch conditional total correlation	Population excess-loss neighborhood of a two-round target
Matched statement here	Different oracle problem	Dependence-sensitive execution characterization	Fixed-depth dimension exponent in the fair exchangeable class

The classical theory of finite exchangeability distinguishes finite symmetric laws from infinitely extendible product mixtures (Diaconis & Freedman, 1980). Our upper proof keeps the finite-population structure through total-count shells and the hypergeometric comparison of Dümbgen et al. (2021). It does not replace the model by an infinite de Finetti mixture.

M.3 What does a large output TV mean?

The lower bound detects a systematic mismatch-rate shift by aggregating over n pairs. Its test separates a mean shift of order $n\delta_r$ from fluctuations of order $\sqrt{n}$. Consequently, a per-pair drift tending to zero can still be distinguishable at the level of the full vector. This is the sense in which accurate local predictions can coexist with poor sequence-level sampling.

TV permits every measurable full-output test. It therefore addresses a stronger requirement than accuracy of a few selected marginals or an average token statistic. The theorem does not establish corresponding failures for every local or application-specific metric. Frequency, repetition, and length statistics in language generation are possible motivations for studying aggregate bias, but our experiments use synthetic pairs and make no empirical claim about these language statistics. Appendix L gives the conclusions that do follow for Hellinger distance and forward KL.

The averaged-loss threshold is $\Theta(n^{-1-1/[2(K-1)]})$. It is a worst-case sufficient precision when the loss budget is the only control on the model within our class. Structured errors can be much easier: product mismatch laws are generated exactly in two rounds, and (97) also covers small mismatch dependence. A sharp interpolation between this regime and the worst-case class remains an open question. A mixing-measure parameter would apply only to the subclass with a product-mixture representation; finite exchangeability by itself does not supply such a representation.

M.4 The choice of mismatch rate

Fixing $\tau = 1/4$ makes the constants transparent. The fair-pair argument extends to each fixed $\tau \in (0, 1/2)$ with $v = \tau(1-\tau)$ and $0 < \Delta \le 1/2 - \tau$. The first-round co-sample mean $1/2$ then dominates every component, as required by (16). In the upper proof, the initial mean estimate uses $|S/n - \tau| \le 1 - \tau$, and the shell argument uses the window condition $t/\sqrt{c} \le \tau/4$. Its likelihood-ratio and projection estimates retain the same powers of B, c, and n. The fixed-depth dimension exponent is therefore unchanged, with constants depending on τ.

This statement is not uniform at the endpoints. As $\tau \uparrow 1/2$, the available perturbation range Δ vanishes; at $\tau = 1/2$ the target has independent fair coordinates and one round of the model's fair

single-coordinate marginals is exact. As $\tau \downarrow 0$, the upper estimates involving $v^{-1/2}$ and $\exp(t^2/v)$ deteriorate. The heterogeneous-bias constants in Appendix E remain stated at $\tau = 1/4$.

M.5 Unresolved constants and the biased warm start

The constants in Theorem 2 may depend on both K and the desired error. The proof chooses a central window wide enough to make $2(r-1)e^{-2t^2}$ small, at the cost of the factor $e^{t^2/v}$ in the loss-linear term. This tradeoff explains the poor displayed dependence on the error tolerance. We do not determine the critical error curve at a fixed rescaled budget or the sharp leading constants. Appendix J tracks the finite estimates uniformly for one growing-depth regime.

The extra component in the biased lower construction supplies a first-round drift guarantee for informative B-only observations. In the fair model, these observations have identical likelihood under every component, so the proof can start with co-sample completions. A prescribed non-fair representative law changes this starting step. The resulting exponent gap is a limitation of the current bounds; we do not know whether a sharper warm-start argument closes it or whether some biases admit faster policies.