TEMPORAL ALIASING IS ARITHMETIC: OPTIMAL SOURCE-TIME DESIGN FOR UNSEEN-TIME CONTROL

Anonymous authors
Paper under double-blind review

ABSTRACT

A controlled finite-state continuous-time Markov process may be observed only through endpoint transitions at legal source holding times, while deployment uses an unqueryable duration. Let $\mathcal{T}(D)$ be the closed additive group generated by the source times. Equality of source skeletons propagates to $\mathcal{T}(D)$: if $\mathcal{T}(D) = \mathbb{R}$, the generator is identified, while $\mathcal{T}(D) = h\mathbb{Z}$ identifies one base skeleton and its compositional times. Under an explicit rate–time condition, every noncompositional target admits irreducible three-state models with identical adaptive source transcripts but opposite target actions. We characterize the resulting Bellman and arithmetic answer quotients. On finite observable quotients with known legal-source laws, a sparse target-excluding Track-and-Stop rule matches the answer-information lower bound, while Diophantine distance to harmful resonances controls finite-sample conditioning. Finally, on a coupled unknown-damping/logarithm-branch family with known parametric form, a profile-design confidence-sequence learner operates on raw categorical endpoints and attains the pointwise answer-information constant even though exact branch information is zero.

1 INTRODUCTION

We study an *endpoint generative model*: reset a controlled finite-state CTMC to a chosen state, hold one action for an allowed source duration, and observe only the endpoint. Jump times, intermediate states, and source rewards are hidden; the deployment reward is known; and the target duration is not queryable. This models a simulator or resettable test system, not a passive trajectory. If action a has generator Q_a, a source duration Δ reveals a row of $e^{\Delta Q_a}$, whereas deployment uses $e^{\tau_\star Q_a}$.

For $D = \{\Delta_1, \dots, \Delta_m\}$ define the raw integer span and its closed temporal information set

$$\mathcal{G}_0(D) := \left\{ \sum_{j=1}^{m} z_j \Delta_j : z_j \in \mathbb{Z} \right\}, \qquad \mathcal{T}(D) := \overline{\mathcal{G}_0(D)}. \tag{1}$$

Our paper-level claim is:

> Legal source times determine a temporal information group; under an explicit rate–time condition, targets outside a discrete source lattice can conceal the optimal action; and the statistically optimal learner resolves only aliases that cross the target-answer boundary. (2)

Contributions. (i) Temporal information law and control converse. Equality of source skeletons propagates to $\mathcal{T}(D)$. A dense closed group identifies the generator; a discrete group $h\mathbb{Z}$ identifies one base skeleton and every compositional target. Under $\Lambda h \geq 2\pi/\sqrt{3}$, every positive noncompositional target admits irreducible source-indistinguishable three-state models with opposite actions.

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

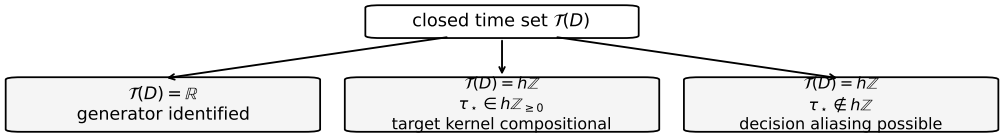

Figure 1: Complete temporal-group trichotomy. If $\mathcal{T}(D) = \mathbb{R}$, the generator is identified. If $\mathcal{T}(D) = h\mathbb{Z}$, compositional targets are identified; noncompositional targets can be decision-aliased under the stated dimension and bandwidth condition. The independent short-duration condition $\Delta\Lambda < \pi$ also identifies the generator.

(ii) Bellman/arithmetic answer quotient and optimal legal design. Fiberwise zero regret is equivalent to one policy being optimal throughout the observed source fiber. Bellman action-gap operators identify the control-relevant image of this fiber, and a circulant family yields exact integer-annihilator source and target quotients. On finite observable quotients with known legal-source laws, a sparse target-excluding adaptive rule matches the answer-information lower bound.

(iii) Stable end-to-end quotient learning. Diophantine distance to harmful rational resonances determines finite-sample answer information. On a coupled unknown-damping/logarithm-branch family with known parametric form, a profile-design confidence-sequence learner is not supplied the true damping, branch, endpoint probabilities, or quotient table; it localizes the observable fiber and attains the pointwise answer-information constant while exact branch information remains zero.

What should the learner do?

$\exists \Delta \in D:\ \Delta\Lambda < \pi$	$\Rightarrow$	reconstruct the generator;
$\mathcal{T}(D) = h\mathbb{Z},\ \tau_\star \in h\mathbb{Z}_{\geq 0}$	$\Rightarrow$	estimate e^{hQ_a} and compose;
otherwise	$\Rightarrow$	learn the target-answer quotient using legal sources.

(3)

Scope. The global fiber results are information theoretic. The exact legal-design theorem assumes a finite observable quotient and known quotient source laws; the end-to-end theorem removes that table on one compact continuous–discrete family. We do not claim tractability for arbitrary matrix-exponential fibers or exact complexity for overlapping correct-answer sets.

2 Model and the Closed Source-Time Group

Let $\mathcal{X} = [S]$, $\mathcal{A} = [A]$, and $\theta = (Q_a^\theta)_{a \in \mathcal{A}}$ belong to a nonempty compact subset of

$$\mathsf{Gen}_S(\Lambda) := \{Q : Q(x,y) \geq 0\ (x \neq y),\ Q\mathbf{1} = 0,\ \max_x[-Q(x,x)] \leq \Lambda\}.$$

A source query $e = (x, a, j)$ resets the process to x, holds a for Δ_j, and returns

$$Y \sim p_e^\theta, \qquad p_e^\theta(y) = e_x^\top e^{\Delta_j Q_a^\theta} e_y. \tag{4}$$

Queries may be adaptive. The source skeleton and fiber are

$$\Phi_{\mathcal{D}}(\theta) := (e^{\Delta_j Q_a^\theta})_{a,j}, \qquad F(s) := \{\vartheta : \Phi_{\mathcal{D}}(\vartheta) = s\}.$$

Deployment uses $e^{\tau_\star Q_a^\theta}$, a fixed known normalized finite-horizon task, and no target interaction. Let Π_H be the finite deterministic nonstationary Markov-policy class. We use uniform Bellman regret

$$\mathrm{Reg}_{\tau_\star}(\pi; \theta) := \max_{h,x}\{V_h^{\star,\theta}(x) - V_h^{\pi,\theta}(x)\},$$

which upper-bounds every fixed-start regret. A randomized output is an ex ante lottery over complete deterministic policies.

For temporal design, write $u = \Delta/h$ and $r = \tau_\star/h$. The legal source domain is a declared nonempty compact set

$$\mathcal{U}_{\mathrm{src}} \subset [u_{\min}, u_{\max}] \setminus (r - \varepsilon_{\mathrm{tgt}}, r + \varepsilon_{\mathrm{tgt}}), \qquad u_{\min} > 0, \quad \varepsilon_{\mathrm{tgt}} > 0. \tag{5}$$

Every non-oracle theorem, algorithm, and experiment samples only from $\mathcal{U}_{\mathrm{src}}$.

Theorem 2.1 (Closed source-time group law)**.** *Let $D = \{\Delta_1, \dots, \Delta_m\}$. If $e^{\Delta_j Q} = e^{\Delta_j \widetilde{Q}}$ for every j, then*

$$e^{tQ} = e^{t\widetilde{Q}} \qquad \forall t \in \mathcal{T}(D). \tag{6}$$

Exactly one of the following holds:

1. *$\mathcal{T}(D) = \mathbb{R}$, in which case $Q = \widetilde{Q}$;*
2. *$\mathcal{T}(D) = \mathcal{G}_0(D) = h\mathbb{Z}$ for a unique $h > 0$. Writing $\Delta_j = n_j h$ with $\gcd(n_1, \dots, n_m) = 1$, every Bézout vector $\sum_j b_j n_j = 1$ reconstructs*

$$e^{hQ} = \prod_j (e^{\Delta_j Q})^{b_j},$$

and hence every e^{nhQ}, $n \geq 0$.

The result holds actionwise for controlled generators.

Proof mechanism. For $t = \sum_j z_j \Delta_j \in \mathcal{G}_0(D)$, exponentials of one matrix commute and are invertible, so $e^{tQ} = \prod_j (e^{\Delta_j Q})^{z_j}$; continuity gives equation 6. A closed additive subgroup of $\mathbb{R}$ either has arbitrarily small positive elements and equals $\mathbb{R}$, or has a least positive element h and equals $h\mathbb{Z}$. In the dense case, take $t_n \downarrow 0$ in $\mathcal{G}_0(D)$ and use $Q = \lim_n (e^{t_n Q} - I)/t_n$. Appendix A gives the complete proof.

Proposition 2.2 (Short-duration injectivity)**.** *If $\Delta\Lambda < \pi$, then $Q \mapsto e^{\Delta Q}$ is injective on $\mathsf{Gen}_S(\Lambda)$.*

Gershgorin places the spectrum of ΔQ in the principal-logarithm strip, so $\mathrm{Log}(e^{\Delta Q}) = \Delta Q$. This spectral mechanism and the multirate dense-group mechanism are complementary. Exact identification need not be statistically stable: Bézout products may use ill-conditioned inverses, and dense-group recovery differentiates semigroup values near zero.

3 Rate-Qualified Aliasing at Every Noncompositional Target

Assume $\mathcal{T}(D) = h\mathbb{Z}$ and let $r := \tau_\star/h \notin \mathbb{Z}$. Define

$$J = \frac{1}{3}\begin{bmatrix} -2 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{bmatrix}, \qquad A_{\mathrm{rot}} = \frac{\pi}{\sqrt{3}}\begin{bmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{bmatrix}.$$

Choose $\alpha \geq \pi/\sqrt{3}$, $\beta > 0$, and $2\max\{\alpha, \beta\}/h \leq \Lambda$. Feasibility at the boundary requires $\Lambda h \geq 2\pi/\sqrt{3}$; strictly positive risky off-diagonal rates require strict inequality. Let

$$Q_{\mathrm{safe}} = 3\beta J/h, \qquad Q^{\pm}_{\mathrm{risk}} = h^{-1}(3\alpha J \pm A_{\mathrm{rot}}).$$

Every action generator is irreducible.

Theorem 3.1 (Rate-qualified decision aliasing at every noncompositional target)**.** *For the normalized one-step task with terminal reward $g = (1/2, 1, 0)^\top$, the two risky generators can be labeled so that their complete adaptive source-transcript laws agree for every budget, but their unique target actions at state 1 are opposite. Their common action gap is*

$$\gamma_{\alpha,r} = \frac{e^{-3\alpha r}|\sin(\pi r)|}{\sqrt{3}} > 0, \tag{7}$$

and every N-query learner satisfies

$$\inf_{\mathsf{A}} \sup_{\theta \in \{\theta_+, \theta_-\}} \mathbb{E}_\theta \operatorname{Reg}_{\tau_\star}(\widehat{\mu}_N; \theta) \geq \frac{\gamma_{\alpha,r}}{2}. \tag{8}$$

No almost-surely finite source-only procedure is uniformly η-correct for $\eta < 1/2$.

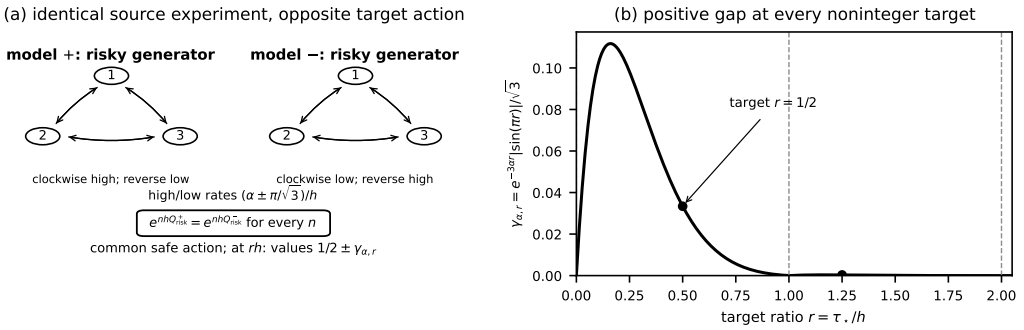

Figure 2: Rate-qualified arbitrary-target aliasing. Under $\Lambda h \geq 2\pi/\sqrt{3}$, the risky generators exchange the physical rates $(\alpha \pm \pi/\sqrt{3})/h$, agree at every source time nh, and straddle the common safe value at every noninteger target ratio.

The closed algebra $J^2 = -J$, $JA_{\mathrm{rot}} = A_{\mathrm{rot}}J = -A_{\mathrm{rot}}$, and $A_{\mathrm{rot}}^2 = \pi^2 J$ gives

$$e^{tQ^{\pm}_{\mathrm{risk}}} = I + \left[1 - e^{-3\alpha t/h}\cos(\pi t/h)\right] J \pm e^{-3\alpha t/h}\frac{\sin(\pi t/h)}{\pi} A_{\mathrm{rot}}. \tag{9}$$

At every source time nh, the rotating term vanishes. Hence every conditional endpoint law agrees, and induction over the adaptive history gives equal transcript laws. At rh, the safe value is $1/2$, while the risky values are $1/2 \pm e^{-3\alpha r}\sin(\pi r)/\sqrt{3}$; relabeling by the sine sign gives opposite actions. If the common output law chooses risky with probability p, the two regrets are at least $(1-p)\gamma_{\alpha,r}$ and $p\gamma_{\alpha,r}$, proving equation 8. Equality of every finite prefix gives the stopped-time claim. Appendix A proves the theorem and shows that three states are minimal for exact generator aliasing under full endpoint observation from every starting state at one fixed positive duration.

4 Fiberwise Decisions and the Bellman Quotient

For a realizable source skeleton s, define its robust decision radius

$$\mathfrak{a}_\star(s) := \min_{\mu \in \Delta(\Pi_H)} \max_{\theta \in F(s)} \mathrm{Reg}_{\tau_\star}(\mu; \theta).$$

Theorem 4.1 (Fiberwise oracle decision)**.** *Under finite state, action, and horizon sets, compact Θ, a fixed known normalized target task, and endpoint generative source access,*

$$\mathfrak{R}^\star_{\mathrm{or}} = \max_{s \in \Phi_{\mathcal{D}}(\Theta)} \mathfrak{a}_\star(s).$$

Moreover, $\mathfrak{R}^\star_{\mathrm{or}} = 0$ iff every fiber admits one common target-optimal policy:

$$\bigcap_{\theta \in F(s)} \Pi^\star_{\tau_\star}(\theta) \neq \varnothing \qquad \forall s.$$

Balanced finite-query estimation and robust planning over shrinking skeleton confidence sets converge to this oracle value.

The lower bound holds because every source rule induces one output law on a fiber; measurable ε-optimal fiberwise selectors give the converse. Randomization cannot rescue zero risk: deterministic-policy regrets are nonnegative, so every policy in the lottery support must be optimal throughout the fiber. Compactness and continuity then yield finite-query closure without assuming a continuous inverse of the matrix exponential. The construction is information theoretic; no efficient optimizer for arbitrary generator fibers is claimed.

For target kernels $P^\theta_a = e^{\tau_\star Q^\theta_a}$, a continuation space $\mathcal{V} \subseteq \mathbb{R}^S$, and reference action a_0, define

$$\mathfrak{B}_{\mathcal{V}}(\theta) := \left((P^\theta_a - P^\theta_{a_0})|_{\mathcal{V}}\right)_{a \neq a_0}.$$

Theorem 4.2 (One-step Bellman quotient)**.** *For two models, the following are equivalent over the universal one-step task class with arbitrary real action offsets and continuations $v \in \mathcal{V}$: (i) all pairwise action gaps agree; (ii) every greedy-action correspondence agrees; (iii) the model difference is an action-independent linear shift on $\mathcal{V}$; and (iv) $\mathfrak{B}_{\mathcal{V}}(\theta) = \mathfrak{B}_{\mathcal{V}}(\vartheta)$. If $\mathbf{1} \in \mathcal{V}$, every failure has a nonnegative normalized separating task.*

Any component shared by every action cancels from all action comparisons; conversely, a failed pairwise equality is separated by placing an action-reward offset between the two continuation gaps. A macro–micro Kronecker-sum construction then gives source-equivalent models with distinct target micro kernels but identical control for every macro-measurable task. Hence generator identification, target-kernel identification, and target-policy coherence are strictly different requirements. Appendix B contains the normalized witness and a sufficient, not necessary, finite-horizon certificate.

5 Controlled Arithmetic Answer Quotients

Fix a finite branch set $\mathcal{K} \subset \mathbb{Z}$ and

$$\alpha \geq \frac{\pi}{\sqrt{3}} \max_{k \in \mathcal{K}} |2k+1|, \qquad \frac{2\alpha}{h} \leq \Lambda. \tag{10}$$

For every $k \in \mathcal{K}$, define the controlled model

$$\theta_k = (Q_{\mathrm{safe}}, Q^{(k)}_{\mathrm{risk}}), \qquad Q_{\mathrm{safe}} := 0, \qquad Q^{(k)}_{\mathrm{risk}} := h^{-1}\big(3\alpha J + (2k+1)A_{\mathrm{rot}}\big). \tag{11}$$

The target is the normalized one-step task with zero stage reward and terminal reward $g = (1/2, 1, 0)^\top$. For normalized source ratios $\mathcal{U} = (u_1, \dots, u_m)$ and target ratio $r > 0$, write

$$\mathrm{Ann}(\mathcal{U}) = \{d \in \mathbb{Z} : du_j \in \mathbb{Z}\ \forall j\}, \qquad \mathrm{Ann}(r) = \{d \in \mathbb{Z} : dr \in \mathbb{Z}\}.$$

Theorem 5.1 (Controlled arbitrary-target arithmetic quotient)**.** *Let $P_{\alpha,k}(u) = e^{huQ^{(k)}_{\mathrm{risk}}}$ and $\mathsf{S}_{\mathcal{U}}(k) = (P_{\alpha,k}(u_j))_{j=1}^m$. Then*

$$\|P_{\alpha,k}(u) - P_{\alpha,\ell}(u)\|_{\mathrm{F}}^2 = 8e^{-6\alpha u}\sin^2(\pi(k-\ell)u). \tag{12}$$

Consequently,

$$\mathsf{S}_{\mathcal{U}}(k) = \mathsf{S}_{\mathcal{U}}(\ell) \iff k - \ell \in \mathrm{Ann}(\mathcal{U}), \qquad P_{\alpha,k}(r) = P_{\alpha,\ell}(r) \iff k - \ell \in \mathrm{Ann}(r),$$

and the target kernel is identified on every source fiber iff

$$\mathrm{Ann}(\mathcal{U}) \cap (\mathcal{K} - \mathcal{K}) \subseteq \mathrm{Ann}(r). \tag{13}$$

For $a = e^{-3\alpha r}$ and $\vartheta_k = (2k+1)\pi r$, the risky-minus-safe Bellman gaps are

$$G_1(k, r) = \frac{a}{\sqrt{3}}\sin\vartheta_k, \tag{14}$$

$$G_2(k, r) = \frac{1}{2}\left[a\left(\cos\vartheta_k - \frac{\sin\vartheta_k}{\sqrt{3}}\right) - 1\right] < 0, \tag{15}$$

$$G_3(k, r) = \frac{1}{2}\left[1 - a\left(\cos\vartheta_k + \frac{\sin\vartheta_k}{\sqrt{3}}\right)\right] > 0. \tag{16}$$

Thus safe is uniquely optimal at state 2, risky is uniquely optimal at state 3, and at state 1 the optimal-action set is risky, safe, or both according as G_1 is positive, negative, or zero. A source fiber is common-policy coherent iff it contains no positive-G_1 branch and negative- G_1 branch simultaneously; zero-gap branches are compatible with either strict side.

Proof mechanism. Fourier diagonalization gives the branch semigroup, and Frobenius orthogonality of J and A_{rot} gives equation 12; equality is therefore precisely the integer annihilator condition. Multiplying the closed-form kernel by g gives the three gaps. Admissibility implies $3\alpha r \geq \sqrt{3}|\vartheta_k|$, while

$$\left|\cos z \pm \frac{\sin z}{\sqrt{3}}\right| \leq 1 + \frac{|z|}{\sqrt{3}} \leq e^{|z|/\sqrt{3}}.$$

Since $\vartheta_k \neq 0$, this proves $G_2 < 0 < G_3$. Thus the action ordering comes from the explicit Bellman gaps, not irreducibility. Appendix C gives the full derivation and the minimax consequence for incoherent fibers.

If $r = p/q$ is reduced, $\mathrm{Ann}(r) = q\mathbb{Z}$; if r is irrational, $\mathrm{Ann}(r) = \{0\}$. Half-time parity is the case $q = 2$. Kernel equivalence has exact subgroup structure, whereas the answer quotient is its image through the three-valued Bellman-margin map and may be strictly coarser.

6 Optimal Legal Design and Temporal Conditioning

Let $\mathcal{E}_{\mathrm{src}} := \mathcal{X} \times \mathcal{A} \times \mathcal{U}_{\mathrm{src}}$. Start from a finite latent family, collapse models having identical laws on every legal experiment, and let $\mathcal{Q}$ be the resulting finite observable quotient. In this section the learner knows $\mathcal{Q}$, every legal-source law p_e^q, the strict answer map $c : \mathcal{Q} \to \mathcal{C}$, and a valid target policy π_c for each answer; only the true q is unknown. Put

$$\mathcal{H}(q) := \{q' : c(q') \neq c(q)\}, \qquad D_\xi(q, q') := \int D_{\mathrm{KL}}(p_e^q \| p_e^{q'})\, \xi(de),$$

and, for $\mathcal{H}(q) \neq \varnothing$,

$$C^\star_{\mathrm{ans}}(q) := \sup_{\xi \in \mathcal{P}(\mathcal{E}_{\mathrm{src}})} \inf_{q' \in \mathcal{H}(q)} D_\xi(q, q'). \tag{17}$$

Assume continuous endpoint laws, model-independent support, a uniform positive probability bound, and legal separation of distinct quotient classes. An optimal design with at most $|\mathcal{H}(q)| + 1$ support points exists. Fix the canonical optimizer $\xi_q^\star$ obtained by taking the lexicographically smallest sorted sparse optimal representation under a fixed order on (x, a, u).

For observations $(E_s, Y_s)_{s \le t}$ define

$$\ell_t(q) := \sum_{s=1}^{t} \log p^q_{E_s}(Y_s), \qquad \widehat{q}_t := \min_{\preceq} \arg\max_{q \in \mathcal{Q}} \ell_t(q),$$

$$Z_t(q, q') := \ell_t(q) - \ell_t(q'), \qquad \widehat{\xi}_t := \frac{1}{t} \sum_{s=1}^{t} \delta_{E_s}. \tag{18}$$

Let ξ_0 be a legal finite-support design separating every pair, $\gamma_t = (t+1)^{-1/3}$, and

$$\lambda_{t+1} := (1 - \gamma_t)\xi^\star_{\widehat{q}_t} + \gamma_t \xi_0, \qquad b_\eta := \log\left(\frac{|\mathcal{Q}| - 1}{\eta}\right).$$

Sample $E_{t+1} \sim \lambda_{t+1}$ and stop when

$$\tau_\eta := \inf\left\{ t \ge 1 : \min_{q' \in \mathcal{H}(\widehat{q}_t)} Z_t(\widehat{q}_t, q') \ge b_\eta \right\}. \tag{19}$$

Return $\widehat{c}_{\tau_\eta} := c(\widehat{q}_{\tau_\eta})$ and $\widehat{\pi}_{\tau_\eta} := \pi_{\widehat{c}_{\tau_\eta}}$. If the family has one answer class, stop immediately.

Theorem 6.1 (Exact target-excluding design complexity)**.** *Every uniformly η-correct legal source rule, meaning $\Pr_q(\widehat{c}_\tau = c(q)) \ge 1 - \eta$ for every q, satisfies*

$$\mathbb{E}_q[\tau] \ge \frac{\mathrm{kl}(1 - \eta, \eta)}{C^\star_{\mathrm{ans}}(q)}. \tag{20}$$

The supremum in equation 17 is attained and obeys

$$C^\star_{\mathrm{ans}}(q) = \min_{\lambda \in \Delta(\mathcal{H}(q))} \max_{e \in \mathcal{E}_{\mathrm{src}}} \sum_{q' \in \mathcal{H}(q)} \lambda_{q'} D_{\mathrm{KL}}(p_e^q \| p_e^{q'}). \tag{21}$$

The rule above is time-uniformly η-correct and, under true q,

$$\widehat{q}_t = q \quad \textit{eventually a.s.}, \tag{22}$$

$$\widehat{\xi}_t \Rightarrow \xi_q^\star \quad \textit{a.s.}, \tag{23}$$

$$\frac{1}{t} \min_{q' \in \mathcal{H}(q)} Z_t(q, q') \to C^\star_{\mathrm{ans}}(q) \quad \textit{a.s.}, \tag{24}$$

$$\frac{\mathbb{E}_q[\tau_\eta]}{\log(1/\eta)} \longrightarrow \frac{1}{C^\star_{\mathrm{ans}}(q)}. \tag{25}$$

Every sampled duration is legal, so target exclusion holds pathwise.

The lower bound normalizes the expected stopped sampling measure and applies binary data processing to the returned answer. Compactness, Sion minimax, and Carathéodory give existence, duality, and sparse support. Forced legal exploration stabilizes the quotient MLE; bounded likelihood increments give allocation tracking, evidence growth, summable lower tails, and uniform integrability. Appendix C proves each step. This is a continuous-experiment temporal specialization of fixed-confidence optimal design and Track-and-Stop (Chernoff, 1959; Garivier & Kaufmann, 2016).

For the circulant family, define the harmful displacements and conditioning quantities

$$\mathcal{D}_k := \{k-\ell : c(\ell) \neq c(k)\}, \qquad D_k^{\max} := \max_{d\in\mathcal{D}_k} |d|,$$
$$\kappa_{\mathrm{ans}}(k,u) := \min_{d\in\mathcal{D}_k} \|du\|_{\mathbb{Z}}, \qquad \nu_u := \min_{j\in\mathcal{K},\, x,y} P_{\alpha,j}(u)(x,y),$$
$$c_{\mathrm{BA}}(u) := \inf_{n\geq 1} n\|nu\|_{\mathbb{Z}}.$$

If the legal jitter set is $\mathcal{I} = \bigcup_{j=1}^{m_{\mathcal{I}}} I_j$, define

$$A_{\mathcal{I}}(\mathcal{D}_k) := \sum_{d\in\mathcal{D}_k} \left(2 + \frac{4m_{\mathcal{I}}}{|d|\,|\mathcal{I}|}\right). \tag{26}$$

Let $C_{\mathrm{ans}}(k,u) := \min_{\ell:c(\ell)\neq c(k)} D_{\mathrm{KL}}(p_{1,k}(u)\|p_{1,\ell}(u))$. Empty $\mathcal{D}_k$ means the answer is known and no samples are required.

Theorem 6.2 (Finite-sample temporal conditioning)**.** *For every legal $u > 0$ with nonempty $\mathcal{D}_k$,*

$$\frac{16}{3} e^{-6\alpha u} \kappa_{\mathrm{ans}}(k,u)^2 \leq C_{\mathrm{ans}}(k,u) \leq \frac{8\pi^2}{3\nu_u} e^{-6\alpha u} \kappa_{\mathrm{ans}}(k,u)^2. \tag{27}$$

If $u = a/L + \varepsilon$ and a harmful d is divisible by L, then every uniformly η-correct fixed-clock learner satisfies

$$\mathbb{E}_k[\tau] \geq \frac{3\nu_u e^{6\alpha u}}{8\pi^2 d^2 \varepsilon^2} \,\mathrm{kl}(1-\eta,\eta). \tag{28}$$

For uniform U on $\mathcal{I}$, with probability at least $1-\delta$,

$$C_{\mathrm{ans}}(k,U) \geq \frac{16}{3} e^{-6\alpha u_{\max}} \left(\frac{\delta}{A_{\mathcal{I}}(\mathcal{D}_k)}\right)^2. \tag{29}$$

Every badly approximable legal clock satisfies

$$C_{\mathrm{ans}}(k,u) \geq \frac{16}{3} e^{-6\alpha u} \frac{c_{\mathrm{BA}}(u)^2}{(D_k^{\max})^2}. \tag{30}$$

The proof combines the exact row separation with $2\|x\|_{\mathbb{Z}} \leq |\sin(\pi x)| \leq \pi\|x\|_{\mathbb{Z}}$. Thus irrationality governs exact identifiability, whereas Diophantine quality governs finite-sample identifiability.

7 Stable End-to-End Quotient Learning

We remove the supplied quotient table on a compact family in which a continuous nuisance and logarithm branch affect the same categorical source law. Let

$$\theta = (\alpha, k), \quad \alpha \in [\alpha_-, \alpha_+], \quad k \in \{-3,-2,1,2\},$$

with legal source ratios $\{1/4, 3/4\}$ and target ratio $1/2$. The learner knows this parametric CTMC family, the legal durations, and the target reward, but not the true damping, branch, endpoint probabilities, or observable quotient table. Define

$$Q_{\mathrm{safe}} = 0, \qquad Q_{\mathrm{risk}}^{\alpha,k} = h^{-1}(3\alpha J + (2k+1)A_{\mathrm{rot}}).$$

Scale the terminal reward by R and give the safe action state-1 reward b. Assume

$$\alpha_- > \frac{5\pi}{\sqrt{3}}, \qquad \frac{2\alpha_+}{h} \leq \Lambda, \qquad R + b \leq 1, \tag{31}$$

$$\frac{Re^{-3\alpha_+/2}}{\sqrt{3}} + \Delta < b < \frac{Re^{-3\alpha_-/2}}{\sqrt{3}} - \Delta. \tag{32}$$

The state-1 target margin and strict parameter set are

$$\Gamma(\alpha, k) := \frac{Re^{-3\alpha/2}}{\sqrt{3}}(-1)^k - b,$$
$$\Theta_\Delta := \{(\alpha, k) : \alpha \in [\alpha_-, \alpha_+],\ k \in \{-3, -2, 1, 2\},\ |\Gamma(\alpha, k)| \geq \Delta\}. \tag{33}$$

Both coordinates enter each observed three-category endpoint law. Exact legal-source fibers are $\{(\alpha, -3), (\alpha, 1)\}$ and $\{(\alpha, -2), (\alpha, 2)\}$, and each fiber has one strict answer.

For $w \in \Delta_2$, define the answer-information objective

$$\mathcal{I}(\theta, w) := \inf_{c(\vartheta) \neq c(\theta)} \sum_{e=1}^{2} w_e D_{\mathrm{KL}}(p_e^\theta \| p_e^\vartheta), \qquad C^\star_{\mathrm{ans}}(\theta) := \max_{w \in \Delta_2} \mathcal{I}(\theta, w), \tag{34}$$

and $\mathcal{W}^\star(\theta) := \arg\max_w \mathcal{I}(\theta, w)$. Given data $(E_s, Y_s)_{s \leq t}$, let

$$\ell_t(\vartheta) := \sum_{s=1}^{t} \log p^\vartheta_{E_s}(Y_s), \qquad \widehat{\theta}_t := \min_{\preceq} \arg\max_{\vartheta \in \Theta_\Delta} \ell_t(\vartheta).$$

With $\rho_t = (t+1)^{-1/6}$, $\gamma_t = (t+1)^{-1/3}$, and $w_0 = (1/2, 1/2)$, set

$$w_t := \arg\max_{w \in \Delta_2} \{\mathcal{I}(\widehat{\theta}_t, w) - \rho_t \|w\|_2^2\}, \qquad E_{t+1} \sim (1-\gamma_t) w_t + \gamma_t w_0, \qquad \widehat{w}_t := \frac{1}{t} \sum_{s=1}^{t} e_{E_s}. \tag{35}$$

For counts $N_{e,y}(t)$ and $N = \sum_y n_y$, define $m(n_1, n_2, n_3) := 2 \prod_y n_y!/(N+2)!$ and

$$M_t(\vartheta) := \prod_{e=1}^{2} \frac{m((N_{e,y}(t))_y)}{\prod_y p_e^\vartheta(y)^{N_{e,y}(t)}}, \qquad \mathcal{C}_t(\eta) := \{\vartheta : M_t(\vartheta) < \eta^{-1}\}. \tag{36}$$

Stop at the first t for which $\{c(\vartheta) : \vartheta \in \mathcal{C}_t(\eta)\}$ has at most one element; return that answer, using $c(\widehat{\theta}_t)$ only if the confidence set is empty.

Theorem 7.1 (Coupled end-to-end temporal quotient learning)**.** *The procedure is uniformly η-correct and stops almost surely. For every true $\theta \in \Theta_\Delta$,*

$$\mathrm{dist}_{\mathrm{src}}(\widehat{\theta}_t, F(\theta)) \to 0 \qquad \textit{a.s.}, \tag{37}$$
$$\mathrm{dist}(\widehat{w}_t, \mathcal{W}^\star(\theta)) \to 0 \qquad \textit{a.s.}, \tag{38}$$
$$\frac{1}{t} \inf_{c(\vartheta) \neq c(\theta)} \log M_t(\vartheta) \to C^\star_{\mathrm{ans}}(\theta) \qquad \textit{a.s.}, \tag{39}$$
$$\frac{\mathbb{E}_\theta[\tau_\eta]}{\log(1/\eta)} \to \frac{1}{C^\star_{\mathrm{ans}}(\theta)} \qquad \textit{for each fixed } \theta \textit{ as } \eta \downarrow 0. \tag{40}$$

Moreover,

$$\boxed{C^\star_{\mathrm{branch}}(\theta) = 0 < C^\star_{\mathrm{ans}}(\theta).} \tag{41}$$

Thus the result is pointwise first-order asymptotically optimal on Θ_Δ, not uniform at the answer boundary.

The source-row radius identifies α, while the two legal phases identify k only modulo four. Dirichlet-mixture martingales give anytime coverage; on that coverage event the confidence set is never empty, so the fallback is used only on the failure event. Forced exploration localizes the observable fiber, and strict regularization gives a unique continuous design selector. A uniform likelihood-and-compensator net argument yields answer-evidence growth; a formal stable-design-time lemma and summable lower tails give uniform integrability. Appendix D proves measurability, continuity, the matching lower bound, and every limit. Mixture-martingale and Track-and-Stop methods are tools; the coupled temporal quotient and zero-branch/positive-answer separation are specific to this problem (Kaufmann & Koolen, 2021; Garivier & Kaufmann, 2016).

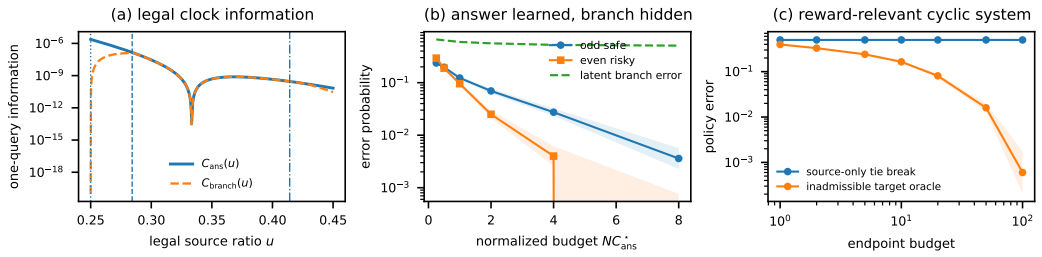

Figure 3: Three theorem-driven mechanisms. (a) A legal quotient-matched clock retains same-answer aliases. (b) Answer error vanishes while branch error remains near one half. (c) In a reward-relevant four-state cycle, source-only tie-breaking cannot beat one-half error; the target-data curve is an inadmissible oracle.

8 Theorem-Driven Experiments

All non-oracle studies use reset endpoint observations at legal source ratios, master seed 20270925, and 95% intervals; no jump time, intermediate state, or target-duration transition is observed. They illustrate the predicted mechanisms and are not used as proof.

For target ratio $1/2$ and branches $\{-3, -2, 1, 2\}$, the legal clock $u = 1/4$ has $C_{\text{ans}} = 2.36 \times 10^{-6}$ and $C_{\text{branch}} = 0$, whereas the branch-optimal legal clock has $C_{\text{ans}} = 1.45 \times 10^{-7}$ and $C_{\text{branch}} = 1.27 \times 10^{-7}$. Thus answer-aware design gives $16.3\times$ more answer information while retaining a harmless alias. At $NC^{\star}_{\text{ans}} \approx 8$, coupled-family answer errors are 0.0036 and 0, branch errors are 0.507 and 0.493, mean generator error exceeds 17, and mean target-kernel error is below 4×10^{-7}. Appendix E reports the actual stopping rule over five confidence levels, all metrics, raw outputs, and supplementary plots.

9 Related Work, Scope, and Conclusion

Positioning. Nonuniform and multirate sampling can remove continuous-time aliases (Ding et al., 2009; Crommelin & Vanden-Eijnden, 2009; Yue et al., 2016; Baake & Sumner, 2020), and recent Koopman work studies irregular-sampling cancellation (Cho & Sowers, 2026). We do not claim multirate reconstruction or nonunique logarithms as new. Our contribution is their unseen-time control synthesis: adaptive source equivalence, a rate-qualified action converse, Bellman images of the source fiber, and answer-optimal legal design. Partial identification theory studies robust decisions over identified sets (Yata, 2021; Montiel Olea et al., 2026); robust MDPs instead begin with externally specified ambiguity sets (Iyengar, 2005; Nilim & El Ghaoui, 2005), while value equivalence studies Bellman preservation in a specified family (Grimm et al., 2020; 2021).

Scope and conclusion. Track-and-Stop and mixture martingales are established tools (Garivier & Kaufmann, 2016; Kaufmann & Koolen, 2021); overlapping answers require a different game (Degenne & Koolen, 2019). Our exact complexity results use strict answers. The group trichotomy is complete, but the bounded-rate gap remains open. Legal design assumes known finite-quotient laws; the coupled theorem removes the table only for one compact family, where answers are learned without identifying the logarithm branch.

AI USE STATEMENT

Generative AI tools were used for research ideation, literature-search assistance, manuscript editing, algebraic checking, code drafting, and critical review of proof arguments. Every AI-assisted mathematical claim, proof, and empirical script included in this submission was independently rederived and verified by the authors. Deterministic checks and fixed-seed reproductions are described in Appendix F. The authors take full responsibility for the text, claims, proofs, citations, code, and artifacts.

REPRODUCIBILITY STATEMENT

Complete proofs are provided in Appendices A– D. Appendix E specifies every experimental family, legal duration, baseline, seed, budget, confidence interval, and failure rule. The anonymous archive regenerates all figures and tables, runs theorem consistency checks, compiles the paper, audits citations and anonymity, and records source hashes with one command.

REFERENCES

Michael Baake and Jeremy Sumner. Notes on Markov embedding. *Linear Algebra and its Applications*, 594:262–299, 2020. doi: 10.1016/j.laa.2020.02.016.

Herman Chernoff. Sequential design of experiments. *The Annals of Mathematical Statistics*, 30(3): 755–770, 1959. doi: 10.1214/aoms/1177706205.

Younghwan Cho and Richard Sowers. Koopman representations for non-vanishing time intervals: An optimization approach and sampling effects, 2026.

Daan T. Crommelin and Eric Vanden-Eijnden. Data-based inference of generators for markov jump processes using convex optimization. *Multiscale Modeling & Simulation*, 7(4):1751–1778, 2009. doi: 10.1137/080735977.

Rémy Degenne and Wouter M. Koolen. Pure exploration with multiple correct answers. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Feng Ding, Li Qiu, and Tongwen Chen. Reconstruction of continuous-time systems from their non-uniformly sampled discrete-time systems. *Automatica*, 45(2):324–332, 2009. doi: 10.1016/j.automatica.2008.08.007.

Aurélien Garivier and Emilie Kaufmann. Optimal best arm identification with fixed confidence. In *Proceedings of the 29th Conference on Learning Theory*, volume 49 of *Proceedings of Machine Learning Research*, pp. 998–1027. PMLR, 2016.

Christopher Grimm, André Barreto, Satinder P. Singh, and David Silver. The value equivalence principle for model-based reinforcement learning. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

Christopher Grimm, André Barreto, Greg Farquhar, David Silver, and Satinder P. Singh. Proper value equivalence. In *Advances in Neural Information Processing Systems*, volume 34, pp. 7773–7786, 2021.

Nicholas J. Higham. *Functions of Matrices: Theory and Computation*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2008. doi: 10.1137/1.9780898717778.

Garud N. Iyengar. Robust dynamic programming. *Mathematics of Operations Research*, 30(2): 257–280, 2005. doi: 10.1287/moor.1040.0129.

Emilie Kaufmann and Wouter M. Koolen. Mixture martingales revisited with applications to sequential tests and confidence intervals. *Journal of Machine Learning Research*, 22(246):1–44, 2021.

José Luis Montiel Olea, Chen Qiu, and Jörg Stoye. Decision theory for treatment choice problems with partial identification. *The Review of Economic Studies*, pp. rdag015, 2026. doi: 10.1093/restud/rdag015.

Arnab Nilim and Laurent El Ghaoui. Robust control of markov decision processes with uncertain transition matrices. *Operations Research*, 53(5):780–798, 2005. doi: 10.1287/opre.1050.0216.

Kohei Yata. Optimal decision rules under partial identification, 2021. Revised March 2025.

Zuogong Yue, Johan Thunberg, and Jorge Gonçalves. Inverse problems for matrix exponential in system identification: System aliasing. In *Proceedings of the 22nd International Symposium on Mathematical Theory of Networks and Systems*, 2016.

A Global Temporal and Fiber Proofs

A.1 Source-time group law

Proof of Theorem 2.1. Take $t = \sum_j z_j \Delta_j \in \mathcal{G}_0(D)$. Matrix exponentials are invertible, and exponentials of scalar multiples of one matrix commute, so

$$e^{tQ} = \prod_j (e^{\Delta_j Q})^{z_j} = \prod_j (e^{\Delta_j \widetilde{Q}})^{z_j} = e^{t\widetilde{Q}}.$$

Negative exponents mean ordinary inverses. Continuity in t extends equality to $\mathcal{T}(D)$.

Let $H := \mathcal{T}(D)$ and $\rho := \inf(H \cap (0, \infty))$. If $\rho = 0$, then for every $x \in \mathbb{R}$ and $\varepsilon > 0$ choose $g \in H$ with $0 < g < 2\varepsilon$ and an integer n nearest to x/g; then $|x - ng| < \varepsilon$. Thus H is dense, hence $H = \mathbb{R}$ because it is closed. If $\rho > 0$, take $g_n \downarrow \rho$ in H; closedness gives $\rho \in H$. For any $x \in H$, Euclidean division $x = q\rho + r$, $0 \le r < \rho$, gives $r \in H$, so minimality forces $r = 0$. Hence $H = \rho\mathbb{Z}$. Since $\mathcal{G}_0(D)$ is dense in this discrete group, it equals it; write $h = \rho$.

In the dense case choose positive $t_n \in \mathcal{G}_0(D)$ with $t_n \downarrow 0$. Then

$$Q = \lim_n \frac{e^{t_n Q} - I}{t_n} = \lim_n \frac{e^{t_n \widetilde{Q}} - I}{t_n} = \widetilde{Q}.$$

In the discrete case, $\Delta_j = n_j h$ and $h \in \mathcal{G}_0(D)$, so Bézout gives integers b_j with $\sum_j b_j n_j = 1$ and

$$e^{hQ} = \prod_j (e^{\Delta_j Q})^{b_j}.$$

All forward powers follow. Applying the argument actionwise proves the controlled statement. □

Proof of Proposition 2.2. For $q_x = -Q(x,x) \le \Lambda$, Gershgorin places every eigenvalue of Q in a disc centered at $-q_x$ of radius q_x, hence $|\operatorname{Im} \lambda(Q)| \le \Lambda$. If $\Delta\Lambda < \pi$, the spectrum of ΔQ lies in the open principal-logarithm strip. Therefore $e^{\Delta Q}$ has no eigenvalue on the closed negative real axis and $\operatorname{Log}(e^{\Delta Q}) = \Delta Q$; see Higham (2008). Equal exponentials have equal principal logs. □

A.2 Fiberwise oracle theorem

Proof of Theorem 4.1. Fix a realizable skeleton s. Every finite or infinite source experiment has the same law under all $\theta \in F(s)$ because every conditional endpoint law in equation 4 is the same. Hence an oracle rule that observes s chooses one lottery $\mu(s)$ for the complete fiber, and its worst risk is at least $\mathfrak{a}_\star(s)$. Taking the worst fiber gives the lower bound.

The policy simplex is compact and regret is continuous, so each fiberwise robust game attains its value. A countable dense subset of the simplex and measurable first-ε selection yield Borel ε-optimal fiberwise rules; sending $\varepsilon \downarrow 0$ gives the reverse inequality. Because deterministic-policy regrets are nonnegative, a lottery has zero regret throughout a fiber iff every policy in its support is optimal throughout that fiber. This proves the common-policy criterion.

For finite-query closure, sample every finite source experiment in balanced fashion. Finite categorical concentration gives uniform convergence of the empirical skeleton. Optimize worst-case regret over the compact set of models whose source skeleton lies in a shrinking confidence neighborhood of the estimate. Compactness and continuity imply convergence of these thickened fiber games to the exact fiber game, uniformly over true models. The lower bound follows because the exact-skeleton oracle can simulate any finite-query rule. This is an existence proof and makes no general computational claim. □

A.3 ARBITRARY-TARGET DECISION ALIASING

The matrices J and A_{rot} satisfy

$$J^2 = -J, \qquad JA_{\mathrm{rot}} = A_{\mathrm{rot}}J = -A_{\mathrm{rot}}, \qquad A_{\mathrm{rot}}^2 = \pi^2 J.$$

Expanding the exponential on this closed algebra gives

$$e^{th^{-1}(3\alpha J \pm A_{\mathrm{rot}})} = I + [1 - e^{-3\alpha t/h}\cos(\pi t/h)]J \pm e^{-3\alpha t/h}\frac{\sin(\pi t/h)}{\pi}A_{\mathrm{rot}}. \tag{42}$$

The off-diagonal rates are $(\alpha \pm \pi/\sqrt{3})/h$ in the two cyclic orientations, so the risky generators are valid when $\alpha \geq \pi/\sqrt{3}$, irreducible at equality, and interior at strict inequality. Their maximal exit rate is $2\alpha/h$; the safe generator has off-diagonal rate β/h and exit rate $2\beta/h$.

Proof of Theorem 3.1. At every source duration $n_j h$, the sine term in equation 42 vanishes. The safe action is common, so every state–action–duration endpoint law agrees. For an adaptive transcript $z_N = (e_1, y_1, \ldots, e_N, y_N)$,

$$\mathbb{P}^{(N)}_{\theta,\sigma}(z_N) = \prod_{t=1}^{N} \sigma_t(e_t \mid z_{t-1}) p^{\theta}_{e_t}(y_t).$$

The strategy factors are the same predictable functions of the observed history and all endpoint factors agree, proving complete finite-transcript equality. For an almost-surely finite stopping time, apply finite-prefix equality to $B \cap \{\tau \leq n\}$ and let $n \to \infty$.

Let $g = (1/2, 1, 0)^\top$. Since $e_1^\top Jg = 0$, the safe value is $1/2$ at every duration. Also $e_1^\top A_{\mathrm{rot}} g = \pi/\sqrt{3}$, so the raw risky values at rh are

$$\frac{1}{2} \pm \frac{e^{-3\alpha r}\sin(\pi r)}{\sqrt{3}}.$$

Relabel the two models by $\mathrm{sign}(\sin \pi r)$, yielding opposite unique actions and common gap $\gamma_{\alpha,r}$. If p is the common probability that the output policy chooses risky at the witness state, the two expected regrets are at least $(1-p)\gamma_{\alpha,r}$ and $p\gamma_{\alpha,r}$. Their maximum is at least $\gamma_{\alpha,r}/2$. At a stopping time, the common output law cannot put probability exceeding one half on both disjoint success sets, proving the fixed-confidence impossibility. □

Proposition A.1 (Two-state injectivity and minimality)**.** *For every $t > 0$, $Q \mapsto e^{tQ}$ is injective on two-state generators.*

Proof. Every two-state generator is $Q(a,b) = \left[\begin{smallmatrix} -a & a \\ b & -b \end{smallmatrix}\right]$ and satisfies $Q^2 = -(a+b)Q$. If $s = a + b > 0$, $e^{tQ} = I + (1 - e^{-st})Q/s$. The trace identifies s, after which the matrix identifies Q; the identity kernel implies $Q = 0$. Thus three states are minimal under complete endpoint observation at one fixed positive duration. □

B BELLMAN QUOTIENT PROOFS

Fix a continuation space $\mathcal{V}$ and reference action a_0. For a model θ, the pairwise one-step continuation gap between actions a, b is

$$G^{\theta}_{ab} v := (P^{\theta}_a - P^{\theta}_b)v.$$

Proof of Theorem 4.2. Equality of $\mathfrak{B}_{\mathcal{V}}$ is equivalent to equality of every reference-action difference, hence of every pairwise action-gap operator by subtraction. If all pairwise gaps agree, adding arbitrary action rewards produces identical pairwise Bellman comparisons and therefore identical greedy sets. Conversely, suppose some pair a, b, state x, and continuation $v \in \mathcal{V}$ give two different scalar gaps $d \neq d'$. An action-reward offset strictly between $-d$ and $-d'$ makes the two models choose opposite actions, contradicting equality of every greedy correspondence.

The pairwise equalities are also equivalent to the existence of an action-independent map K on $\mathcal{V}$: set $K = (P^{\theta}_{a_0} - P^{\vartheta}_{a_0})|_{\mathcal{V}}$ and use equality of all reference differences to obtain $(P^{\theta}_a - P^{\vartheta}_a)v = Kv$ for every action. Such a common shift cancels from every comparison.

When $\mathbf{1} \in \mathcal{V}$, a failed equality has a normalized nonnegative witness. Replace v by

$$g = \frac{v - (\min_y v(y))\mathbf{1}}{4\,\mathrm{sp}(v)},$$

so $0 \le g \le \mathbf{1}/4$. Let d, d' be the resulting pairwise continuation gaps and $m = (d + d')/2$. At the witness state choose rewards $r(a) = 1/2 - m/2$, $r(b) = 1/2 + m/2$, and give every other action reward zero. The two witness actions are uniquely optimal in opposite models and the total return is at most $7/8$. □

B.1 Sufficient finite-horizon certificate

Let $\mathcal{V}_h$ be linear continuation spaces and $\mathcal{N}_h \subseteq \mathcal{V}_h$ nuisance spaces. The following conditions are sufficient on a source fiber:

1. stage rewards and terminal reward lie in the declared value spaces;
2. all models have the same action-gap operators on $\mathcal{V}_{h+1}$;
3. every nuisance direction in $\mathcal{N}_{h+1}$ is invisible to all action comparisons;
4. model-dependent common shifts propagate into $\mathcal{N}_h$.

Backward induction shows that optimal values remain in $\mathcal{V}_h$, their model differences remain in $\mathcal{N}_h$, and every state–stage optimal action correspondence agrees. We do not claim necessity for arbitrary coherent fibers.

B.2 Strict target-kernel/policy separation

Let $\bar{Q}_a$ be nontrivial macro generators and let R^+, R^- be the exact three-state micro aliases from Appendix A. Define $Q_a^{\pm} = \bar{Q}_a \oplus R^{\pm}$. Kronecker-sum factorization gives

$$e^{tQ_a^{\pm}} = e^{t\bar{Q}_a} \otimes e^{tR^{\pm}}.$$

The source kernels agree, while target micro kernels may differ. For every macro-measurable reward and continuation Uv with $U = I \otimes \mathbf{1}_3$, however, $Q_a^{\pm}U = U\bar{Q}_a$ and therefore $e^{tQ_a^{\pm}}U = Ue^{t\bar{Q}_a}$. The complete Bellman recursion is the same in both models. Taking a two-state macro death process gives a unique, state-dependent nontrivial action, proving that target-kernel identification is strictly stronger than policy coherence for the declared task class.

C Arithmetic Quotient and Legal Temporal-Design Proofs

For $Q_{\alpha,k}^{(h)} = h^{-1}(3\alpha J + (2k+1)A_{\mathrm{rot}})$, the algebra in Appendix A yields

$$P_{\alpha,k}(u) = I + [1 - e^{-3\alpha u}\cos((2k+1)\pi u)]J + e^{-3\alpha u}\frac{\sin((2k+1)\pi u)}{\pi}A_{\mathrm{rot}}. \tag{43}$$

Using $\langle J, A_{\mathrm{rot}}\rangle_{\mathrm{F}} = 0$, $\|J\|_{\mathrm{F}}^2 = 2$, $\|A_{\mathrm{rot}}\|_{\mathrm{F}}^2 = 2\pi^2$, and the chord identities gives

$$\|P_{\alpha,k}(u) - P_{\alpha,\ell}(u)\|_{\mathrm{F}}^2 = 8e^{-6\alpha u}\sin^2(\pi(k-\ell)u). \tag{44}$$

This vanishes iff $(k-\ell)u \in \mathbb{Z}$, proving the source and target annihilator statements in Theorem 5.1. If $r = p/q$ is reduced, $dr \in \mathbb{Z}$ iff $q \mid d$; irrational r admits no nonzero integer annihilator.

C.1 The fully specified target-control task

The safe action has generator $Q_{\mathrm{safe}} = 0$ and the risky action has generator $Q_{\mathrm{risk}}^{(k)} = Q_{\alpha,k}^{(h)}$. The stage reward is zero and the terminal reward is

$$g = (1/2, 1, 0)^{\top}.$$

Let $a = e^{-3\alpha r}$ and $\vartheta_k = (2k+1)\pi r$. Since

$$Jg = (0, -1/2, 1/2)^{\top}, \qquad \pi^{-1}A_{\mathrm{rot}}g = (1/\sqrt{3}, -1/(2\sqrt{3}), -1/(2\sqrt{3}))^{\top},$$

substitution in equation 43 gives the risky-minus-safe Bellman gaps

$$G_1(k,r) = \frac{a}{\sqrt{3}} \sin \vartheta_k, \tag{45}$$

$$G_2(k,r) = \frac{1}{2}\left[a\left(\cos\vartheta_k - \frac{\sin\vartheta_k}{\sqrt{3}}\right) - 1\right], \tag{46}$$

$$G_3(k,r) = \frac{1}{2}\left[1 - a\left(\cos\vartheta_k + \frac{\sin\vartheta_k}{\sqrt{3}}\right)\right]. \tag{47}$$

We now prove the strict state-2 and state-3 comparisons rather than inferring them from irreducibility. Branch admissibility implies

$$3\alpha r \geq \sqrt{3}|\vartheta_k|.$$

For every real z,

$$\left|\cos z \pm \frac{\sin z}{\sqrt{3}}\right| \leq |\cos z| + \frac{|\sin z|}{\sqrt{3}} \leq 1 + \frac{|z|}{\sqrt{3}} \leq e^{|z|/\sqrt{3}}.$$

Because $r > 0$ and $2k+1 \neq 0$, $\vartheta_k \neq 0$. Therefore

$$\begin{aligned} a\left|\cos\vartheta_k \pm \frac{\sin\vartheta_k}{\sqrt{3}}\right| &\leq e^{-3\alpha r} e^{|\vartheta_k|/\sqrt{3}} \\ &\leq e^{-\sqrt{3}|\vartheta_k| + |\vartheta_k|/\sqrt{3}} = e^{-2|\vartheta_k|/\sqrt{3}} < 1. \end{aligned}$$

Equations equation 46–equation 47 now give $G_2(k,r) < 0$ and $G_3(k,r) > 0$. Thus safe is uniquely optimal at state 2 and risky is uniquely optimal at state 3 for every branch. At state 1, risky is uniquely optimal when $G_1 > 0$, safe is uniquely optimal when $G_1 < 0$, and both are optimal when $G_1 = 0$.

Consequently two branches have equal complete optimal-policy sets exactly when their three-valued signs $\operatorname{sign}(G_1(k,r)) \in \{-1,0,1\}$ agree. A source fiber has at least one common optimal policy exactly when it does not contain both a positive-gap and a negative-gap branch. A zero-gap branch imposes no restriction at state 1 and is compatible with either strict side.

If source-equivalent branches k, ℓ have opposite signs, let $\gamma_j = |G_1(j,r)|$. Their complete adaptive source transcript laws are equal. If the common output law chooses risky at state 1 with probability p, the two regrets are at least $(1-p)\gamma_k$ and $p\gamma_\ell$. Hence every learner incurs worst-case regret at least

$$\min_{p\in[0,1]} \max\{(1-p)\gamma_k, p\gamma_\ell\} = \frac{\gamma_k\gamma_\ell}{\gamma_k + \gamma_\ell}.$$

C.2 Rowwise separation and KL comparison

Fix any retained start row x. By cyclic symmetry every row has the same Euclidean separation, and direct calculation from equation 43 gives

$$\|p_{k,u} - p_{\ell,u}\|_2^2 = \frac{8}{3} e^{-6\alpha u} \sin^2(\pi(k-\ell)u). \tag{48}$$

For probability vectors p, q, $\mathrm{TV}(p,q) \geq \|p-q\|_2/2$, so Pinsker yields

$$D_{\mathrm{KL}}(p\|q) \geq 2\,\mathrm{TV}(p,q)^2 \geq \frac{1}{2}\|p-q\|_2^2.$$

If every coordinate of q is at least ν,

$$D_{\mathrm{KL}}(p\|q) \leq \chi^2(p\|q) = \sum_y \frac{(p_y - q_y)^2}{q_y} \leq \nu^{-1}\|p-q\|_2^2.$$

For the present finite branch family and compact legal domain with $u_{\min} > 0$, every risky generator is irreducible, every endpoint kernel is entrywise positive, and continuity gives a uniform support

constant $\nu > 0$. For a finite duration vector $\mathcal{U} = (u_j)$ and allocation w, define

$$\chi_{\mathrm{ans}}(k;\mathcal{U},w) := \min_{\ell:c(\ell)\neq c(k)} \sum_j w_j e^{-6\alpha u_j} \sin^2(\pi(k-\ell)u_j),$$

$$C_{\mathrm{ans}}(k;\mathcal{U},w) := \min_{\ell:c(\ell)\neq c(k)} \sum_j w_j D_{\mathrm{KL}}(p_{k,u_j} \| p_{\ell,u_j}).$$

Substituting equation 48, integrating against the design, and minimizing over harmful alternatives gives

$$\frac{4}{3}\chi_{\mathrm{ans}}(k;\mathcal{U},w) \le C_{\mathrm{ans}}(k;\mathcal{U},w) \le \frac{8}{3\nu}\chi_{\mathrm{ans}}(k;\mathcal{U},w). \tag{49}$$

All summands defining χ_{ans} are nonnegative. Hence it is positive exactly when each harmful ℓ has some j with $w_j > 0$ and $(k-\ell)u_j \notin \mathbb{Z}$. The fixed-duration object $C_{\mathrm{ans}}(k;\mathcal{U},w)$, its allocation optimum $C^{\star}_{\mathrm{ans}}(k;\mathcal{U})$, and its duration optimum $C^{\star,\mathrm{dur}}_{\mathrm{ans}}(k)$ are distinct. The first is continuous in w; the latter two attain their maxima by compactness and continuity. If the harmful set is empty, the answer is known and the zero-sample rule is correct.

C.3 Denominator matching

If $u = a/L$ is reduced, then $du \in \mathbb{Z}$ iff $L \mid d$, hence $\mathrm{Ann}(u) = L\mathbb{Z}$. For target $r = p/q$, $q \mid L$ implies $L\mathbb{Z} \subseteq q\mathbb{Z}$, so every source alias is target-kernel harmless. If $L \le 2K$ on $\mathcal{K} = \{-K,\dots,K\}$, a distinct pair separated by L remains source aliased, proving strict branch/target separation. This claim applies only when the chosen rational duration lies in the legal domain $\mathcal{U}_{\mathrm{src}}$.

C.4 Exact target-excluding design complexity

We prove Theorem 6.1. Begin with the finite latent family and define source equivalence by

$$\theta \sim_{\mathrm{src}} \vartheta \quad \Longleftrightarrow \quad p^{\theta}_e = p^{\vartheta}_e \text{ for every } e \in \mathcal{E}_{\mathrm{src}}.$$

The theorem assumes answer compatibility: $\theta \sim_{\mathrm{src}} \vartheta$ implies $c(\theta) = c(\vartheta)$. Hence the endpoint laws and answer map descend to the finite observable quotient $\mathcal{Q} := \Theta/\sim_{\mathrm{src}}$. Distinct elements of $\mathcal{Q}$ are separated by at least one legal experiment by construction. The proof is stated on this quotient; exact same-answer latent aliases therefore remain unresolved throughout. The legal experiment space itself may be continuous. Let $\mathcal{Y}$ be the finite endpoint outcome set. For each legal experiment $e \in \mathcal{E}_{\mathrm{src}}$, let $\mathcal{S}_e \subseteq \mathcal{Y}$ be a support that is common to all $q \in \mathcal{Q}$, and assume

$$p^q_e(y) = 0 \quad (y \notin \mathcal{S}_e), \qquad p^q_e(y) \ge \nu > 0 \quad (y \in \mathcal{S}_e) \tag{50}$$

uniformly in (q,e). Thus every legal log-likelihood ratio is bounded by

$$\left|\log \frac{p^q_e(y)}{p^{q'}_e(y)}\right| \le B := \log(1/\nu). \tag{51}$$

The query-selection kernel of an adaptive learner is required to be the same predictable function of the observed history under every model. This is the usual source-only experiment: the unknown model changes endpoint laws, not the learner's randomization rule.

For a probability measure ξ on $\mathcal{E}_{\mathrm{src}}$, write

$$D_{\xi}(q,q') := \int D_{\mathrm{KL}}(p^q_e \| p^{q'}_e)\, \xi(de).$$

Continuity of $e \mapsto p^q_e$ and the common-support bound make $e \mapsto D_{\mathrm{KL}}(p^q_e \| p^{q'}_e)$ continuous and bounded.

C.4.1 Lower bound for adaptive continuous experiment selection

Lemma C.1 (Stopped adaptive KL identity)**.** *Let τ be a stopping time with $\mathbb{E}_q[\tau] < \infty$. For every $q' \in \mathcal{Q}$,*

$$D_{\mathrm{KL}}(\mathbb{P}^{\mathcal{F}_\tau}_q \| \mathbb{P}^{\mathcal{F}_\tau}_{q'}) = \mathbb{E}_q\left[\sum_{t=1}^{\tau} D_{\mathrm{KL}}(p^q_{E_t} \| p^{q'}_{E_t})\right]. \tag{52}$$

Proof. For $\tau_n := \tau \wedge n$, the query-selection factors cancel from the likelihood ratio because they are the same predictable kernels under q and q'. Hence

$$\log \frac{d\mathbb{P}_q^{\mathcal{F}_{\tau_n}}}{d\mathbb{P}_{q'}^{\mathcal{F}_{\tau_n}}} = \sum_{t=1}^{\tau_n} \log \frac{p_{E_t}^q(Y_t)}{p_{E_t}^{q'}(Y_t)}.$$

Taking $\mathbb{E}_q$ and conditioning on $\mathcal{F}_{t-1}$ gives the finite n version of equation 52. By equation 51, the absolute value of the stopped log likelihood ratio is at most $B\tau$, which is integrable. Dominated convergence therefore lets $n \to \infty$ on both the log-likelihood and its conditional-KL expectation, proving the identity. □

Proof of Theorem 6.1, part 1. If $\mathbb{E}_q[\tau] = +\infty$, the claim is immediate. Otherwise define the finite expected sampling measure

$$\Xi_q(B) := \mathbb{E}_q\left[\sum_{t=1}^{\tau} \mathbf{1}\{E_t \in B\}\right], \qquad B \subseteq \mathcal{E}_{\mathrm{src}} \text{ Borel.} \tag{53}$$

Its total mass is $\Xi_q(\mathcal{E}_{\mathrm{src}}) = \mathbb{E}_q[\tau]$. Let B_q be the event that the returned answer is $c(q)$. Uniform η-correctness gives

$$\mathbb{P}_q(B_q) \geq 1 - \eta, \qquad \mathbb{P}_{q'}(B_q) \leq \eta \quad (q' \in \mathcal{H}(q)).$$

Binary data processing and Lemma C.1 yield

$$\begin{aligned} \mathrm{kl}(1-\eta, \eta) &\leq D_{\mathrm{KL}}(\mathbb{P}_q^{\mathcal{F}_\tau} \| \mathbb{P}_{q'}^{\mathcal{F}_\tau}) \\ &= \int_{\mathcal{E}_{\mathrm{src}}} D_{\mathrm{KL}}(p_e^q \| p_e^{q'})\, \Xi_q(de) \qquad (q' \in \mathcal{H}(q)). \end{aligned} \tag{54}$$

Normalize Ξ_q to $\bar{\xi}_q := \Xi_q / \mathbb{E}_q[\tau] \in \mathcal{P}(\mathcal{E}_{\mathrm{src}})$. Taking the minimum in equation 54 over harmful alternatives gives

$$\mathrm{kl}(1-\eta, \eta) \leq \mathbb{E}_q[\tau] \min_{q' \in \mathcal{H}(q)} D_{\bar{\xi}_q}(q, q') \leq \mathbb{E}_q[\tau] C_{\mathrm{ans}}^\star(q),$$

which is equation 20. The proof permits arbitrary adaptive, randomized choices in the continuous legal experiment space. □

C.4.2 Existence, duality, and sparse support

Lemma C.2 (Optimal legal designs exist and are sparse)**.** *For every q with $m_q := |\mathcal{H}(q)| > 0$, the maximum defining $C_{\mathrm{ans}}^\star(q)$ is attained. It satisfies the dual identity equation 21, and one optimizer has support size at most $m_q + 1$.*

Proof. The set $\mathcal{P}(\mathcal{E}_{\mathrm{src}})$ is compact in the weak topology because $\mathcal{E}_{\mathrm{src}}$ is compact. Each map $\xi \mapsto D_\xi(q, q')$ is continuous and affine. Their finite minimum is continuous, so the maximum is attained.

For $\lambda \in \Delta(\mathcal{H}(q))$, define the bilinear continuous payoff

$$F_q(\xi, \lambda) := \sum_{q' \in \mathcal{H}(q)} \lambda_{q'} D_\xi(q, q').$$

The minimum of the coordinates of a vector equals its minimum over the simplex, hence

$$C_{\mathrm{ans}}^\star(q) = \max_{\xi \in \mathcal{P}(\mathcal{E}_{\mathrm{src}})} \min_{\lambda \in \Delta(\mathcal{H}(q))} F_q(\xi, \lambda).$$

Sion's minimax theorem applies to the two compact convex sets and gives

$$C_{\mathrm{ans}}^\star(q) = \min_{\lambda \in \Delta(\mathcal{H}(q))} \max_{\xi \in \mathcal{P}(\mathcal{E}_{\mathrm{src}})} F_q(\xi, \lambda).$$

For fixed λ, the maximum of a linear functional over probability measures is attained by a point mass, so

$$\max_{\xi} F_q(\xi, \lambda) = \max_{e \in \mathcal{E}_{\mathrm{src}}} \sum_{q' \in \mathcal{H}(q)} \lambda_{q'} D_{\mathrm{KL}}(p_e^q \| p_e^{q'}),$$

proving equation 21.

For sparse support, define the continuous information-vector map

$$d_q(e) := \left(D_{\mathrm{KL}}(p_e^q \| p_e^{q'})\right)_{q' \in \mathcal{H}(q)} \in \mathbb{R}^{m_q}.$$

If $\xi_q^\star$ is any optimizer, its information vector $z_q^\star := \int d_q(e)\, \xi_q^\star(de)$ lies in $\operatorname{conv} d_q(\mathcal{E}_{\mathrm{src}})$. Carathéodory's theorem represents $z_q^\star$ as a convex combination of at most $m_q + 1$ points of $d_q(\mathcal{E}_{\mathrm{src}})$. The corresponding finitely supported design has exactly the same harmful-alternative information vector and is therefore also optimal. □

Lemma C.3 (Canonical sparse optimizer)**.** *Fix the lexicographic order on experiments $e = (x, a, u)$ and put $m_q := |\mathcal{H}(q)| + 1$. Among all optimal representations $\sum_{i=1}^{m_q} w_i \delta_{e_i}$ with $e_1 \preceq \cdots \preceq e_{m_q}$, zero weights and repeated support points allowed, there is a unique lexicographically smallest tuple $(e_1, \ldots, e_{m_q}, w_1, \ldots, w_{m_q})$. Its measure is the canonical optimizer $\xi_q^\star$ used by the algorithm.*

Proof. The ordered representation space is a closed subset of $\mathcal{E}_{\mathrm{src}}^{m_q} \times \Delta_{m_q}$ and is therefore compact. The map from a representation to every harmful-alternative information coordinate is continuous, so the subset representing optimal designs is nonempty and compact by Lemma C.2. Minimize the first scalar coordinate over this set, then the second over the remaining compact minimizers, and continue through the finite coordinate list. The final nested compact set contains exactly one ordered representation, which is the declared lexicographic minimum. Since $\mathcal{Q}$ is finite, this also gives a measurable, reproducible selection $q \mapsto \xi_q^\star$. □

Pairwise legal distinguishability implies positivity. Indeed, choose for every unordered pair $q \neq q'$ one legal experiment $e_{q,q'}$ with $p^q_{e_{q,q'}} \neq p^{q'}_{e_{q,q'}}$, and let ξ_0 be the uniform measure on this finite collection. Then

$$\kappa_0 := \min_{q \neq q'} D_{\xi_0}(q, q') > 0. \tag{55}$$

In particular $C^\star_{\mathrm{ans}}(q) > 0$ whenever $\mathcal{H}(q) \neq \varnothing$.

C.4.3 A LEGAL CONTINUOUS-EXPERIMENT TRACK-AND-STOP RULE

Use the canonical sparse optimizer $\xi_q^\star$ from Lemma C.3 and the baseline ξ_0 from equation 55. Let $\gamma_t := (t+1)^{-1/3}$. Given observations through time t, define

$$\ell_t(q) := \sum_{s=1}^{t} \log p^q_{E_s}(Y_s), \qquad \widehat{q}_t := \min \arg\max_{q \in \mathcal{Q}} \ell_t(q),$$

where the minimum refers to a fixed deterministic ordering of $\mathcal{Q}$. Conditional on $\mathcal{F}_t$, draw E_{t+1} from

$$\lambda_{t+1} := (1 - \gamma_t)\xi^\star_{\widehat{q}_t} + \gamma_t \xi_0.$$

All support points of both measures belong to $\mathcal{E}_{\mathrm{src}}$, so target exclusion holds pathwise. Write $Z_t(q, q') := \ell_t(q) - \ell_t(q')$ and $b_\eta := \log((|\mathcal{Q}| - 1)/\eta)$. The stopping rule is equation 19; it returns $c(\widehat{q}_{\tau_\eta})$ and the declared policy $\pi_{c(\widehat{q}_{\tau_\eta})}$.

Lemma C.4 (Time-uniform correctness)**.** *The rule is uniformly η-correct.*

Proof. Fix the true model q. For any $q' \neq q$,

$$R_t(q', q) := \exp Z_t(q', q)$$

is a nonnegative martingale under q. Indeed, conditional on $\mathcal{F}_{t-1}$ and the predictable experiment E_t,

$$\mathbb{E}_q\left[\frac{p^{q'}_{E_t}(Y_t)}{p^q_{E_t}(Y_t)}\middle| \mathcal{F}_{t-1}, E_t\right] = 1.$$

If the procedure stops with a wrong answer, then for some model q' with $c(q') \neq c(q)$ we have $\widehat{q}_t = q'$ and, because $q \in \mathcal{H}(q')$, $Z_t(q', q) \geq b_\eta$. Ville's inequality and a union bound give

$$\mathbb{P}_q(\text{wrong answer}) \leq \sum_{q' \neq q} \mathbb{P}_q\left(\sup_t R_t(q', q) \geq e^{b_\eta}\right) \leq (|\mathcal{Q}| - 1)e^{-b_\eta} = \eta.$$

□

Lemma C.5 (MLE stabilization with a summable tail)**.** *For every true q, $\widehat{q}_t = q$ eventually almost surely. Moreover, for the stabilization time*

$$T_q := 1 + \sup\{t \geq 0 : \widehat{q}_t \neq q\},$$

there exist constants $A_q, a_q > 0$ such that

$$\mathbb{P}_q(T_q > n) \leq A_q e^{-a_q n^{1/3}} \qquad (n \geq 1). \tag{56}$$

Proof. Fix $q' \neq q$. The conditional mean increment of $Z_t(q, q')$ at round s is

$$D_{\lambda_s}(q, q') \geq \gamma_{s-1} D_{\xi_0}(q, q') \geq \gamma_{s-1}\kappa_0.$$

Let

$$A_t(q, q') := \sum_{s=1}^{t} D_{\lambda_s}(q, q'), \qquad M_t(q, q') := Z_t(q, q') - A_t(q, q').$$

Then M_t is a martingale and its increments have absolute value at most $2B$ by equation 51. Also

$$A_t(q, q') \geq \kappa_0 \sum_{s=1}^{t} s^{-1/3} \geq c_0 t^{2/3}$$

for a constant $c_0 > 0$. Azuma–Hoeffding therefore gives

$$\mathbb{P}_q\big(Z_t(q, q') \leq 0\big) \leq \exp\left(-\frac{c_0^2 t^{1/3}}{8B^2}\right). \tag{57}$$

If $\widehat{q}_t \neq q$, then $Z_t(q, q') \leq 0$ for some $q' \neq q$. The union bound, followed by summation of equation 57 over $t \geq n$, proves equation 56; the polynomial factor from the integral test is absorbed into $A_q e^{-a_q n^{1/3}}$ after adjusting the constants. The tail is summable, so Borel–Cantelli gives almost-sure stabilization. □

Lemma C.6 (Allocation and evidence convergence)**.** *Under true q, the empirical experiment measure converges weakly to $\xi_q^\star$, and for every $q' \neq q$,*

$$\frac{Z_t(q, q')}{t} \longrightarrow D_{\xi_q^\star}(q, q') \qquad \textit{a.s.}$$

Consequently equation 23 and equation 24 hold.

Proof. By Lemma C.5, after an almost-surely finite time,

$$\lambda_{t+1} = (1 - \gamma_t)\xi_q^\star + \gamma_t \xi_0.$$

For any bounded continuous f on the compact experiment space,

$$\frac{1}{t}\sum_{s=1}^{t}\left(f(E_s) - \int f\, d\lambda_s\right) \longrightarrow 0 \quad \text{a.s.}$$

by the martingale strong law. Since $\gamma_t \to 0$, Cesàro averaging gives $t^{-1}\sum_s \int f\, d\lambda_s \to \int f\, d\xi_q^\star$. Applying this to a countable convergence-determining class of continuous functions proves weak convergence of the empirical experiment measure.

For each q', the centered log-likelihood-ratio increments are bounded martingale differences. Hence

$$\frac{Z_t(q, q') - \sum_{s=1}^{t} D_{\lambda_s}(q, q')}{t} \longrightarrow 0 \quad \text{a.s.}$$

The conditional-mean average converges to $D_{\xi_q^\star}(q, q')$. Taking the minimum over the finite harmful set gives the evidence limit $C_{\text{ans}}^\star(q)$. □

Lemma C.7 (Pathwise characteristic-time limit)**.** *With all confidence levels coupled through one infinite source trajectory,*

$$\frac{\tau_\eta}{b_\eta} \longrightarrow \frac{1}{C^\star_{\mathrm{ans}}(q)} \qquad \textit{a.s.}$$

Proof. By stabilization, the MLE equals q eventually. By Lemma C.6, the correct stopping statistic divided by t converges to $C^\star_{\mathrm{ans}}(q) > 0$. For every wrong-answer model q', $Z_t(q', q) = -Z_t(q, q') \to -\infty$, so the supremum of its likelihood ratio over all t is finite on each sample path. Hence sufficiently large thresholds cannot be crossed by a wrong answer, and the first crossing of the correct statistic occurs at $b_\eta / C^\star_{\mathrm{ans}}(q)(1 + o(1))$. □

For a generic threshold $b > 0$, let

$$\tau(b) := \inf\left\{ t \geq 1 : \min_{q' \in \mathcal{H}(\widehat{q}_t)} Z_t(\widehat{q}_t, q') \geq b \right\}. \tag{58}$$

Thus the algorithm in the main text has $\tau_\eta = \tau(b_\eta)$.

Lemma C.8 (Expected stopping-time upper bound)**.** *For every true q,*

$$\limsup_{\eta \downarrow 0} \frac{\mathbb{E}_q[\tau_\eta]}{b_\eta} \leq \frac{1}{C^\star_{\mathrm{ans}}(q)}.$$

Proof. Write $C := C^\star_{\mathrm{ans}}(q)$ and fix $\varepsilon > 0$. Let

$$n_b := \left\lceil \frac{(1+\varepsilon)b}{C} \right\rceil, \qquad m_n := \lfloor n^{3/4} \rfloor.$$

On the event $\{T_q \leq m_n\}$, every allocation after m_n is based on the true model. For any harmful q' the cumulative conditional mean satisfies

$$\begin{aligned} A_n(q, q') &:= \sum_{s=1}^{n} D_{\lambda_s}(q, q') \\ &\geq \sum_{s=m_n+1}^{n} (1 - \gamma_{s-1}) D_{\xi^\star_q}(q, q') \\ &\geq C \left(n - m_n - \sum_{s=m_n+1}^{n} \gamma_{s-1} \right). \end{aligned} \tag{59}$$

The fraction subtracted from n in equation 59 is $o(n)$, uniformly for $n \geq n_b$ as $b \to \infty$. Thus for all sufficiently large b and every $n \geq n_b$,

$$A_n(q, q') - b \geq \delta_\varepsilon n, \qquad \delta_\varepsilon := \frac{C\varepsilon}{2(1+\varepsilon)}.$$

If $T_q \leq m_n$ and $\tau(b) > n$, then at time n the MLE is q and some harmful q' satisfies $Z_n(q, q') < b$. Since $Z_n(q, q') - A_n(q, q')$ is a martingale with increments bounded by $2B$, Azuma–Hoeffding and Lemma C.5 imply

$$\mathbb{P}_q(\tau(b) > n) \leq A_q e^{-a_q n^{1/4}} + |\mathcal{H}(q)| \exp\left(-\frac{\delta_\varepsilon^2 n}{8B^2} \right) \qquad (n \geq n_b). \tag{60}$$

The first exponent uses $m_n^{1/3} \asymp n^{1/4}$. Both tails are summable and their sum over $n \geq n_b$ is $o(b)$. Therefore

$$\mathbb{E}_q[\tau(b)] = \sum_{n \geq 0} \mathbb{P}_q(\tau(b) > n) \leq n_b + o(b).$$

Divide by b, let $b \to \infty$, and then let $\varepsilon \downarrow 0$. □

Completion of Theorem 6.1. Existence, duality, and sparse support are Lemma C.2. Correctness, stabilization, allocation, and evidence convergence are Lemmas C.4– C.6. The adaptive lower bound and $\mathrm{kl}(1-\eta,\eta)/\log(1/\eta) \to 1$ give

$$\liminf_{\eta\downarrow 0} \frac{\mathbb{E}_q[\tau_\eta]}{\log(1/\eta)} \geq \frac{1}{C^\star_{\mathrm{ans}}(q)}.$$

Lemma C.8, together with $b_\eta/\log(1/\eta) \to 1$, gives the reverse limsup. This proves equation 25; Lemma C.7 proves the coupled almost-sure limit. Every design in the algorithm is supported on $\mathcal{E}_{\mathrm{src}}$, so the target and its excluded neighborhood are never queried. Because the procedure operates on $\mathcal{Q}$, it never attempts to distinguish latent models that have identical legal source laws and the same target answer. □

C.5 Finite-sample temporal conditioning

Fix a strict branch answer $c(k)$ and define

$$\mathcal{D}_k := \{k-\ell : c(\ell) \neq c(k)\}, \qquad D_k^{\max} := \max_{d\in\mathcal{D}_k} |d|, \qquad c_{\mathrm{BA}}(u) := \inf_{n\geq 1} n\|nu\|_{\mathbb{Z}}.$$

Throughout this subsection $\mathcal{D}_k$ is nonempty. For a fixed source ratio $u > 0$, let $p_{x,k}(u)$ denote row x of the risky endpoint kernel, let $\nu_u := \min_{j\in\mathcal{K},\, x,y} P_{\alpha,j}(u)(x,y) > 0$, and put

$$C_{\mathrm{ans}}(k,u) := \min_{\ell: c(\ell)\neq c(k)} D_{\mathrm{KL}}(p_{1,k}(u)\|p_{1,\ell}(u))\,. \tag{61}$$

Because the kernels are circulant, every corresponding row pair is a cyclic permutation of the first-row pair and has the same directional KL. Thus equation 61 is also the maximum information an adaptive learner can obtain from one query when the duration is fixed and the safe action is common to all branches.

Lemma C.9 (Row separation and Diophantine distance)**.** *For every pair k,ℓ and every $u > 0$,*

$$\|p_{1,k}(u) - p_{1,\ell}(u)\|_2^2 = \frac{8}{3} e^{-6\alpha u} \sin^2(\pi(k-\ell)u). \tag{62}$$

Moreover, for every $x \in \mathbb{R}$,

$$2\|x\|_{\mathbb{Z}} \leq |\sin(\pi x)| \leq \pi\|x\|_{\mathbb{Z}}. \tag{63}$$

Proof. Equation equation 62 is the rowwise version of the exact Frobenius identity: the three row differences are cyclic permutations and the sum of their squared norms is $8e^{-6\alpha u}\sin^2(\pi(k-\ell)u)$. Hence each row contributes one third. For equation 63, let $z = \|x\|_{\mathbb{Z}} \in [0, 1/2]$. Periodicity and symmetry give $|\sin(\pi x)| = \sin(\pi z)$. Concavity of sine on $[0,\pi/2]$ places it above the chord joining $(0,0)$ and $(\pi/2,1)$, yielding $\sin(\pi z) \geq 2z$; the elementary bound $\sin y \leq y$ gives the upper bound. □

Proof of Theorem 6.2. Let

$$\kappa := \kappa_{\mathrm{ans}}(k,u) = \min_{d\in\mathcal{D}_k} \|du\|_{\mathbb{Z}}.$$

Pinsker's inequality and $\|v\|_1 \geq \|v\|_2$ give, for every harmful ℓ,

$$D_{\mathrm{KL}}(p_{1,k}(u)\|p_{1,\ell}(u)) \geq \frac{1}{2}\|p_{1,k}(u) - p_{1,\ell}(u)\|_2^2.$$

Using Lemma C.9,

$$D_{\mathrm{KL}}(p_{1,k}(u)\|p_{1,\ell}(u)) \geq \frac{4}{3} e^{-6\alpha u}\sin^2(\pi(k-\ell)u) \geq \frac{16}{3} e^{-6\alpha u}\kappa^2.$$

Taking the harmful minimum proves the lower half of equation 27.

The chi-square comparison gives

$$D_{\mathrm{KL}}(p\|q) \leq \sum_y \frac{(p_y - q_y)^2}{q_y} \leq \nu_u^{-1}\|p-q\|_2^2.$$

Choose a harmful displacement $d_\star$ attaining κ. Then

$$C_{\mathrm{ans}}(k,u) \le \frac{8}{3\nu_u} e^{-6\alpha u} \sin^2(\pi d_\star u) \le \frac{8\pi^2}{3\nu_u} e^{-6\alpha u} \kappa^2,$$

which proves the upper half.

Now suppose $u = a/L + \varepsilon$ and $d = mL \in \mathcal{D}_k$. Since $da/L = ma \in \mathbb{Z}$,

$$\sin(\pi d u) = (-1)^{ma} \sin(\pi d \varepsilon).$$

The preceding chi-square bound and $|\sin z| \le |z|$ imply

$$C_{\mathrm{ans}}(k,u) \le \frac{8\pi^2}{3\nu_u} e^{-6\alpha u} d^2 \varepsilon^2. \tag{64}$$

For any uniformly η-correct rule whose source duration is fixed at u, the stopped change-of-measure inequality gives

$$\mathbb{E}_k[\tau]\, C_{\mathrm{ans}}(k,u) \ge \mathrm{kl}(1-\eta, \eta).$$

Combining this with equation 64 proves equation 28. This bound permits adaptive starting states and actions: the safe action has zero branch information and all risky rows have the same KL by circulant symmetry.

We next prove the legal-jitter statement. Write the legal set as a disjoint union of m closed intervals and let $|\mathcal{I}|$ denote its total length. For a nonzero integer d and $\epsilon \in [0, 1/2]$, define

$$B_{d,\epsilon} := \{u \in \mathcal{I} : \|du\|_{\mathbb{Z}} \le \epsilon\}.$$

On an interval I of length L, the change of variables $v = |d|u$ gives an interval of length $|d|L$. Every complete unit period contains bad-set length 2ϵ, and at most two incomplete boundary periods remain, each contributing at most 2ϵ. Therefore

$$\mathrm{Leb}(B_{d,\epsilon} \cap I) \le 2\epsilon L + \frac{4\epsilon}{|d|}.$$

Summing over the m components and dividing by $|\mathcal{I}|$ yields

$$\mathbb{P}\big(\|dU\|_{\mathbb{Z}} \le \epsilon\big) \le \epsilon \left(2 + \frac{4m}{|d|\,|\mathcal{I}|}\right). \tag{65}$$

A union bound over $d \in \mathcal{D}_k$ gives

$$\mathbb{P}\left(\kappa_{\mathrm{ans}}(k,U) \le \epsilon\right) \le \epsilon A_{\mathcal{I}}(\mathcal{D}_k).$$

Since $A_{\mathcal{I}}(\mathcal{D}_k) \ge 2$ and $\delta \in (0,1)$, the choice $\epsilon = \delta / A_{\mathcal{I}}(\mathcal{D}_k)$ lies in $(0, 1/2)$. With probability at least $1 - \delta$,

$$\kappa_{\mathrm{ans}}(k,U) \ge \frac{\delta}{A_{\mathcal{I}}(\mathcal{D}_k)}.$$

Apply the lower half of equation 27 and use $U \le u_{\max}$ to obtain equation 29. The bound includes the endpoint corrections created by removing the target neighborhood.

Finally, if u is badly approximable, then for every $d \in \mathcal{D}_k$,

$$\|du\|_{\mathbb{Z}} \ge \frac{c_{\mathrm{BA}}(u)}{|d|} \ge \frac{c_{\mathrm{BA}}(u)}{D_k^{\max}}.$$

Substitution into the lower half of equation 27 proves equation 30. □

Remark C.10 (What the conditioning theorem says). The exact source-time theorem distinguishes equality from non-equality. The quantity κ_{ans} refines that binary distinction: it measures the distance to the nearest answer-changing temporal resonance. In particular, an irrational source ratio can be exactly identifying yet statistically useless when it lies $O(\varepsilon)$ from a harmful rational, in which case the fixed-confidence cost is $\Omega(\varepsilon^{-2} \log(1/\eta))$.

D COUPLED CONTINUOUS–DISCRETE QUOTIENT LEARNING

This appendix proves Theorem 7.1. The continuous damping coordinate and the discrete logarithm branch enter the same categorical endpoint law, and the target answer depends on both. The learner knows this parametric family, the legal source ratios, and the target reward, but not the true damping, branch, endpoint probabilities, or observable quotient table.

D.1 THE COUPLED CONTROLLED FAMILY

Let

$$\mathcal{K}_{\mathrm{cpl}} := \{-3,-2,1,2\}, \qquad \mathcal{U}_{\mathrm{cpl}} := \{u_1,u_2\} := \{1/4,3/4\}, \qquad r := 1/2.$$

Fix $0 < R < 1$, $b > 0$, $\Delta > 0$, and $\alpha_- < \alpha_+$ satisfying

$$\alpha_- > 5\pi/\sqrt{3}, \qquad 2\alpha_+/h \le \Lambda, \qquad R + b \le 1, \tag{66}$$

$$\frac{Re^{-3\alpha_+/2}}{\sqrt{3}} + \Delta < b < \frac{Re^{-3\alpha_-/2}}{\sqrt{3}} - \Delta. \tag{67}$$

For $\theta = (\alpha, k)$ define

$$Q^{\theta}_{\mathrm{safe}} := 0, \qquad Q^{\theta}_{\mathrm{risk}} := h^{-1}\big(3\alpha J + (2k+1)A_{\mathrm{rot}}\big).$$

The one-step target task has terminal reward $R(1/2,1,0)^{\top}$, zero risky stage reward, safe stage reward b at state 1, and zero stage reward at states $2,3$. Its state-1 risky-minus-safe margin is

$$\Gamma(\alpha,k) = \frac{Re^{-3\alpha/2}}{\sqrt{3}}(-1)^k - b. \tag{68}$$

The strict-margin parameter set is

$$\Theta_\Delta := \{(\alpha,k) : \alpha \in [\alpha_-,\alpha_+],\ k \in \mathcal{K}_{\mathrm{cpl}},\ |\Gamma(\alpha,k)| \ge \Delta\}. \tag{69}$$

It is a finite union of compact intervals. Let $c(\theta) \in \{\mathrm{s},\mathrm{r}\}$ be the sign answer determined by Γ.

At a source ratio $u \in \mathcal{U}_{\mathrm{cpl}}$, the learner resets to state 1, applies the risky action, and observes the endpoint $Y \in \{1,2,3\}$. Write

$$p_u^{\alpha,k} := e_1^{\top}\exp(u[3\alpha J + (2k+1)A_{\mathrm{rot}}])\,. \tag{70}$$

Both α and k enter this same three-category law.

Lemma D.1 (Target policy and strict answer classes)**.** *Under equation 66– equation 69, safe is uniquely optimal at state* 2*, risky is uniquely optimal at state* 3*, and the unique state-*1 *action is* $c(\alpha,k)$*. The answer map has a margin at least* Δ*.*

Proof. The terminal reward is the reward in Theorem 5.1 scaled by R. Therefore the state-2 and state-3 risky-minus-safe gaps are R times equation 15 and equation 16, and retain their strict negative and positive signs. At state 1, scaling the continuation reward gives $Re^{-3\alpha/2}(-1)^k/\sqrt{3}$ and the safe stage reward subtracts b, which is equation 68. The definition of Θ_Δ excludes ties and supplies the declared margin. The normalization follows from $R + b \le 1$. □

Lemma D.2 (Exact legal-source fibers)**.** *For* $\theta = (\alpha,k)$ *and* $\vartheta = (\alpha',\ell)$ *in* Θ_Δ*,*

$$p_u^{\alpha,k} = p_u^{\alpha',\ell} \quad \forall u \in \{1/4,3/4\} \quad \Longleftrightarrow \quad \alpha = \alpha' \text{ and } k - \ell \in 4\mathbb{Z}. \tag{71}$$

Thus every exact source fiber is either $\{(\alpha,-3),(\alpha,1)\}$ *or* $\{(\alpha,-2),(\alpha,2)\}$*, and the answer is constant on each fiber.*

Proof. Let $c_0 = (1/3,1/3,1/3)$ be the first row of the averaging matrix. From the closed-form circulant kernel,

$$p_u^{\alpha,k} - c_0 = -e^{-3\alpha u}\cos((2k+1)\pi u)e_1^{\top}J + e^{-3\alpha u}\frac{\sin((2k+1)\pi u)}{\pi}e_1^{\top}A_{\mathrm{rot}}.$$

The two row vectors on the right are orthogonal and both have squared norm $2/3$. Hence

$$\|p_u^{\alpha,k} - c_0\|_2^2 = \frac{2}{3} e^{-6\alpha u}. \tag{72}$$

Equality of one positive-duration source row therefore forces $\alpha = \alpha'$. With the damping fixed, the single-time equality criterion from Theorem 5.1 says $(k - \ell)u \in \mathbb{Z}$. Requiring this at $u = 1/4$ and $u = 3/4$ is equivalent to $k - \ell \in 4\mathbb{Z}$. The four-branch set has exactly the two pairs listed above. Since congruence modulo four preserves parity, it also preserves equation 68 and therefore the answer. □

The strict inequality $\alpha_- > 5\pi/\sqrt{3}$ makes every risky generator irreducible with positive off-diagonal rates. Compactness and positive source durations therefore give a common outcome-support constant

$$\nu := \min_{\theta \in \Theta_\Delta,\, u \in \mathcal{U}_{\mathrm{cpl}},\, y} p_u^\theta(y) > 0. \tag{73}$$

In particular all log-likelihood ratios and their α derivatives are uniformly bounded.

D.2 Answer information and latent nonidentifiability

Let $\mathcal{E} := \{1, 2\}$ index the two legal source ratios. For $w \in \Delta_2$, define

$$D_w(\theta, \vartheta) := \sum_{e=1}^{2} w_e D_{\mathrm{KL}}(p_{u_e}^\theta \| p_{u_e}^\vartheta),$$

$$\mathcal{I}(\theta, w) := \inf_{\vartheta : c(\vartheta) \neq c(\theta)} D_w(\theta, \vartheta), \qquad C_{\mathrm{ans}}^\star(\theta) := \max_{w \in \Delta_2} \mathcal{I}(\theta, w), \tag{74}$$

with optimal-design set $\mathcal{W}^\star(\theta) := \arg\max_w \mathcal{I}(\theta, w)$.

Lemma D.3 (Uniform likelihood and compensator regularity)**.** *There are finite constants $L_{\log}, L_D$ such that, for every experiment e, endpoint y, branch k, and α, α' in the same branch component of Θ_Δ,*

$$\left| \log p_{u_e}^{\alpha,k}(y) - \log p_{u_e}^{\alpha',k}(y) \right| \le L_{\log} |\alpha - \alpha'|, \tag{75}$$

$$|D_w(\theta, (\alpha, k)) - D_w(\theta, (\alpha', k))| \le L_D |\alpha - \alpha'| \qquad \forall w \in \Delta_2. \tag{76}$$

For any predictable design sequence (λ_s), define $A_t(\vartheta) := \sum_{s=1}^t D_{\lambda_s}(\theta, \vartheta)$. Then

$$|Z_t(\theta, (\alpha, k)) - Z_t(\theta, (\alpha', k))| \le t L_{\log} |\alpha - \alpha'|,$$
$$|A_t(\alpha, k) - A_t(\alpha', k)| \le t L_D |\alpha - \alpha'|.$$

Proof. For fixed k, e, y, the map $\alpha \mapsto p_{u_e}^{\alpha,k}(y)$ is C^1 on a compact interval. Equation equation 73 bounds it away from zero, so the derivative of its logarithm is uniformly bounded. This proves equation 75. With the true parameter fixed,

$$D_w(\theta, \vartheta) = \sum_{e,y} w_e p_{u_e}^\theta(y) \log \frac{p_{u_e}^\theta(y)}{p_{u_e}^\vartheta(y)}.$$

Differentiating only the second logarithm and using the same uniform bound proves equation 76, uniformly over the simplex. Summation over t rounds gives the final two inequalities. □

Lemma D.4 (Positive answer information, zero branch information)**.** *For every $\theta \in \Theta_\Delta$, $C_{\mathrm{ans}}^\star(\theta) > 0$. On the other hand, if*

$$C_{\mathrm{branch}}^\star(\theta) := \max_{w \in \Delta_2} \inf_{\vartheta : k(\vartheta) \neq k(\theta)} D_w(\theta, \vartheta),$$

then $C_{\mathrm{branch}}^\star(\theta) = 0$.

Proof. Take the uniform design $w_0 = (1/2, 1/2)$. The opposite-answer set is compact. If its minimum D_{w_0} were zero, both legal source rows would agree, so Lemma D.2 would place the alternative in the same source fiber as θ, contradicting answer compatibility. Continuity and compactness therefore give a strictly positive minimum, and hence $C^{\star}_{\mathrm{ans}}(\theta) > 0$.

Every branch in $\mathcal{K}_{\mathrm{cpl}}$ has a different branch partner at displacement four and the same damping. Lemma D.2 shows that this partner has exactly the same law under both legal experiments. Thus the branch-alternative infimum is zero for every w. □

D.3 Algorithm: regularized profile design and categorical mixture confidence

For observations $(E_s, Y_s)_{s \le t}$, let

$$\ell_t(\theta) := \sum_{s=1}^{t} \log p^{\theta}_{u_{E_s}}(Y_s), \qquad \widehat{\theta}_t := \min \arg\max_{\theta \in \Theta_\Delta} \ell_t(\theta),$$

where the minimum is taken in a fixed lexicographic order over branch and continuous coordinate. Put

$$\rho_t := (t+1)^{-1/6}, \qquad \gamma_t := (t+1)^{-1/3}, \qquad w_0 := (1/2, 1/2),$$

and define the unique regularized optimizer

$$w_t := \arg\max_{w \in \Delta_2} \{\mathcal{I}(\widehat{\theta}_t, w) - \rho_t \|w\|_2^2\}. \tag{77}$$

The objective is strictly concave because $\mathcal{I}(\theta, \cdot)$ is concave. Conditional on the history, draw

$$E_{t+1} \sim \lambda_{t+1} := (1 - \gamma_t) w_t + \gamma_t w_0. \tag{78}$$

Both source ratios are legal and target excluding.

For experiment e and endpoint y, let $N_{e,y}(t)$ be the count through time t and $N_e(t) = \sum_y N_{e,y}(t)$. For counts $n = (n_1, n_2, n_3)$ with $N = \sum_y n_y$, define the uniform-Dirichlet mixture probability

$$m(n) := \int_{\Delta_3} \prod_{y=1}^{3} q_y^{n_y} \, 2 \, dq = \frac{2 \prod_{y=1}^{3} n_y!}{(N+2)!}. \tag{79}$$

For a candidate ϑ, set

$$M_t(\vartheta) := \prod_{e=1}^{2} \frac{m((N_{e,y}(t))_{y=1}^{3})}{\prod_{y=1}^{3} p^{\vartheta}_{u_e}(y)^{N_{e,y}(t)}}. \tag{80}$$

Define

$$\mathcal{C}_t(\eta) := \{\vartheta \in \Theta_\Delta : M_t(\vartheta) < \eta^{-1}\}.$$

The stopping time is the first t for which $\{c(\vartheta) : \vartheta \in \mathcal{C}_t(\eta)\}$ has cardinality at most one. If it is a singleton, return its element; if the confidence set is empty, return $c(\widehat{\theta}_t)$.

Lemma D.5 (Measurable implementation)**.** *The lexicographically selected MLE $\widehat{\theta}_t$, the regularized design w_t, the sampling kernel λ_{t+1}, the stopping time, and the output answer are measurable functions of the observed history. Moreover, w_t is the unique maximizer in equation 77 and depends continuously on $\widehat{\theta}_t$ within each compact branch component.*

Proof. At fixed t the history space is finite. For every history, $\theta \mapsto \ell_t(\theta)$ is continuous on the finite union of compact intervals Θ_Δ, so its argmax is nonempty and compact. Choosing first the smallest branch in the declared order and then the smallest continuous maximizer defines a deterministic Borel selector. The function $(\theta, w) \mapsto \mathcal{I}(\theta, w)$ is continuous by uniform source positivity and Berge's maximum theorem, and it is concave in w. Subtracting $\rho_t \|w\|_2^2$ makes the objective strictly concave on the compact simplex; hence the optimizer is unique, and Berge's theorem gives continuity on each branch component.

For an answer $a \in \{\mathrm{s}, \mathrm{r}\}$, let $\Theta_a := \{\vartheta : c(\vartheta) = a\}$. Continuity of $M_t(\vartheta)$ and compactness of Θ_a imply that $\min_{\vartheta \in \Theta_a} M_t(\vartheta)$ is a measurable function of the finite history. Answer a is represented in $\mathcal{C}_t(\eta)$ exactly when this minimum is below η^{-1}. Therefore the stopping event and the stated deterministic output rule are measurable. □

Lemma D.6 (Mixture martingale and universal-coding bounds)**.** *For every fixed ϑ, $(M_t(\vartheta))_{t\geq 0}$ is a nonnegative unit-mean martingale under ϑ. Let $\widehat{p}_{e,t}(y) = N_{e,y}(t)/N_e(t)$ when $N_e(t) > 0$ and define*

$$Q_t(\vartheta) := \sum_{e=1}^{2} N_e(t) D_{\mathrm{KL}}(\widehat{p}_{e,t} \| p^{\vartheta}_{u_e}).$$

Then, with $R_t := 4\log(t+3)$,

$$Q_t(\vartheta) - R_t \leq \log M_t(\vartheta) \leq Q_t(\vartheta). \tag{81}$$

Moreover, under true θ,

$$\log M_t(\vartheta) \geq Z_t(\theta,\vartheta) - R_t, \qquad Z_t(\theta,\vartheta) := \ell_t(\theta) - \ell_t(\vartheta). \tag{82}$$

Proof. Using the uniform Dirichlet prior independently for the two experiments, equation 80 is an integral of likelihood ratios $\prod_{s\leq t} q_{E_s}(Y_s)/p^{\vartheta}_{E_s}(Y_s)$. Predictable experiment selection does not enter the likelihood ratio, and the conditional expectation of every update is one under ϑ. Tonelli therefore gives the martingale claim.

For one experiment, let $\widehat{L} = \prod_y (n_y/N)^{n_y}$ be the saturated maximum likelihood. Because a Dirichlet mixture is an average, $m(n) \leq \widehat{L}$. Conversely, the multinomial probability of the type n under its empirical law is at most one:

$$\frac{N!}{\prod_y n_y!}\widehat{L} \leq 1.$$

Hence

$$m(n) = \frac{2\prod_y n_y!}{(N+2)!} \geq \frac{2N!}{(N+2)!}\widehat{L} = \frac{2}{(N+1)(N+2)}\widehat{L}.$$

Taking logarithms, summing over two experiments, and using $N_e(t) \leq t$ gives equation 81. Finally, $Q_t(\vartheta)$ is the saturated log likelihood minus $\ell_t(\vartheta)$, which is at least $\ell_t(\theta) - \ell_t(\vartheta)$, proving equation 82. □

Lemma D.7 (Uniform time-valid correctness)**.** *The confidence-set procedure is uniformly η-correct.*

Proof. Under true θ, Ville's inequality and Lemma D.6 give

$$\mathbb{P}_\theta\left(\sup_t M_t(\theta) \geq \eta^{-1}\right) \leq \eta.$$

On the complementary event, $\theta \in \mathcal{C}_t(\eta)$ for every t. Hence the confidence set is never empty there. Whenever the procedure stops on this event, the represented answer set is nonempty and its unique answer equals $c(\theta)$. The empty-set fallback can therefore be activated only on the coverage-failure event and cannot increase the total error beyond η. □

D.4 Fiber consistency and optimal-design tracking

Let

$$d_{\mathrm{src}}(\theta,\vartheta) := \max_{e=1,2} \|p^{\theta}_{u_e} - p^{\vartheta}_{u_e}\|_1, \qquad d_{\mathrm{src}}(\vartheta, F(\theta)) := \inf_{\zeta\in F(\theta)} d_{\mathrm{src}}(\vartheta,\zeta).$$

The functions $\log p^{\alpha,k}_{u_e}(y)$ are uniformly Lipschitz in α on the compact family; write L for a common constant and $B := \log(1/\nu)$ for a log-likelihood bound.

Lemma D.8 (MLE localization with summable tail)**.** *For every $\epsilon > 0$ there exist $A_\epsilon, a_\epsilon > 0$ such that*

$$\mathbb{P}_\theta\Big(d_{\mathrm{src}}(\widehat{\theta}_t, F(\theta)) > \epsilon\Big) \leq A_\epsilon t^2 e^{-a_\epsilon t^{1/3}}. \tag{83}$$

Consequently $d_{\mathrm{src}}(\widehat{\theta}_t, F(\theta)) \to 0$ almost surely.

Proof. If $d_{\mathrm{src}}(\theta, \vartheta) \geq \epsilon$, one of the two experiments has ℓ_1 distance at least ϵ. Pinsker and the uniform baseline give

$$D_{w_0}(\theta, \vartheta) \geq \frac{1}{4}\epsilon^2.$$

At round s, the sampling distribution in equation 78 therefore has

$$D_{\lambda_s}(\theta, \vartheta) \geq \gamma_{s-1}\epsilon^2/4.$$

The cumulative conditional drift of $Z_t(\theta, \vartheta)$ is at least $c\epsilon^2 t^{2/3}$. Its centered increments are bounded by $2B$. Cover each of the four compact α intervals by a grid of mesh t^{-2}; the grid has at most Ct^2 points. Azuma–Hoeffding at every grid point, followed by a union bound, gives $Ct^2 \exp(-c'\epsilon^4 t^{1/3})$. The Lipschitz error between a parameter and its grid representative is at most Lt^{-1} in total log likelihood and is absorbed by half the $t^{2/3}$ drift for all large t.

If the MLE is farther than ϵ from the true source fiber, some such parameter has likelihood at least that of θ, i.e. its log-likelihood ratio is nonpositive. The preceding bound proves equation 83, after changing constants for finitely many small times. The right-hand side is summable, so Borel–Cantelli gives the almost-sure conclusion. □

The function $\mathcal{I}(\theta, w)$ in equation 74 is continuous in (θ, w), constant on exact source fibers, and concave in w. The first two facts follow from uniform positivity, compactness of the opposite-answer set, and Berge's maximum theorem. Answer compatibility and compactness also imply a positive source-law distance between a source fiber and the opposite-answer set.

Lemma D.9 (Asymptotically optimal legal design)**.** *Under true θ,*

$$\mathcal{I}(\theta, w_t) \longrightarrow C^{\star}_{\mathrm{ans}}(\theta) \qquad \textit{a.s.} \tag{84}$$

If

$$\widehat{w}_t := \frac{1}{t}\sum_{s=1}^{t} e_{E_s} \in \Delta_2$$

is the empirical experiment-frequency vector, then

$$\mathrm{dist}(\widehat{w}_t, \mathcal{W}^{\star}(\theta)) \to 0 \qquad \textit{a.s.} \tag{85}$$

Proof. By Lemma D.8, every limit point of $\widehat{\theta}_t$ belongs to $F(\theta)$, and all such points have the same answer and the same function $\mathcal{I}$. Uniform continuity therefore gives $\sup_w |\mathcal{I}(\widehat{\theta}_t, w) - \mathcal{I}(\theta, w)| \to 0$. Let $w^{\star} \in \mathcal{W}^{\star}(\theta)$. Optimality in equation 77 gives

$$\mathcal{I}(\widehat{\theta}_t, w_t) - \rho_t \|w_t\|^2 \geq \mathcal{I}(\widehat{\theta}_t, w^{\star}) - \rho_t \|w^{\star}\|^2.$$

Since $\rho_t \to 0$, this proves equation 84.

The conditional sampling distribution λ_t differs from w_{t-1} by at most $2\gamma_{t-1}$ in ℓ_1, and $\gamma_t \to 0$. Martingale strong laws for the bounded experiment indicators imply that empirical frequencies and the Cesàro mean of λ_t have the same limit points. Concavity of $\mathcal{I}(\theta, \cdot)$ gives

$$\mathcal{I}\left(\theta, \frac{1}{t}\sum_{s=1}^{t}\lambda_s\right) \geq \frac{1}{t}\sum_{s=1}^{t}\mathcal{I}(\theta, \lambda_s) \longrightarrow C^{\star}_{\mathrm{ans}}(\theta).$$

Every limit point is therefore optimal. Compactness of the simplex yields equation 85. □

Lemma D.10 (Stable answer-optimal design time)**.** *Fix a true θ and write $C = C^{\star}_{\mathrm{ans}}(\theta) > 0$. For every $\varepsilon \in (0, C)$ there exist constants $A_\varepsilon, a_\varepsilon > 0$ and an almost surely finite random time T_ε such that*

$$\mathbb{P}_\theta(T_\varepsilon > n) \leq A_\varepsilon e^{-a_\varepsilon n^{1/3}}, \tag{86}$$

and, for every $t \geq T_\varepsilon$,

$$\inf_{c(\vartheta) \neq c(\theta)} D_{\lambda_{t+1}}(\theta, \vartheta) \geq C - \varepsilon. \tag{87}$$

Proof. Because $\mathcal{I}$ is continuous and constant on the exact source fiber, there is $\delta > 0$ such that

$$d_{\mathrm{src}}(\zeta, F(\theta)) \le \delta \quad \Longrightarrow \quad \sup_{w \in \Delta_2} |\mathcal{I}(\zeta, w) - \mathcal{I}(\theta, w)| \le \varepsilon/8.$$

Let $w^\star \in \mathcal{W}^\star(\theta)$. On this source-neighborhood, optimality of w_t gives

$$\mathcal{I}(\theta, w_t) \ge C - \varepsilon/4 - \rho_t,$$

because $\|w\|_2^2 \le 1$ on the simplex. Choose a deterministic t_0 so that $\rho_t \le \varepsilon/4$ and $\gamma_t(C+1) \le \varepsilon/2$ for all $t \ge t_0$. Concavity and nonnegativity of $\mathcal{I}$ then imply

$$\mathcal{I}(\theta, \lambda_{t+1}) \ge (1 - \gamma_t)\mathcal{I}(\theta, w_t) \ge C - \varepsilon.$$

Define T_ε as the maximum of t_0 and the first time after which the MLE remains in the preceding source-neighborhood forever. By Lemma D.8 and a union bound over all $t \ge n$,

$$\mathbb{P}_\theta(T_\varepsilon > n) \le \sum_{t \ge n} A_\delta t^2 e^{-a_\delta t^{1/3}} \le A_\varepsilon e^{-a_\varepsilon n^{1/3}}$$

after adjusting constants. Finally, $\mathcal{I}(\theta, \lambda) = \inf_{c(\vartheta) \ne c(\theta)} D_\lambda(\theta, \vartheta)$, which proves equation 87. □

D.5 Evidence growth and stopping complexity

For an alternative ϑ, let

$$A_t(\vartheta) := \sum_{s=1}^{t} D_{\lambda_s}(\theta, \vartheta).$$

The centered process $Z_t(\theta, \vartheta) - A_t(\vartheta)$ is a bounded-increment martingale.

Lemma D.11 (Uniform answer-evidence rate)**.** *Under true θ,*

$$\frac{1}{t} \inf_{c(\vartheta) \ne c(\theta)} Z_t(\theta, \vartheta) \longrightarrow C^\star_{\mathrm{ans}}(\theta) \qquad \textit{a.s.}, \tag{88}$$

and

$$\frac{1}{t} \inf_{c(\vartheta) \ne c(\theta)} \log M_t(\vartheta) \longrightarrow C^\star_{\mathrm{ans}}(\theta) \qquad \textit{a.s.} \tag{89}$$

Proof. For each s,

$$\inf_{c(\vartheta) \ne c(\theta)} D_{\lambda_s}(\theta, \vartheta) \ge (1 - \gamma_{s-1})\mathcal{I}(\theta, w_{s-1}) + \gamma_{s-1}\mathcal{I}(\theta, w_0).$$

Lemma D.9 and Cesàro averaging imply

$$\liminf_t \frac{1}{t} \inf_{c(\vartheta) \ne c(\theta)} A_t(\vartheta) \ge C^\star_{\mathrm{ans}}(\theta).$$

Conversely, $t^{-1} \sum_s \lambda_s$ is a feasible design, so the same infimum divided by t is at most $C^\star_{\mathrm{ans}}(\theta)$. Thus the conditional-drift minimum converges to the characteristic information.

For uniform martingale convergence, cover each compact continuous branch component by a mesh of size t^{-2}. The number of grid points is $O(t^2)$. Azuma–Hoeffding controls an ϵt deviation at one grid point by $2e^{-c\epsilon^2 t}$; the union over the polynomial grid is summable. Lemma D.3 controls both the empirical log-likelihood ratio and its predictable compensator between a parameter and its grid representative. Their total off-grid errors are respectively at most $L_{\log}/t$ and L_D/t. Borel–Cantelli therefore gives

$$\sup_{\vartheta \in \Theta_\Delta} \frac{|Z_t(\theta, \vartheta) - A_t(\vartheta)|}{t} \to 0 \quad \text{a.s.}$$

This proves equation 88.

By equation 82, the mixture-evidence infimum is at least the likelihood-ratio infimum minus $R_t = o(t)$. For the reverse inequality, the upper coding bound and the identity

$$Q_t(\vartheta) = Q_t(\theta) + Z_t(\theta, \vartheta)$$

give

$$\inf_{c(\vartheta) \ne c(\theta)} \log M_t(\vartheta) \le Q_t(\theta) + \inf_{c(\vartheta) \ne c(\theta)} Z_t(\theta, \vartheta).$$

For each experiment, empirical endpoint frequencies converge to the true law; hence $Q_t(\theta)/t \to 0$. This proves equation 89. □

Equation equation 89 and positivity of $C^\star_{\mathrm{ans}}(\theta)$ imply almost-sure finite stopping: eventually every opposite-answer candidate is excluded, while same-answer candidates may remain in the confidence set. This is quotient learning rather than exact parameter recovery.

Lemma D.12 (Adaptive information lower bound)**.** *Every uniformly η-correct source-only procedure with finite expected stopping time under θ satisfies*

$$\mathbb{E}_\theta[\tau] \geq \frac{\mathrm{kl}(1-\eta,\eta)}{C^\star_{\mathrm{ans}}(\theta)}. \tag{90}$$

Proof. For every opposite-answer ϑ, stopped likelihood-ratio factorization and binary data processing on the returned answer give

$$\sum_{e=1}^{2} \mathbb{E}_\theta[N_e(\tau)] D_{\mathrm{KL}}(p_e^\theta \| p_e^\vartheta) \geq \mathrm{kl}(1-\eta,\eta).$$

Normalize the expected counts by $\mathbb{E}_\theta[\tau]$. The resulting vector is a feasible design in Δ_2; minimizing over opposite-answer parameters and then comparing with equation 74 proves the claim. If the expectation is infinite, the inequality is immediate. □

Lemma D.13 (Expected stopping-time upper bound)**.** *For every true θ,*

$$\limsup_{\eta \downarrow 0} \frac{\mathbb{E}_\theta[\tau_\eta]}{\log(1/\eta)} \leq \frac{1}{C^\star_{\mathrm{ans}}(\theta)}. \tag{91}$$

Proof. Write $C = C^\star_{\mathrm{ans}}(\theta) > 0$ and $b = \log(1/\eta)$. Fix $\epsilon \in (0, C/4)$ and let T_ϵ be the stable-design time from Lemma D.10. Thus

$$\mathbb{P}_\theta(T_\epsilon > n) \leq A_\epsilon e^{-a_\epsilon n^{1/3}},$$

and every round after T_ϵ has opposite-answer information at least $C - \epsilon$.

Let $m_n = \lfloor n^{3/4} \rfloor$. On $\{T_\epsilon \leq m_n\}$,

$$\inf_{c(\vartheta) \neq c(\theta)} A_n(\vartheta) \geq (C-\epsilon)(n - m_n). \tag{92}$$

Choose

$$n_b := \left\lceil \frac{(1+4\epsilon/C)b}{C} \right\rceil.$$

Since $m_n = o(n)$ and $R_n = O(\log n)$, for all sufficiently large b and $n \geq n_b$, the right-hand side of equation 92 exceeds $b + R_n + \epsilon n$. If $\tau_\eta > n$ on the good-design event, some opposite-answer ϑ has $\log M_n(\vartheta) < b$. By equation 82, this forces

$$Z_n(\theta,\vartheta) - A_n(\vartheta) \leq -\epsilon n.$$

The same n^{-2} net and bounded-increment Azuma argument used in Lemma D.11 gives constants $C_1, c_1 > 0$ such that

$$\mathbb{P}_\theta\left(\inf_{c(\vartheta) \neq c(\theta)} [Z_n(\theta,\vartheta) - A_n(\vartheta)] \leq -\epsilon n \right) \leq C_1 n^2 e^{-c_1 n}.$$

Hence, for $n \geq n_b$,

$$\mathbb{P}_\theta(\tau_\eta > n) \leq A_\epsilon e^{-a_\epsilon n^{1/4}} + C_1 n^2 e^{-c_1 n}.$$

Both tails are summable and their sums from n_b onward are $o(b)$. The tail-sum formula yields

$$\mathbb{E}_\theta[\tau_\eta] \leq n_b + o(b).$$

Let first $\eta \downarrow 0$ and then $\epsilon \downarrow 0$ to obtain equation 91. □

Proof of Theorem 7.1. Lemma D.5 makes the procedure a well-defined adapted algorithm, and Lemma D.7 gives uniform time-valid correctness. Lemma D.8 proves equation 37; Lemma D.9 proves equation 38; and Lemma D.11 proves equation 39 and almost-sure stopping. Lemma D.12, Lemma D.13, and $\mathrm{kl}(1-\eta,\eta) \sim \log(1/\eta)$ give equation 40. Finally, Lemma D.4 proves equation 41. The algorithm operates on the original continuous–discrete parameter family and never receives the source quotient or its endpoint laws as a table. □

Table 1: Baselines in the empirical package.

Baseline	family	role
full latent MLE	coupled family	estimates (α, k); branch ties remain
answer-profile likelihood	coupled family	profiles over parameters with each answer
confidence-set stopping rule	coupled family	returns once retained models share an answer
uniform legal duration	clock design	fixed legal source allocation
answer-optimal legal duration	clock design	maximizes harmful-answer KL
branch-optimal legal duration	clock design	maximizes all-branch KL
target-duration oracle	cyclic system	inadmissible deployment-time observations

Remark D.14 (Computational form). Every online optimization is finite dimensional. For each of the four branches, the profile likelihood and opposite-answer inner infimum are one-dimensional optimizations over compact α intervals; the outer design is a strictly concave optimization on the two-point simplex. No matrix-logarithm branch is supplied to the learner.

E Empirical Protocols and Additional Results

All non-oracle experiments use reset endpoint observations and reveal neither intermediate states nor jump times. The target duration is excluded from the legal source domain. The master seed is 20270925. Binary error bars are 95% Wilson intervals; intervals for means use the independent-run standard error. Raw CSV files and the machine-readable summary are in `data/`.

E.1 Mechanism validation: exact and near aliases

The exact pair uses $h = 1$, $\alpha = 1.9$, $\beta = 0.8$, and target ratio $r = 1/2$. We evaluate the closed-form kernels on 801 times in $[0, 2]$. The source Frobenius distances at $t = h, 2h$ are numerically zero, while the target distance is 0.163608, the action gap is 0.0333964, and the exact fiberwise minimax regret is 0.0166982.

For near aliasing, the source duration is $1 + \varepsilon$ with $\varepsilon \in \{.02, .05, .10, .15\}$. At each perturbation we compute the maximum one-query KL κ and simulate the equal-prior likelihood-ratio test at eight values of $N\kappa \in [.05, 10]$. Every point uses 4000 independent multinomial runs. The collapse in Figure 3(a) is therefore a direct check of the transcript-information scaling, not a fit in raw sample count.

E.2 Clock conditioning and legal duration design

The target ratio is $1/2$, the legal source interval is $[.25, .45]$, the branch family is $\{-3, -2, 1, 2\}$, the true branch is 1, and $\alpha = 9.1$. On 801 equally spaced legal ratios we report:

$$\kappa_{\mathrm{ans}}(k, u), \quad C_{\mathrm{ans}}(k, u), \quad C_{\mathrm{branch}}(k, u), \quad \frac{\mathrm{kl}(.99, .01)}{C_{\mathrm{ans}}(k, u)},$$

and the distance to the nearest rational with denominator at most 12. The quotient-matched clock $u = 1/4$ leaves the difference-four same-answer branches exactly aliased, so $C_{\mathrm{branch}} = 0$, but has $C_{\mathrm{ans}} = 2.35818 \times 10^{-6}$. The branch-optimal legal clock is $u = 0.28425$, with $C_{\mathrm{branch}} = 1.26680 \times 10^{-7}$ and $C_{\mathrm{ans}} = 1.44820 \times 10^{-7}$. The legal policy clock therefore has $16.28\times$ the answer information while refusing to resolve a harmless branch ambiguity. We also report $u = .251$, $\sqrt{2} - 1$, the shortest legal clock, and 4000 target-excluding uniform jitter draws.

E.3 Coupled unknown-damping/logarithm-branch study

The empirical family matches Theorem 7.1: $k \in \{-3, -2, 1, 2\}$, source ratios $\{1/4, 3/4\}$, target ratio $1/2$, and a 191-point computational grid on $\alpha \in [9.10, 11.00]$. The theorem itself is continuous;

the grid is only the finite numerical optimizer. We set $R = .8$ and choose the safe offset so the even-branch answer boundary is at $\alpha = 10$. Grid points with target margin below 10^{-8} are excluded.

We test an odd safe instance $(9.2, 1)$ and an even risky instance $(9.2, 2)$. For each instance the legal allocation is optimized by a 1001-point weight grid, then fixed-budget profile likelihood is evaluated at $NC^{\star}_{\mathrm{ans}} \in \{.25, .5, 1, 2, 4, 8\}$. Every point uses 5000 independent categorical-count runs. The latent truth is randomized uniformly over the exact source-equivalent branch pair before scoring branch error. We report: answer error and Wilson interval, branch error, damping RMSE, profile-evidence rate, target regret, target-kernel error, and

$$\mathrm{Err}_Q := \|\widehat{Q}_{\mathrm{risk}} - Q_{\mathrm{risk}}\|_{\mathrm{F}}.$$

No branch alignment is applied. At normalized budget eight, answer error is .0036 for the odd-safe instance and zero in 5000 runs for the even-risky instance; branch errors are .5074 and .4930. Mean evidence rates are 1.051 and 1.072 times the numerically computed characteristic information. Mean generator errors remain 18.43 and 17.75, while target-kernel errors are below 4×10^{-7}. Here $\mathrm{Err}_Q = \|\widehat{Q}_{\mathrm{risk}} - Q_{\mathrm{risk}}\|_{\mathrm{F}}$ is the unnormalized Frobenius error, reported as a mean with no alignment over source-equivalent branches.

E.4 Actual confidence-set stopping rule

We additionally implement the stopping rule of Theorem 7.1 at $\eta \in \{.1, .05, .02, .01, .005\}$. The continuous damping interval is represented by the frozen 0.01 grid used above; the lexicographic grid MLE, regularized answer-optimal design, forced exploration, Dirichlet-mixture confidence set, and empty-set fallback otherwise match the theorem. To make large characteristic times computationally tractable, categorical sufficient statistics are updated in batches of 10^6 observations and the stopping condition is checked at batch boundaries. This discretized batched study is implementation validation, not part of the proof.

Each instance uses 500 independent trajectories coupled across all five confidence levels. No trajectory reaches the declared 8×10^8 cap, and no wrong answer is observed; the common zero-error Wilson upper endpoint is 0.00763. Mean completed stopping times range from 4.97×10^7 to 5.46×10^7 on the odd-safe instance and from 3.18×10^8 upward on the even-risky instance. Branch error at stopping remains between 0.50 and 0.544. We report the mean and its interval, median, 90th and 99th percentiles, nontermination, fail-closed error, and $\mathbb{E}[\tau]C^{\star}_{\mathrm{ans}}/\log(1/\eta)$ in `data/coupled_sequential.csv`. The normalized ratios remain well above one at these moderate confidence levels, illustrating the finite-sample coding overhead rather than asserting premature asymptotic calibration.

E.5 Reward-relevant cyclic-maintenance testbed

To avoid an independent tensor nuisance, we also use a four-state directed maintenance cycle. Let C_4 be the clockwise four-cycle permutation and

$$Q_+ = \pi(C_4 - I), \qquad Q_- = \pi(C_4^{-1} - I).$$

At source duration one, $e^{Q_+} = e^{Q_-}$ to numerical precision, while the half-time kernels differ. The safe action is static and the terminal reward is $g = (1/2, 1, 1/2, 0)^{\top}$. From state one the risky values are .603940 and .396060, so the hidden orientation directly reverses the control action with gap .103940. Under a uniform hidden orientation, every source-only deterministic branch tie-break has policy error $1/2$ and mean regret .051970. For comparison only, an explicitly inadmissible half-time-data oracle is simulated at budgets $1, 2, 5, 10, 20, 50, 100$, using 5000 runs per point.

E.6 Reproduction

Run

```
python experiments/run_empirical_package.py
python experiments/postprocess_empirical.py
python experiments/make_architecture_figures.py
```

from the project root. The scripts regenerate every CSV, JSON summary, table, and figure used by the manuscript. No external data are downloaded.

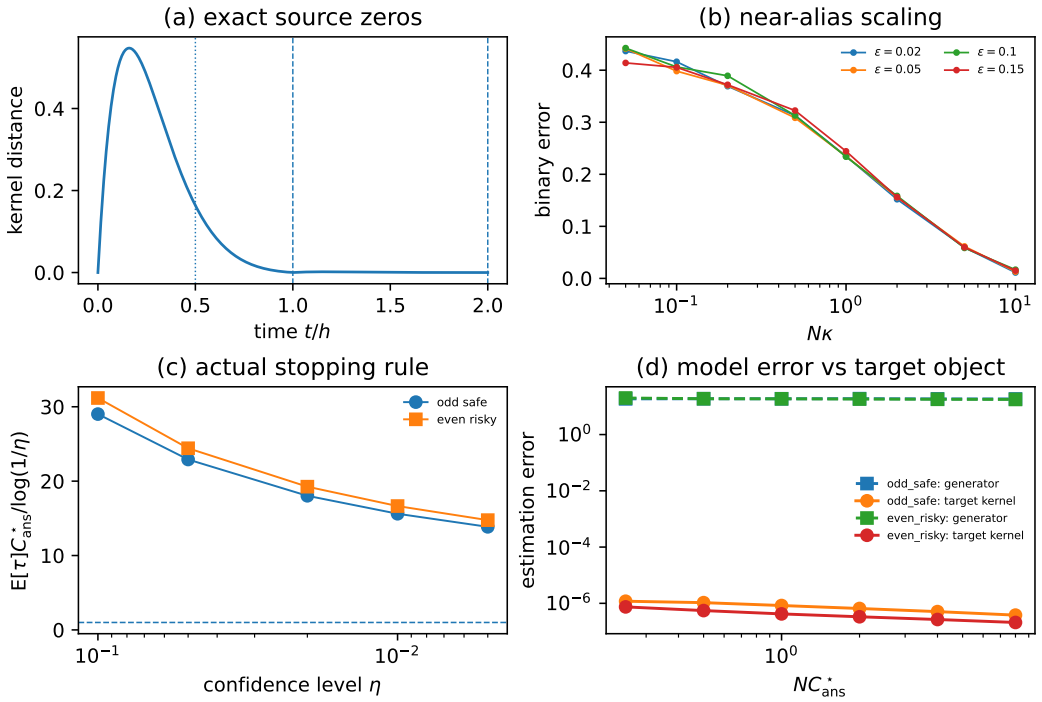

Figure 4: Supplementary diagnostics: exact source-time zeros, near-alias scaling, the actual confidence-set stopping rule, and persistent generator error despite convergence of the target kernel.

F Reproducibility and Claim Audit

The archive contains one focused source tree, all retained raw empirical outputs, complete proofs of the headline results, a one-page theorem-obligation ledger, and a fail-closed one-command runner. The former local-LAN, nonregular-rate, generic multiple-answer, independent-Bernoulli, and sequential-propagation developments are outside the compiled input closure.

The strict runner regenerates all retained data, tables, and figures; compiles the paper; and checks:

1. semigroup algebra, Bézout reconstruction, source aliases, rate bounds, and arbitrary-target action reversal;
2. finite-quotient legal-design lower and upper bounds, primal–dual agreement, sparse support, exploration, and likelihood-ratio calibration;
3. KL–Diophantine conditioning, near-rational hardness, target-excluding jitter, and badly approximable clocks;
4. the coupled family assumptions, exact source fibers, target margins, zero branch/positive answer information, mixture martingale, coding bounds, and the actual confidence-set stopping study;
5. the title, one-paragraph abstract, exactly three contributions, named closed time group, complete algorithm notation, theorem assumptions, measurability/regularity lemmas, related-work boundary, and three retained main figures;
6. citation and label closure, the exact nine-page main-text boundary, anonymous metadata and sources, absence of stale artifacts, and official style-file hashes.

The analytic proofs are authoritative. Floating-point and fixed-seed results are labeled consistency or mechanism-validation checks and are not used as proof. `THEOREM_AUDIT.md` maps every headline theorem to assumptions, the critical proof step, anonymous human audit role, and deterministic script. Run

```
bash reproducibility/reproduce_all.sh
```

from the project root; any machine-detectable failure returns a nonzero status.

G Related-Work Claim Matrix

Literature	observed object	latent ambiguity	decision target	additional object supplied here
Markov embedding	sampled stochastic kernels	generator logarithms	model recovery	unseen-time target-policy consequences
System aliasing	sampled continuous-time dynamics	aliased generators	system identification	equality of all adaptive endpoint transcripts and opposite control actions
Partial identification	identified set	payoff-relevant parameter	robust decision	semigroup-generated identified set and legal temporal design
Robust MDPs	externally specified ambiguity set	transitions	robust policy	nonconvex inverse image of source-time data evaluated at another time
Temporal-resolution RL	managed observation/control clocks	discretization or interaction cost	value/control efficiency	zero-target-data answer identifiability from legal source clocks
Value equivalence	specified model, function, and policy classes	planning-equivalent models	Bellman operations	statistically induced temporal fiber and target-time quotient
Pure exploration	known statistical model family	answer uncertainty	answer identification	answer classes induced by an unseen-time control task and legal duration choice
This work	endpoint kernels at legal source times	generator fiber	unseen-time policy	group law, control converse, arithmetic conditioning, and coupled answer-optimal learning

Table 2: Novelty boundary. Robust decision-making over an identified set, nonunique matrix logarithms, Bellman equivalence, and sequential-design tools are prior concepts; the contribution is their temporal CTMC synthesis and the matching target-policy experiment design.