ONLY LEARN WHAT CHANGES THE DECISION: EXACT QUOTIENT GEOMETRY FOR ACTIVE ACQUISITION

Anonymous authors
Paper under double-blind review

ABSTRACT

Decision-focused acquisition can certify a decision without recovering every latent coordinate. Our decision-observability quotient groups parameters with identical query means and preserves regret and acceptable answers exactly when the query span contains every decision difference. Under Gaussian observations with positive tolerance, it preserves the information game whose characteristic time gives the exact leading-order fixed-confidence sample complexity. A practical adaptive rule provides finite-confidence guarantees and matches this coefficient under the stated conditions, including exact allocations and an inactive parameter ball. On a fully observable family with n divisible by four, coordinate-wise identification has characteristic time exactly n times that of decision identification, with a stopping-time ratio between n and $4n$ in the stated self-normalized ellipsoid regime. Synthetic experiments exhibit the predicted stopping-time separation. On 18 public rating panels, the quotient test stops in $18{,}993$ of $64{,}800$ matched runs, while the coordinate-wise test never stops on the same samples and confidence events. No incorrect recommendation is observed on held-out raters.

1 INTRODUCTION

Decision makers often choose actions for many individuals or contexts using noisy measurements of a shared latent parameter. A platform promotes one of K items to its users; a scheduler picks a daily load profile; a league ranks teams from game margins. These choices depend on differences between means, but the learner observes individual ratings, meter readings, or game scores. An active learner selects which measurements to collect. It should focus on distinctions that can change the decision and stop once that decision can be certified at a stated confidence level.

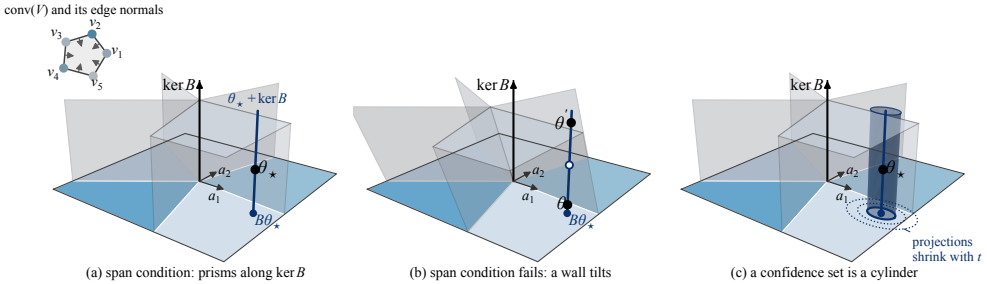
We study this problem in a general linear model. Each policy specifies a decision for every context, with loss linear in a shared parameter $\theta \in \mathbb{R}^d$. Each query observes a linear functional of θ, and the library of query means is $B\theta$. Parameters with the same $B\theta$ cannot be distinguished by any sampling rule. If the queries span every difference between candidate decisions, those parameters also induce the same decision comparisons. We therefore identify them and work on the resulting *decision-observability quotient*.

We ask which measurements preserve every decision comparison when the latent parameter is only partially identifiable and how much sampling is spent on recovering coordinates that the decision does not need. The confidence boundary must also cover several directions that may bind simultaneously at stopping.

Once the observable model is specified, fixed-confidence identification provides methods for problems with several acceptable answers. Track-and-Stop finds a characteristic allocation through a change-of-measure game (Garivier & Kaufmann, 2016); its extension to multiple correct answers introduced the sticky rule (Degenne & Koolen, 2019), and non-asymptotic guarantees now cover both variants (Poiani et al., 2026). Transductive linear bandits distinguish measurement vectors from target vectors (Soare et al., 2014; Xu et al., 2018; Fiez et al., 2019; Degenne et al., 2020). Time-uniform boundaries for linear models (Abbasi-Yadkori et al., 2011; Kaufmann & Koolen, 2021; Howard et al., 2021) and bounded observations (Waudby-Smith & Ramdas, 2024; 2020) provide stopping guarantees, and boundaries restricted to decision directions pay a union over those directions (Kaufmann & Koolen,

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: **Decision-observability quotient.** Queries a_1, a_2 span the observable plane orthogonal to $\ker B$. (a) Fibers stay within decision regions under the span condition. (b) A fiber crosses a decision boundary when it fails. (c) A cylindrical confidence set projects into one decision region.



(a) span condition: prisms along $\ker B$ (b) span condition fails: a wall tilts (c) a confidence set is a cylinder

2021; Mutný & Krause, 2022; Lindon et al., 2026). Work on decision-focused active acquisition also includes methods without fixed-confidence stopping guarantees (Li et al., 2022; Brunskill et al., 2026; Wan et al., 2026; Hübotter et al., 2024).

We derive the observable model from the decision problem itself. The span condition characterizes when the quotient preserves decision comparisons and their information-theoretic difficulty. Under Gaussian observations, the quotient characteristic time gives the exact leading-order fixed-confidence complexity. On explicit families, we establish an n-fold cost of identifying observable coordinates beyond what the decision requires and a necessary union penalty for fixed direction-wise confidence bounds with independent binding directions.

Contributions. *(i) Exact quotient theory.* We establish when the observable quotient preserves the decision. Under the span condition, it preserves regret and the set of ϵ-correct answers; when the condition fails, indistinguishable parameters require different answers below an explicit tolerance (Proposition 3.2). For Gaussian observations, the quotient also preserves the multiple-answer information game, and its characteristic time H_ϵ gives the exact leading-order fixed-confidence sample complexity, even for singular designs (Theorems 3.3 and 3.4). We develop an adaptive rule with finite-confidence correctness and stopping-time guarantees and establish conditions under which it attains the same leading coefficient (Proposition 3.5, Corollary 3.6).

(ii) The cost of identifying more than the decision requires. Coordinate-wise identification has characteristic time exactly n times that of decision identification on a constructed family with $n \in 4\mathbb{N}$, even though every coordinate is observable (Theorem 4.1). In the regime of Theorem 4.2, the self-normalized ellipsoid yields a ratio of coordinate-wise to quotient stopping times between n and $4n$. When the design has a non-trivial kernel, coordinate-wise uncertainty has a nonvanishing floor, while the quotient test can stop on the same confidence event. For fixed per-direction radii, we establish a union penalty on a family with D independent binding directions (Proposition 4.3). We introduce headroom $\rho_{\mathrm{adapt}} \in [1, K]$ to bound the asymptotic gain available from adaptive allocation over round robin.

(iii) Empirical validation on synthetic and real designs. Synthetic experiments show the practical rule approaching characteristic-time scaling and the measured separation between coordinate-wise and quotient tests matching the theoretical bounds. On 18 public rating panels, the coordinate-wise test stops in none of $64{,}800$ runs, whereas the quotient test stops in $18{,}993$ matched runs, with no incorrect recommendation on held-out raters. Controlled experiments show that omitting the direction union exceeds the target error level and that headroom tracks the gain from adaptive allocation over round robin. We also run existing transductive allocation rules directly in the quotient space without modification.

2 Setting

The model couples local decisions through a shared latent parameter and a finite library of queries. Figure 1 shows the relation between query directions, decision comparisons, and the observable quotient.

Decisions and aggregates. Let contexts $i = 1, \dots, n$ have known weights $p_i > 0$, with $\sum_i p_i = 1$. At context i, the policy selects a vertex π_i of a compact polytope $\mathcal{Z}_i$, and its loss is linear in the unknown parameter $\theta \in \mathbb{R}^d$ through a known matrix A_i. The policy $\pi = (\pi_1, \dots, \pi_n)$ induces the aggregate $\phi(\pi) = \sum_i p_i A_i^\top \pi_i$. Policies with the same aggregate have the same loss for every θ, so we identify them. Let $\mathcal{V}$ be the finite set of distinct aggregates and $\mathcal{K} = \mathrm{conv}(\mathcal{V})$. Define regret for $v \in \mathcal{V}$ and the set of acceptable aggregates by

$$r_\theta(v) = \langle \theta, v \rangle - \min_{u \in \mathcal{K}} \langle \theta, u \rangle = \max_{u \in \mathcal{V}} \langle \theta, v - u \rangle, \qquad \mathcal{V}_\epsilon(\theta) = \{v \in \mathcal{V} : r_\theta(v) \le \epsilon\}. \tag{1}$$

Here $\epsilon > 0$ is the fixed tolerance, so $\mathcal{V}_\epsilon(\theta)$ contains the ϵ-correct aggregates. We assume that θ lies in the ball $\Theta = \{\theta : \|\theta\|_2 \le S\}$ for known $S > 0$.

Queries and the span condition. A library of K query vectors $a_1, \dots, a_K \in \mathbb{R}^d$ has mean operator $(B\theta)_j = \langle a_j, \theta \rangle$. At round t, the learner selects query J_t and observes $Y_t = \langle a_{J_t}, \theta_\star \rangle + \xi_t$, where ξ_t is conditionally σ-sub-Gaussian; results requiring Gaussian noise state that assumption separately. Let $\mathcal{Q} = \mathrm{span}\{a_j\}$ be the observable subspace and $\mathcal{L} = \mathrm{span}(\mathcal{V} - \mathcal{V})$ the subspace of decision differences. The central *span condition* is

$$\mathcal{L} \subseteq \mathcal{Q}. \tag{2}$$

The span condition requires only decision comparisons to be observable, so $\ker B$ may be nontrivial and the full parameter need not be identifiable. Parameters with $B\theta = B\theta'$ give the same mean for every query and cannot be distinguished by any adaptive allocation.

Stopping rules. An acquisition procedure selects $J_1, J_2, \dots$ adaptively, stops at time τ, and returns $\hat{v}_\tau \in \mathcal{V}$. Uniform (ϵ, δ)-correctness requires $\Pr_\theta(\tau < \infty,\ \hat{v}_\tau \in \mathcal{V}_\epsilon(\theta)) \ge 1 - \delta$ for every $\theta \in \Theta$. Let $\mathcal{G} = \{v - u : v \ne u \in \mathcal{V}\}$ be the decision directions, $V_t = \lambda I + \sum_{s \le t} a_{J_s} a_{J_s}^\top$ the ridge design matrix, and $\hat{\theta}_t$ the ridge estimator. Every stopping rule in this paper uses the time-uniform confidence event

$$|\langle \hat{\theta}_t - \theta_\star, g \rangle| \le R_t(\delta, g)\, \|g\|_{V_t^{-1}} \qquad \forall t \ge 1,\ \forall g \in \mathcal{G}. \tag{3}$$

The radius is set by a time-uniform boundary $\beta(\delta, t, g)$ (Section A.1). A candidate v is certified once the confidence region excludes every parameter for which its regret exceeds ϵ. We apply two tests to the same event. The *coordinate-wise test* resolves the latent coordinates, whereas the *quotient test* resolves only directions in $\mathcal{G}$. Their comparison isolates the cost of the identification target from the cost of the confidence boundary.

3 Exact Quotient Theory

3.1 The quotient preserves the decision

Proposition 3.1 (Fibers of B are decision-equivalent)**.** *Under (2), $B\theta = B\theta'$ implies $r_\theta(v) = r_{\theta'}(v)$ for every $v \in \mathcal{V}$. Hence, for every $\epsilon \ge 0$, the observable parameter space $\mathcal{M}_{\mathrm{obs}} = B\Theta$ and the answer correspondence $i_\epsilon(B\theta) = \mathcal{V}_\epsilon(\theta)$ are well defined.*

Indeed, $\theta - \theta' \in \ker B = \mathcal{Q}^\perp$ is orthogonal to every decision direction $v - u \in \mathcal{L} \subseteq \mathcal{Q}$, so every pairwise comparison, and therefore every regret, is constant along a fiber of B. The span condition is also necessary.

Proposition 3.2 (The span condition is necessary)**.** *Let P be the orthogonal projection onto $\ker B$ and $\epsilon_0 = \frac{S}{2} \max_{v,u \in \mathcal{V}} \|P(v - u)\|_2$. Then $\epsilon_0 > 0$ if and only if $\mathcal{L} \not\subseteq \mathcal{Q}$, and for every $\epsilon < \epsilon_0$ there exist $\theta, \theta' \in \Theta$ with $B\theta = B\theta'$ and $\mathcal{V}_\epsilon(\theta) \cap \mathcal{V}_\epsilon(\theta') = \varnothing$. Thus no procedure that stops almost surely is (ϵ, δ)-correct at both parameters for any $\delta < 1/2$, whatever the noise law or allocation. Under Gaussian observations, a uniformly (ϵ, δ)-correct procedure exists for every $\epsilon > 0$ if and only if $\mathcal{L} \subseteq \mathcal{Q}$.*

The indistinguishable pair in Proposition 3.2 is $\pm Sz$ for a unit $z \in \ker B$ along which two aggregates differ most; Theorem 3.4 establishes sufficiency. We next characterize the information needed to identify an acceptable answer. For an allocation q in the simplex Δ_K, let $M(q) = \sum_j q_j a_j a_j^\top$, which

may be singular, and let $\mathrm{Alt}_\epsilon(v) = \{\theta' \in \Theta : r_{\theta'}(v) > \epsilon\}$ contain the parameters under which v is not ϵ-correct. Define

$$I_q(\theta, v) = \inf_{\theta' \in \mathrm{Alt}_\epsilon(v)} \frac{\|\theta' - \theta\|^2_{M(q)}}{2\sigma^2}, \qquad \Gamma_\epsilon(\theta) = \max_{v \in \mathcal{V}_\epsilon(\theta)} \max_{q \in \Delta_K} I_q(\theta, v), \qquad H_\epsilon(\theta) = \Gamma_\epsilon(\theta)^{-1}. \tag{4}$$

Here $\inf \varnothing = +\infty$. For the observable Gaussian model, let $D(q, \mu, \mu') = \sum_j q_j(\mu_j - \mu'_j)^2/2\sigma^2$, $\neg_\epsilon v = \{\mu' \in \mathcal{M}_{\mathrm{obs}} : v \notin i_\epsilon(\mu')\}$, and $\mathcal{D}_\epsilon(\mu, v) = \max_q \inf_{\mu' \in \neg_\epsilon v} D(q, \mu, \mu')$.

Theorem 3.3 (Exact quotient information identity)**.** *Under finite weighted polyhedral geometry, a compact parameter ball, a finite query library, and (2), $\neg_\epsilon v = B\,\mathrm{Alt}_\epsilon(v)$, and replacing the strict observable alternative by its relative closure leaves every divergence infimum unchanged. Under independent Gaussian observations, $\mathcal{D}_\epsilon(B\theta, v) = \max_q I_q(\theta, v)$ for $v \in \mathcal{V}_\epsilon(\theta)$ and $\mathcal{D}_\epsilon(B\theta, v) = 0$ otherwise, so that $\max_{v \in \mathcal{V}} \mathcal{D}_\epsilon(B\theta, v) = \Gamma_\epsilon(\theta)$. These identities are invariant to policy duplication, local relabeling, and the choice of representative within an aggregate class.*

Proposition 3.1 makes regret constant on fibers of B, while Gaussian divergence depends on $\theta' - \theta$ only through $B(\theta' - \theta)$. Latent and observable alternatives therefore give the same divergence infima. The quotient discards unobservable directions, all of which cancel from decision comparisons under the span condition. The outer maximum over $\mathcal{V}_\epsilon(\theta)$ captures the freedom to select the easiest answer to distinguish when several are ϵ-correct.

3.2 The characteristic time

Theorem 3.4 (Matching Gaussian characteristic time)**.** *Assume the conditions of Theorem 3.3, $\epsilon > 0$, and $Y_t \sim N(\langle a_{J_t}, \theta\rangle, \sigma^2)$ independently given J_t. Every uniformly (ϵ, δ)-correct procedure satisfies*

$$\liminf_{\delta \downarrow 0} \frac{\mathbb{E}_{\theta_\star}[\tau_\delta]}{\log(1/\delta)} \;\geq\; H_\epsilon(\theta_\star), \tag{5}$$

and Sticky Track-and-Stop run on the observable model, with the oracle allocation and candidate set of Section B.1, attains

$$\limsup_{\delta \downarrow 0} \frac{\mathbb{E}_{\theta_\star}[\tau_\delta]}{\log(1/\delta)} \;\leq\; H_\epsilon(\theta_\star). \tag{6}$$

If some answer is ϵ-correct on all of Θ, then $H_\epsilon = 0$ and it is returned without sampling; otherwise $0 < H_\epsilon(\theta_\star) < \infty$.

The lower and upper bounds match with leading constant one. Theorem 3.3 transfers the multiple-answer information game to the observable model, where the change-of-measure lower bound and Sticky Track-and-Stop analysis of Degenne & Koolen (2019), together with the non-asymptotic guarantees of Poiani et al. (2026), apply even to singular designs.

3.3 The practical rule

The practical rule (Algorithm 1) stops when a direction-dependent confidence bound certifies that the current answer has regret at most ϵ. It replaces the oracle components used to attain the rate in Theorem 3.4 with a ridge estimator, the plug-in candidate set $\mathcal{V}_\epsilon(\hat{\theta}_t)$, and sticky answer selection with hysteresis. It tracks an estimated allocation by C-tracking with forced exploration $f_t = 1/(2\sqrt{K^2 + t})$ (Garivier & Kaufmann, 2016). Because the stopping test works directly with decision directions and does not intersect its confidence region with Θ, its characteristic time uses half-space alternatives. With $\gamma_u(v; \theta) = \epsilon - \langle \theta, v - u\rangle$ and $\rho_M(g) = g^\top M^\dagger g$ for $g \in \mathrm{range}(M)$, $+\infty$ otherwise, define

$$H_v(\theta, q) = \max_{u \neq v} \frac{2\sigma^2 \rho_{M(q)}(v - u)}{\gamma_u(v; \theta)^2_+}, \qquad H_v(\theta) = \min_{q \in \Delta_K} H_v(\theta, q), \qquad H^{\mathrm{hs}}_\epsilon(\theta) = \min_{v \in \mathcal{V}_\epsilon(\theta)} H_v(\theta). \tag{7}$$

The final minimum selects an easiest ϵ-correct answer $v_\star$. Restricting alternatives to Θ can only add information, so $H_\epsilon \leq H^{\mathrm{hs}}_\epsilon$, with equality when the ball is inactive (Lemma B.1). To stabilize answer selection, the rule retains v while $\eta H_v(\hat{\theta}_t) \leq \min_{u \in \mathcal{V}_\epsilon(\hat{\theta}_t)} H_u(\hat{\theta}_t)$ and switches to the empirically easiest candidate after h consecutive failures; $\eta = 0$ recovers the plain sticky rule.

Proposition 3.5 (Finite-confidence stopping time). *Assume conditionally σ-sub-Gaussian noise, $\|\theta_\star\|_2 \le S$, (2), $H_\epsilon^{\mathrm{hs}}(\theta_\star) < \infty$, the margin $\gamma_\star = \min_{v \neq u} |\epsilon - \langle \theta_\star, v - u\rangle| > 0$, and the separation condition (77). Let β satisfy the anytime coverage (78) with a deterministic envelope $\beta(x,t,g) \le b(x,t) \le b_0 + b_1 \log(1/x) + b_2 \log(1+t)$. The stopping test satisfies $\Pr(\tau < \infty,\ v_\tau \notin \mathcal{V}_\epsilon(\theta_\star)) \le \delta$ under every measurable answer-selection rule. With the practical rule, stopping occurs almost surely and $\mathbb{E}[\tau_\delta] \le T^{\star\star} + 1$, where $T^{\star\star}$ is the least $T \ge 4K^2$ such that every $t \ge T$ satisfies*

$$t \ \ge\ T_1(t^{-2}) + \frac{1+\alpha}{c_{\mathrm{fl}}}\, H_\epsilon^{\mathrm{hs}}(\theta_\star)\Big(\sqrt{b(\delta,t)} + \sqrt{b(t^{-2},t)} + \frac{\sqrt{2\lambda}S}{\sigma}\Big)^2. \tag{8}$$

The constants in the bound are explicit (Section B.4). The factor $1+\alpha$ accounts for answer selection and allocation error; $4K^2$ is the burn-in; $c_{\mathrm{fl}} \ge 1/K$ measures the information retained by forced exploration; and T_1 is a freezing time. The right-hand side grows as $O(\log^2 t)$, so $T^{\star\star} < \infty$.

Corollary 3.6 (The practical rule attains the coefficient). *Assume $\gamma_\star > 0$, a unique easiest ϵ-correct answer $v_\star$, and $\eta H_v(\theta_\star) \neq H_\epsilon^{\mathrm{hs}}(\theta_\star)$ for $v \neq v_\star$. Run the rule with the mixing floor $\Pi_f(q) = (1 - Kf)q + f\mathbf{1}$, allocations within a factor $1+\chi$ of optimal, and a boundary with $b(\delta,t)/\log(1/\delta) \to 1$ along $t = O(\log(1/\delta))$. Then for every $\eta \in (0,1]$,*

$$\limsup_{\delta \downarrow 0} \frac{\mathbb{E}_{\theta_\star}[\tau_\delta]}{\log(1/\delta)} \ \le\ (1+\chi)\,\frac{H_\epsilon^{\mathrm{hs}}(\theta_\star)}{\eta}. \tag{9}$$

The bound in Eq. equation 9 improves to $(1+\chi)H_\epsilon^{\mathrm{hs}}(\theta_\star)$ whenever $H_v(\theta_\star) > H_\epsilon^{\mathrm{hs}}(\theta_\star)/\eta$ for all $v \neq v_\star$, in particular at $\eta = 1$. With exact allocations and an inactive ball, this upper bound matches the lower-bound coefficient $H_\epsilon(\theta_\star)$ in (5).

4 The Price of Identification

4.1 Identifying coordinates costs a factor n

The family $\mathcal{F}_n$ isolates the identification target by making every coordinate observable. For $n \ge 4$, take $d = n$, weights $p_i = 1/n$, local decisions $\mathcal{Z}_i = [0,1]$ with $A_i^\top z = z e_i$, and queries $e_1, \dots, e_n$. Set $\theta_\star = \Delta\mathbf{1}$, $\epsilon = \Delta/4$, and $S \ge \Delta\sqrt{n}$. The aggregates are $v_A = \mathbf{1}_A/n$ for $A \subseteq \{1,\dots,n\}$; v_A is ϵ-correct at $\theta_\star$ exactly when $|A| \le n/4$. Since $\mathcal{Q} = \mathcal{L} = \mathbb{R}^n$ and $\ker B = \{0\}$, the quotient removes no dimension. Any separation between the two tasks therefore comes from what they seek to identify. Define the coordinate-wise characteristic time as

$$H_{\mathrm{loc}}^{(n)} = \frac{2\sigma^2}{\Delta^2} \min_{q \in \Delta_n} \max_{i \le n} \rho_{M(q)}(e_i). \tag{10}$$

Eq. equation 10 measures the optimal cost of resolving every coordinate at gap Δ, whereas $H_\epsilon(\theta_\star)$ measures the cost of returning an ϵ-correct aggregate. The decision tolerates coordinated changes. Flipping one coordinate from Δ to $-\Delta$ changes the exact optimum but increases the regret of $v_\varnothing$ by only $\Delta/n \le \epsilon$.

Theorem 4.1 (Exact factor-n gain). *For every $n \in 4\mathbb{N}$, $\Gamma_\epsilon(\theta_\star) = \Delta^2/2\sigma^2$, $H_\epsilon(\theta_\star) = 2\sigma^2/\Delta^2$, and $H_{\mathrm{loc}}^{(n)} = 2n\sigma^2/\Delta^2$, so that*

$$\frac{H_{\mathrm{loc}}^{(n)}}{H_\epsilon(\theta_\star)} = n. \tag{11}$$

For every $n \ge 4$, $n \le H_{\mathrm{loc}}^{(n)}/H_\epsilon(\theta_\star) \le 4\lceil n/4 \rceil \le n+3$. More generally, at tolerance $\epsilon = \omega\Delta$ with $\omega \in (0,1)$ and $\omega n \in \mathbb{N}$, $H_\epsilon(\theta_\star) = \sigma^2/(2\omega\Delta^2)$ and the ratio is $4\omega n$. The constants in these bounds do not depend on n, Δ, or σ.

The decision and coordinate tasks use the same observations but rule out different alternatives. An ϵ-correct aggregate tolerates coordinated changes; for $n \in 4\mathbb{N}$, the cheapest alternative to the easiest answer must move $n/4$ coordinates at once. Round robin spreads information across these coordinates, leaving Γ_ϵ independent of n. Coordinate recovery, by contrast, must resolve every e_i at gap Δ while allocating only a $1/n$ share of the budget to each. The different identification targets create a factor-n characteristic-time gap that no allocation can remove (Section 4.3).

4.2 WHAT THE BOUNDARY COSTS

Theorem 4.1 measures intrinsic information cost; a stopping rule also pays for its confidence boundary. To isolate that effect on $\mathcal{F}_n$, compare the two tests under the same round-robin allocation, which is optimal for both, and replace plug-in gaps by population gaps. Let r_G and r_L be their first stopping cycles and set $\varrho_{\mathrm{gap}} = r_L/r_G$. After r cycles, write $\nu = r + \lambda$. With a direction-independent radius $R(r)$, the quotient and coordinate-wise tests require, up to integer rounding of the support size, $R(r)^2 \le n\Delta^2\nu$ and $R(r)^2 \le \Delta^2\nu$, respectively. The geometry creates the factor n between these budgets; the boundary determines how much of that factor appears in stopping time.

Theorem 4.2 (What the boundary costs). *On $\mathcal{F}_n$ with $\sigma = \Delta = 1$ and $S = 4\sqrt{n}$, let $\bar{\nu}_G = \bar{r}_G + \lambda$ with $\bar{r}_G$ the continuous crossing time of the quotient test. (a) Under the self-normalized ellipsoid, $\varrho_{\mathrm{gap}} = \Theta(n)$. For every $\lambda \in (0,1)$ and $4 \le n \le \lambda^{-3}/4$,*

$$\varrho_{\mathrm{gap}} \le c(\lambda)\, n\Big(1 + \frac{\log(4n)}{\log(\bar{\nu}_G/\lambda)}\Big), \tag{12}$$

where $c(\lambda) \le 1.017$ at $\lambda = 10^{-6}$ accounts for integer cycles; $\varrho_{\mathrm{gap}} \le 4n$ whenever $\lambda \le 1/4$, and $\varrho_{\mathrm{gap}} \ge n$ whenever $\lambda \le 10^{-4}$ and $192\lambda(\frac{2}{n}\log\frac{1}{\delta} + \log\frac{1}{\lambda}) \le (\log n - 1)^2$. (b) Under the answer-aware boundary at $\lambda = 10^{-6}$ and $\delta \le 1/10$, $\varrho_{\mathrm{gap}} = \Theta(\log n)$ for $6.65(\log(5/\delta) + 6.23) \le n \le 1.83 \times 10^7$.

The confidence boundary determines how much of the geometric factor n appears in stopping time. The ellipsoid bound in Eq. equation 12 preserves a nearly linear gap, whereas a direction-aware boundary removes most of it even though the characteristic-time ratio remains n. The budget comparison extends beyond the symmetric allocation. For any answer set, allocation, and direction-independent boundary, the two tests apply one confidence event to different stopping budgets. On $\mathcal{F}_n$, every query-count vector gives

$$n \;\le\; \frac{\mathcal{B}_G}{\mathcal{B}_L} \;\le\; \Big(\frac{n}{4} + 1\Big)^2. \tag{13}$$

Here $\mathcal{B}_G$ and $\mathcal{B}_L$ are the quotient and coordinate-wise stopping budgets, respectively. Section B.5 proves Eq. equation 13.

4.3 THE DIRECTION UNION IS NECESSARY

Direction-dependent boundaries must control errors across all decision directions. To measure this cost, consider D directions $g_1, \dots, g_D$ with disjoint supports and a parameter under which every competitor of a reference answer has regret $\epsilon + \zeta$, with $\zeta > 0$. Queries follow a balanced deterministic allocation with $\lambda = 0$. Let $\mathcal{T}_{\mathrm{bal}}$ be the ends of complete balanced cycles, where $\|g_d\|_{V_t^{-1}}$ is the same for every d. The *live window* comprises balanced checkpoints after which all cross-direction half-widths remain at most ϵ. Radii are *valid* at level δ if, with probability at least $1 - \delta$, no answer with regret above ϵ ever passes the stopping test, as required for δ-correctness under arbitrary answer selection.

Proposition 4.3 (The direction union is necessary). *For $D \ge 2$ and $0 < \delta \le 1/2$, if per-direction radii $R_t(\delta, g_d)$ fixed before the run are valid at level δ, then at every t in the live window*

$$\max_{d \le D} \frac{R_t(\delta, g_d)}{\sigma} \;\ge\; \Phi^{-1}\Big(1 - \frac{\log(1/(1-\delta))}{D}\Big) - \frac{\zeta}{\sigma\|g_1\|_{V_t^{-1}}}, \tag{14}$$

where the first term is $\sqrt{2\log(D/\delta)}\,(1 + o(1))$ as $D/\delta \to \infty$.

At balanced checkpoints, the D decision directions yield independent Gaussian processes, whose maximum gives the bound in Eq. equation 14 (Section B.6). Allocating failure probability unevenly moves the penalty among directions but cannot remove it. A union over a generating set of $\mathcal{G}$ attains the required order; block-structured boundaries pay for the number of independent blocks. In the symmetric construction, if each direction crosses its boundary with probability δ on the balanced live checkpoints, the lower bound on the probability of accepting a wrong answer tends to $1-(1-\delta)^D > \delta$ as $\zeta \downarrow 0$ for every $D \ge 2$.

What an allocation can add. Adaptive allocation helps only to the extent that uniform sampling differs from the best design for the decision. Let H_{unif} be the half-space characteristic time under uniform allocation and define the *headroom* $\rho_{\mathrm{adapt}} = H_{\mathrm{unif}}/H^{\mathrm{hs}}_{\epsilon}(\theta_\star) \in [1, K]$. Headroom attains both endpoints of $[1, K]$ and equals one exactly when round robin is optimal. It bounds the asymptotic gain available from adaptive allocation over round robin, and training data can provide a pre-acquisition estimate. On $\mathcal{F}_n$, $\rho_{\mathrm{adapt}} = 1$ whenever $\alpha n \in \mathbb{N}$, so allocation cannot remove the factor n in Theorem 4.1. On the rating panels of Section 5, median headroom at the held-out truth is 1.97.

5 Experiments

On the fully observable family $\mathcal{F}_n$, the measured ratio of coordinate-wise to quotient stopping times grows linearly in n under the ellipsoid boundary, as predicted by Theorem 4.2. On real rating panels with a kernel, the quotient test stops on matched runs where the coordinate-wise test never stops. The practical rule approaches the $H_\epsilon \log(1/\delta)$ scale, omitting the direction union can violate the target error level, and headroom tracks the gain from adaptive allocation. Section A gives the full protocols and additional results.

Setting. Synthetic experiments use $\mathcal{F}_n$ and random quotient designs with and without a kernel. The real studies use MovieLens and Sushi rating panels (Harper & Konstan, 2015; Kamishima, 2003), National Basketball Association (NBA) seasons (FiveThirtyEight, 2015), and electricity load curves (Trindade, 2015). On each rating panel, the learner identifies the best of K items for a population of raters, and the design kernel represents a common shift between rater and item effects. Raters are partitioned into training, query-pool, and held-out sets. The training set fixes the noise estimate σ, parameter radius S, and pre-run headroom estimates. Query-pool ratings are replayed as an acquisition stream (Li et al., 2011), while held-out raters are used only to score returned recommendations.

The practical rule uses $(\eta, h) = (\frac{1}{2}, 1)$ with the L_∞ projection. We compare the ellipsoid, stitched, mixture, and answer-aware boundaries and a distribution-free boundary for bounded ratings under a common stopping test. Allocation baselines are round robin, a static G-optimal design, RAGE, LinGame, and LinGapE (Fiez et al., 2019; Degenne et al., 2020; Xu et al., 2018), all run in the quotient. The coordinate-wise and quotient tests use the same samples and confidence events to isolate the effect of the identification target.

Rate and separation. On $\mathcal{F}_n$, the measured ratio of coordinate-wise to quotient stopping times grows linearly in n under the ellipsoid boundary, with all 36 ratios inside the $[n, 4n]$ band of Theorem 4.2. With round-robin allocation and the ellipsoid boundary held fixed, the coordinate-wise reading of the same confidence event requires $0.531n$–$0.558n$ times as many samples as the quotient reading. Figure 2(b) uses the local gap test, while panel (c) holds the samples and confidence event fixed and changes only the reading. The identification target therefore accounts for the observed separation.

The practical rule approaches the characteristic-time scale as δ decreases. Under the answer-aware boundary, the median stopping time on the random families, normalized by $H_\epsilon \log(1/\delta)$, falls from 4.83–7.47 at $\delta = 10^{-1}$ to 1.65–2.07 at $\delta = 10^{-8}$ (Figure 2(a); means by boundary in Table 1). At $\delta = 10^{-8}$, a paired decomposition attributes factors of 1.07–1.12 to the rule and horizon and 1.55–1.88 to the boundary. The boundary factor reflects the slow convergence of $\sqrt{2\beta}$ toward $\sqrt{2\log(1/\delta)}$ in Proposition 3.5.

On the same instances and noise streams under the answer-aware boundary, the family median of mean normalized stopping time at $\delta = 10^{-8}$ is 1.76 for the configuration covered by Corollary 3.6 and 1.77 for LϵBAI (Jourdan & Degenne, 2022). The practical rule also has the explicit finite-confidence stopping-time bound of Proposition 3.5 (Section A.3).

Figure 2: **Rate and stopping-time separation.** (a) Median stopping time scaled by $H_\epsilon \log(1/\delta)$ across boundaries (200 runs per point). (b) Local gap to quotient stopping-time ratio on $\mathcal{F}_n$ under the ellipsoid boundary, with the theoretical band. (c) Coordinate-wise to quotient readings of the same confidence event under round robin, with ellipsoid and mixture boundaries.

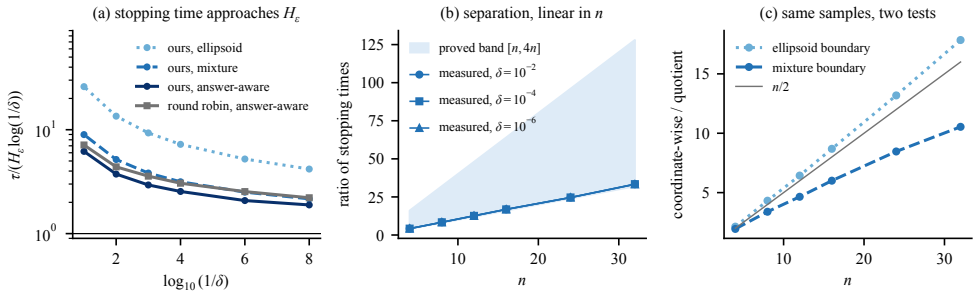


Table 1: **Rate of the practical rule.** Mean $\min(\tau, T)/(H_\epsilon \log(1/\delta))$ on random quotient instances by boundary ($T = 8{,}000$).

boundary	10^{-1}	10^{-2}	10^{-4}	10^{-6}	10^{-8}
ellipsoid	26.95	14.11	7.65	5.52	4.46
stitched	53.78	34.68	24.64	21.07	19.35
mixture	9.18	5.16	3.23	2.56	2.24
answer-aware	6.35	3.93	2.57	2.14	1.87
stitched runs at the horizon	0	0	0	19	40

Feasibility on real designs. The quotient test stops in 18,993 of 64,800 matched runs on the 18 rating panels, while the coordinate-wise test never stops. These runs cover round robin and the practical rule under the ellipsoid and mixture boundaries, with both tests applied to the same samples and confidence events. Table 2 reports the round-robin results, including zero coordinate-wise stops in 32,400 runs, and Figure 3 shows the separation between the tests on paired round-robin samples. The final column pools recommendations returned across the four quotient boundaries, whereas the 1,745 figure below counts panel–order pairs with at least one recommendation.

A design kernel leaves a positive component of coordinate-wise uncertainty along an unobservable direction. The resulting floor depends on the design and does not shrink as ratings accumulate. When it exceeds ϵ, the coordinate-wise test cannot stop regardless of pool size. On all 18 rating panels, the floor exceeds $2{,}843\,\epsilon$. Let U be an orthonormal basis of $\mathcal{Q}$ and $c = \max_i (1 - (UU^\top)_{ii})$. The coordinate-wise half-width is at least

$$\sigma \sqrt{2 \log(1/\delta)} \min_{g \in \mathcal{G}} \|g\|_1 \sqrt{c/\lambda}. \tag{15}$$

At $\delta = 0.05$, round robin with the answer-aware boundary stops in a majority of arrival orders on 12 of 18 panels at $\epsilon = 0.25$ and 17 of 18 at $\epsilon = 0.5$ (Figure 3, lower panel).

Under the answer-aware boundary, the practical rule stops in at least one run on 12 panels at the middle tolerance and on all 18 panels at the largest tolerance. It uses $1.34\times$ fewer ratings than under the mixture boundary. Among the 1,745 panel–order pairs with a recommendation, none is incorrect on held-out raters (Section A.7).

Union, headroom, and allocation. The error rate without the direction union rises with D and reaches 0.1910 at $D = 240$ and $\delta = 0.1$, above the target level. Boundaries that pay the union over all directions or a generating set remain below the target at every tested D (Section A.10).

On 6 panels chosen before acquisition using only training-set headroom, the practical rule wins 40 of 54 panel–tolerance–confidence combinations against round robin under the answer-aware boundary, ties 11, and loses 3. On the wins, round robin uses 1.09–2.43 times as many ratings as

Figure 3: **Real-panel stopping.** Above, paired stopping statistics under round robin and the mixture boundary; bars mark the coordinate-wise floor. Below, the percentage of arrival orders stopped by the quotient test under round robin at $\delta = 0.05$.

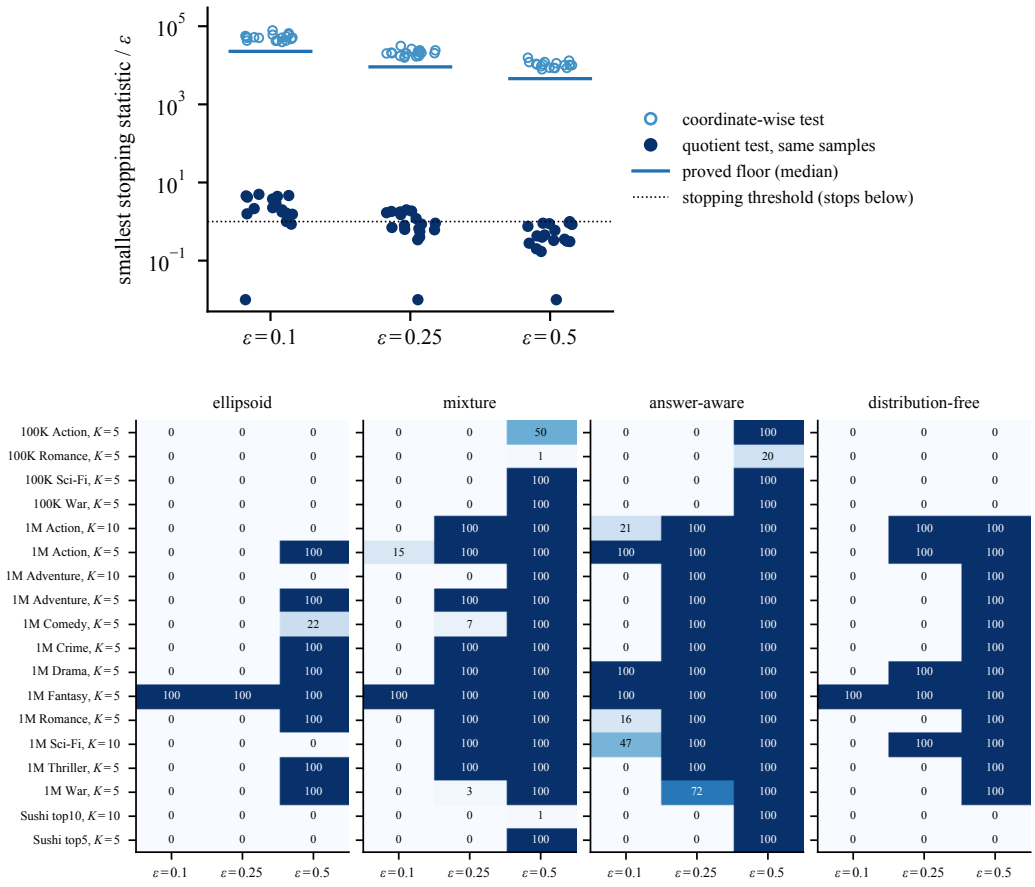


Table 2: **Stopping on real rating panels.**

Matched targets across both allocation rules
Round robin and the practical rule; ellipsoid and mixture boundaries

	coordinate-wise stops	quotient stops
all panels	**0/64,800**	**18,993/64,800**

Round robin by source
Each quotient boundary covers three tolerances, three confidence levels, and 100 arrival orders per panel.

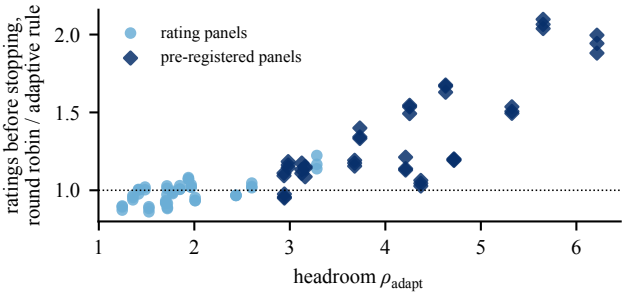
source	panels	coordinate-wise stops (ellipsoid, mixture)	quotient test: runs stopped ellipsoid	mixture	answer-aware	distr.-free	invalid on held-out raters
all panels	**18**	**0/32,400**	**19%**	**48%**	**59%**	**30%**	**0/25,311**
MovieLens 100K	4	0/7,200	0%	18%	27%	0%	0/1,624
MovieLens 1M	12	0/21,600	28%	63%	75%	46%	0/22,784
Sushi	2	0/3,600	0%	17%	33%	0%	0/903

the practical rule (Table 15, Section A.12). Headroom bounds the asymptotic gain available from adaptive allocation over round robin. It equals one on $\mathcal{F}_n$, is at most 2.03 on the random instances, and ranges from 1.25 to 5.42 on the rating panels. On the original rating panels, the paired comparison in Figure 4 gives a Spearman correlation of $+0.87$ between headroom and the practical rule's gain over round robin.

Direction-aware boundaries reduce the distance to stopping on NBA seasons and the query cost on electricity load curves. None of the 36 NBA seasons reaches the stopping threshold, but the closest case finishes 3.44 points above it under the answer-aware boundary, compared with 29.80 under the ellipsoid. On electricity load curves at $\epsilon = 0.05$ and $\delta = 0.05$, the mean query count is 46.10 of the 96 daily queries under the answer-aware boundary, compared with 69.64 under the ellipsoid (Section A.11).

Existing transductive allocation rules operate unchanged in the quotient. On 1,355 of 1,355 paired runs with the same noise streams, implementing the decision-direction test in latent coordinates gives the same stopping time and declared answer as the quotient rule (Section A.8). This comparison keeps the decision test fixed and isolates the cost of computing in different coordinates; the comparison with coordinate recovery measures the sampling cost of recovering individual coordinates before deciding (Table 8). With the quotient dimension fixed at 8, per-round computation stays nearly constant as the latent dimension grows from $d = 8$ to $d = 1{,}032$. At $d = 1{,}032$, the median cost is 0.83 ms per round in quotient coordinates and 1,625.81 ms for the dense latent-coordinate implementation, a difference of three orders of magnitude (Table 16, Section A.13).

Figure 4: **Headroom and realized allocation gain.** Ratio of ratings used by round robin to those used by the practical rule versus headroom on rating panels under the answer-aware boundary.



6 Conclusion

The observable quotient preserves regret and every decision comparison when the query span contains all decision differences. Under Gaussian observations with positive tolerance, it also preserves the information game and the exact leading-order fixed-confidence complexity. The practical rule matches this coefficient in the exact-allocation, inactive-ball case of Corollary 3.6. On $\mathcal{F}_n$ with $n \in 4\mathbb{N}$, coordinate-wise characteristic time is exactly n times that of decision identification, and the self-normalized ellipsoid stopping-time ratio lies between n and $4n$ in the regime of Theorem 4.2. On 18 public rating panels with a kernel, the quotient test stops in 18,993 of 64,800 matched runs, while the coordinate-wise test never stops on the same samples and confidence events. On the independent-direction family, fixed direction-wise confidence radii incur a necessary union penalty, while headroom bounds the asymptotic gain available from adaptive allocation over round robin. The geometry of queries and decision comparisons guides what to measure, how to allocate queries, and when to stop.

AI use statement

Generative AI tools were used for language editing, figure polishing, and limited code assistance. All AI-assisted outputs were reviewed and verified by the authors. The authors take full responsibility for the scientific content, results, and conclusions of this work.

Reproducibility statement

We provide the code, configurations, and scripts needed to reproduce the experiments and reported results. All datasets used in this work are publicly available, and the supplementary material contains the implementation details and experimental settings required for reproduction.

REFERENCES

Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. In *Advances in Neural Information Processing Systems*, volume 24, pp. 2312–2320, 2011.

Emma Brunskill, Ishani Karmarkar, and Zhaoqi Li. Active learning for stochastic contextual linear bandits. In *Proceedings of the 29th International Conference on Artificial Intelligence and Statistics*, volume 300 of *Proceedings of Machine Learning Research*, pp. 577–585. PMLR, 2026. URL https://proceedings.mlr.press/v300/brunskill26a.html.

Sayak Ray Chowdhury and Aditya Gopalan. On kernelized multi-armed bandits. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 844–853. PMLR, 2017. URL https://proceedings.mlr.press/v70/chowdhury17a.html.

Ben Chugg and Aaditya Ramdas. A variational approach to dimension-free self-normalized concentration, 2025. URL https://arxiv.org/abs/2508.06483.

Rémy Degenne and Wouter M. Koolen. Pure exploration with multiple correct answers. In *Advances in Neural Information Processing Systems*, volume 32, pp. 14564–14573, 2019.

Rémy Degenne, Pierre Ménard, Xuedong Shang, and Michal Valko. Gamification of pure exploration for linear bandits. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 2432–2442. PMLR, 2020.

Nicolas Emmenegger, Mojmír Mutný, and Andreas Krause. Likelihood ratio confidence sets for sequential decision making. In *Advances in Neural Information Processing Systems*, volume 36, pp. 26686–26698, 2023. URL https://arxiv.org/abs/2311.04402.

Tanner Fiez, Lalit Jain, Kevin G. Jamieson, and Lillian Ratliff. Sequential experimental design for transductive linear bandits. In *Advances in Neural Information Processing Systems*, volume 32, pp. 10666–10676, 2019. URL https://arxiv.org/abs/1906.08399.

FiveThirtyEight. The complete history of the NBA: nba-elo data set. FiveThirtyEight data repository, nba-elo/nbaallelo.csv, first committed 2015-06-10 ("updated through 6/9/2015"), licence CC BY 4.0; snapshot retrieved 2026-09-02, 2015. URL https://github.com/fivethirtyeight/data/tree/master/nba-elo.

Hamish Flynn, David Reeb, Melih Kandemir, and Jan Peters. Improved algorithms for stochastic linear bandits using tail bounds for martingale mixtures. In *Advances in Neural Information Processing Systems*, volume 36, pp. 45102–45136, 2023. URL https://arxiv.org/abs/2309.14298.

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. URL https://proceedings.mlr.press/v49/garivier16a.html.

F. Maxwell Harper and Joseph A. Konstan. The MovieLens datasets: History and context. *ACM Transactions on Interactive Intelligent Systems*, 5(4):19:1–19:19, 2015. doi: 10.1145/2827872.

Steven R. Howard, Aaditya Ramdas, Jon McAuliffe, and Jasjeet Sekhon. Time-uniform, nonparametric, nonasymptotic confidence sequences. *The Annals of Statistics*, 49(2):1055–1080, 2021. doi: 10.1214/20-AOS1991.

Jonas Hübotter, Bhavya Sukhija, Lenart Treven, Yarden As, and Andreas Krause. Transductive active learning: Theory and applications. In *Advances in Neural Information Processing Systems*, volume 37, pp. 124686–124755, 2024. URL https://arxiv.org/abs/2402.15898.

Yassir Jedra and Alexandre Proutiere. Optimal best-arm identification in linear bandits. In *Advances in Neural Information Processing Systems*, volume 33, pp. 10007–10017, 2020.

Marc Jourdan and Rémy Degenne. Choosing answers in epsilon-best-answer identification for linear bandits. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 10384–10430. PMLR, 2022. URL https://proceedings.mlr.press/v162/jourdan22a.html.

Toshihiro Kamishima. Nantonac collaborative filtering: Recommendation based on order responses. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 583–588. ACM, 2003. doi: 10.1145/956750.956823.

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. URL https://jmlr.org/papers/v22/18-798.html.

Emilie Kaufmann, Olivier Cappé, and Aurélien Garivier. On the complexity of best-arm identification in multi-armed bandit models. *Journal of Machine Learning Research*, 17(1):1–42, 2016. URL https://jmlr.org/papers/v17/kaufman16a.html.

Johannes Kirschner, Andreas Krause, Michele Meziu, and Mojmír Mutný. Confidence estimation via sequential likelihood mixing, 2025. URL https://arxiv.org/abs/2502.14689.

Lihong Li, Wei Chu, John Langford, and Xuanhui Wang. Unbiased offline evaluation of contextual-bandit-based news article recommendation algorithms. In *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, pp. 297–306. ACM, 2011. doi: 10.1145/1935826.1935878.

Zhaoqi Li, Lillian Ratliff, Houssam Nassif, Kevin Jamieson, and Lalit Jain. Instance-optimal PAC algorithms for contextual bandits. In *Advances in Neural Information Processing Systems*, volume 35, pp. 37590–37603, 2022. URL https://arxiv.org/abs/2207.02357.

Michael Lindon, Dae Woong Ham, Martin Tingley, and Iavor Bojinov. Anytime-valid inference in linear models with applications to regression-adjusted causal inference. *Journal of the American Statistical Association*, pp. 1–12, 2026. doi: 10.1080/01621459.2026.2692052. URL https://doi.org/10.1080/01621459.2026.2692052. Published online 27 August 2026.

Kenneth Massey. Statistical models applied to the rating of sports teams. Bachelor's honors thesis, Bluefield College, 1997. URL https://masseyratings.com/theory/massey97.pdf.

Mojmír Mutný and Andreas Krause. Experimental design for linear functionals in reproducing kernel Hilbert spaces. In *Advances in Neural Information Processing Systems*, volume 35, pp. 20175–20188, 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/7f2223201858b6ff4cc1832d8856459b-Abstract-Conference.html.

Riccardo Poiani, Martino Bernasconi, and Andrea Celli. Non-asymptotic analysis of (sticky) track-and-stop. In *International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/6c3490b3616832331d4e2224bb6414c2-Abstract-Conference.html.

Marta Soare, Alessandro Lazaric, and Rémi Munos. Best-arm identification in linear bandits. In *Advances in Neural Information Processing Systems*, volume 27, pp. 828–836, 2014. URL https://arxiv.org/abs/1409.6110.

Artur Trindade. ElectricityLoadDiagrams20112014. UCI Machine Learning Repository, DOI 10.24432/C58C86, 2015. URL https://archive.ics.uci.edu/dataset/321/electricityloaddiagrams20112014.

Beichen Wan, Mo Liu, Paul Grigas, and Zuo-Jun Max Shen. Decision-focused sequential experimental design: A directional uncertainty-guided approach, 2026. URL https://arxiv.org/abs/2602.05340.

Ian Waudby-Smith and Aaditya Ramdas. Confidence sequences for sampling without replacement. In *Advances in Neural Information Processing Systems*, volume 33, pp. 20204–20214, 2020. URL https://arxiv.org/abs/2006.04347.

Ian Waudby-Smith and Aaditya Ramdas. Estimating means of bounded random variables by betting. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 86(1):1–27, 2024. doi: 10.1093/jrsssb/qkad009.

Justin Whitehouse, Zhiwei Steven Wu, and Aaditya Ramdas. Time-uniform self-normalized concentration for vector-valued processes. *The Annals of Applied Probability*, 36(3):1972–2013, 2026. doi: 10.1214/25-AAP2246. URL https://doi.org/10.1214/25-AAP2246.

Liyuan Xu, Junya Honda, and Masashi Sugiyama. A fully adaptive algorithm for pure exploration in linear bandits. In *Proceedings of the 21st International Conference on Artificial Intelligence and Statistics*, volume 84 of *Proceedings of Machine Learning Research*, pp. 843–851. PMLR, 2018.

A Experimental Details and Additional Results

A.1 Setup and protocol

Synthetic families. We evaluate the explicit family $\mathcal{F}_n$ from Section 4 at $n \in \{4, 8, 12, 16, 24, 32\}$, together with random quotient instances with $m \in \{4, 8\}$ decision directions and design-kernel dimension 0 or 4, a best-of-K star family, and a cyclic-arcs answer set. Observations use Gaussian noise with known standard deviation σ. The confidence level ranges over $\delta \in \{10^{-1}, \dots, 10^{-8}\}$, and each experimental cell contains at least 200 runs unless stated otherwise.

Rating panels. We construct 18 rating panels from MovieLens 100K, MovieLens 1M, MovieLens 10M (Harper & Konstan, 2015), and the Sushi preference data (Kamishima, 2003). Each panel contains $K \in \{5, 10\}$ items from one genre or category, and the decision is to identify the best item.

Observation model. Let Y_{ri} denote the rating assigned by rater r to item i. We use the additive model

$$Y_{ri} = u_r + \vartheta_i + \xi_{ri}, \tag{16}$$

where u_r is a rater effect, ϑ_i is an item effect, and ξ_{ri} is observation noise. The within-rater contrast standard deviation σ is estimated from the training raters.

With R raters in the acquisition pool, the raw parameter is $(u, \vartheta) \in \mathbb{R}^{R+K}$, and a raw rating corresponds to the query $e_r + e_i$. The resulting design has rank $R + K - 1$, and its kernel is the common rating-scale shift $(u, \vartheta) \mapsto (u - c, \vartheta + c)$.

The acquisition rule operates directly on within-rater contrasts. The first two revealed items i and j of a rater give $y = Y_{ri} - Y_{rj}$ with query $a = e_i - e_j$. After a set R' of ρ items has been revealed, a new item k contributes the scaled Helmert contrast

$$y = \sqrt{\frac{2\rho}{\rho+1}} \left(Y_{rk} - \bar{Y}_{rR'}\right), \qquad a = \sqrt{\frac{2\rho}{\rho+1}} \left(e_k - \bar{e}_{R'}\right), \tag{17}$$

where $\bar{Y}_{rR'}$ and $\bar{e}_{R'}$ denote the corresponding averages over R'. Under the additive model (16), the contrasts (17) remove the rater effect u_r, have variance σ^2, and are mutually independent.

The resulting decision problem has parameter $\theta = -\vartheta \in \mathbb{R}^K$, quotient $\mathcal{Q} = \mathbf{1}^\perp$, kernel $\ker B = \operatorname{span}(\mathbf{1})$, and decision directions $\mathcal{G} = \{e_i - e_j : i \neq j\}$. Hence $\dim(\mathcal{Q}) = K - 1$ and $\mathcal{G} \subseteq \mathcal{Q}$, so the span condition holds. The parameter-ball radius is set from the training means as $S = \max\{1,\, 2\|\vartheta_{\text{train}} - \bar{\vartheta}_{\text{train}}\mathbf{1}\|_2\}$.

Acquisition cost. Each revealed rating has unit cost. The first contrast from a rater requires two revealed ratings, and each subsequent contrast requires one additional rating. Round robin reveals all available ratings of each arriving rater with at least two observations, using the same unit of cost.

Raters are partitioned by a salted hash into a training set, a query pool, and a held-out set. Training data determine quantities fixed before acquisition, including σ, S, and the pre-run headroom estimates. The query pool supplies the sequential observations. Held-out item means are used only for post-hoc evaluation of the returned recommendations. The pre-registered study uses six additional panels with $K = 8$, selected using training-set headroom alone.

Basketball seasons and electricity. We also evaluate 36 National Basketball Association (NBA) seasons (FiveThirtyEight, 2015) using the pairwise-margin design (Massey, 1997), and electricity load curves (Trindade, 2015) using a task-induced quotient over daily schedules. These datasets provide designs with different latent dimensions and decision-direction structures.

Boundaries. Every stopping rule uses a time-uniform confidence event of the form

$$\left|\langle \hat{\theta}_t - \theta_\star, g\rangle\right| \le R_t(\delta, g)\|g\|_{V_t^{-1}} \qquad \forall t \ge 1,\ \forall g \in \mathcal{G}, \tag{18}$$

with radius $R_t(\delta, g) = \sigma\sqrt{2\beta(\delta, t, g)} + \sqrt{\lambda}\, S$. A confidence boundary is any choice of β for which the event (18) has probability at least $1 - \delta$.

Gaussian boundaries. We evaluate the following Gaussian boundaries.

Ellipsoid. The self-normalized ellipsoid boundary (Abbasi-Yadkori et al., 2011) is

$$\beta_{\mathrm{ell}}(\delta, t) = \log \frac{1}{\delta} + \frac{1}{2} \log \det \left(\frac{V_t}{\lambda I} \right). \tag{19}$$

Mixture. The mixture boundary starts from the fixed-prior Laplace supermartingale construction (Abbasi-Yadkori et al., 2011). With scale matrices $H_k = \lambda 2^k D$ and weights $w_k = 1/((k+1)(k+2))$, where D denotes the reference positive-definite scale matrix used by the mixture, the boundary is

$$\beta_{\mathrm{mix}}(\delta, t) = \min_k c_k^2 \left[\log \frac{1}{\delta w_k} + \frac{1}{2} \log \det \left((H_k + W_t) H_k^{-1} \right) \right], \tag{20}$$

where $W_t = V_t - \lambda I$ and c_k is the metric-mismatch factor. The mixture boundary exceeds the ellipsoid (19) by at most $\log 2$, that is, $\beta_{\mathrm{mix}}(\delta, t) \le \beta_{\mathrm{ell}}(\delta, t) + \log 2$.

Stitched and arm-wise boundaries. The stitched boundary applies the time-uniform construction of (Howard et al., 2021) to each quotient coordinate at level $\delta/(2m)$. The arm-wise boundary applies the per-query construction of (Kaufmann & Koolen, 2021) through the corresponding Pythagorean decomposition.

All Gaussian boundaries assume conditionally σ-sub-Gaussian observations and remain valid under adaptive query selection. Existing vector-valued mixture bounds typically contain dimension-dependent terms (Flynn et al., 2023; Emmenegger et al., 2023; Kirschner et al., 2025; Whitehouse et al., 2026; Chowdhury & Gopalan, 2017; Chugg & Ramdas, 2025; Jedra & Proutiere, 2020). The answer-aware construction below instead allocates its confidence budget using decision directions and design blocks.

The answer-aware boundary. All quantities defining the answer-aware boundary are fixed before acquisition. Let $\mathcal{G}_0 \subseteq \mathcal{G}$ contain one representative from each pair $\{\pm g\}$, and let $\pi(g)$ denote a prior over $\mathcal{G}_0$, chosen uniformly in our experiments.

The quotient coordinates are partitioned into blocks according to the connectivity graph induced by the query rows. Two coordinates belong to the same connected component whenever some query row has nonzero entries in both. Let p denote the number of blocks, and let π_b be the uniform prior over blocks.

We use the scale grid $\rho_k = 2^k$ for $-4 \le k \le 14$ with weights $w_k = \frac{2}{3}/((|k|+1)(|k|+2))$, and the anisotropy grid $\kappa_l \in \{1, 64, 512, 4096\}$ with weights $\nu_l \in \{\frac{1}{4}, \frac{1}{8}, \frac{1}{8}, \frac{1}{2}\}$. Every scale satisfies $\rho_k \ge \lambda$, with $\lambda = 10^{-6}$ in the experiments.

For $g \in \mathcal{G}_0$, let $L(g)$ be the span of the blocks touched by g, and let W_t^L denote the restriction of $W_t = V_t - \lambda I$ to $L(g)$. Write $x = V_t^{-1} g$ and $w = \|g\|^2_{V_t^{-1}}$, and split $\|x\|_2^2$ into its component $p_g = w^2/\|g\|_2^2$ along g and the remainder $q_g = \|x\|_2^2 - p_g$. With P_g the orthogonal projector onto $\mathrm{span}(g)$ and P_L the projector onto $L(g)$, the anisotropic prior matrix is $H_{k,l} = \rho_k P_g + \kappa_l \rho_k (P_L - P_g)$.

The direction-based component is

$$\beta_t^{\mathrm{dir}}(\delta, g) = \min_{k,l} c_{g,k,l}^2 \left[\log \frac{1}{\delta\, \pi(g)\, w_k\, \nu_l} + \frac{1}{2} \log \det{}_L \left((H_{k,l} + W_t^L) H_{k,l}^{-1} \right) \right], \tag{21}$$

with metric-mismatch factor $c_{g,k,l}^2 = 1 + (\rho_k p_g + \kappa_l \rho_k q_g - \lambda \|x\|_2^2)/w$.

For the block-based component, let x_b denote the coordinates of x in block b, and let $W_t^{(b)}$ be the restriction of W_t to that block. With $A_b(k) = x_b^\top W_t x_b + \rho_k \|x_b\|_2^2$ and $B_b(k) = \log(1/(\delta\, \pi_b\, w_k)) + \frac{1}{2} \log \det(I + W_t^{(b)}/\rho_k)$, the block-based component is

$$\beta_t^{\mathrm{blk}}(\delta, g) = \min_k \frac{1}{w} \left[\sum_{b : x_b \neq 0} \sqrt{A_b(k) B_b(k)} \right]^2. \tag{22}$$

The implemented answer-aware boundary combines (21), (22), and the mixture boundary (20) as

$$\beta_t^{\mathrm{aa}}(\delta, g) = \min\left\{\beta_t^{\mathrm{dir}}(0.4\delta, g),\ \beta_t^{\mathrm{blk}}(0.4\delta, g),\ \beta_t^{\mathrm{mix}}(0.2\delta)\right\}. \tag{23}$$

To justify the construction, let $S_t = \sum_{s\le t} a_{J_s}\xi_s$ be the accumulated noise vector. The stochastic error in direction g is $\langle V_t^{-1}g, S_t\rangle$ up to the ridge bias. Since $V_t^{-1}g \in L(g)$, Cauchy–Schwarz in the $(H + W_t^L)$ metric yields

$$|\langle x, S_t\rangle| \le \|x\|_{H+W_t^L}\, \|P_L S_t\|_{(H+W_t^L)^{-1}}. \tag{24}$$

The second factor of (24) is controlled uniformly in time by the corresponding mixture supermartingale, and the first factor is evaluated on the realized design. The three components of (23) each define valid Ville events; their confidence levels sum to δ, so the minimum remains valid after observing the data.

The construction also satisfies $\log(1/(0.4\delta)) \le \beta_t^{\mathrm{aa}}(\delta, g) \le \beta_t^{\mathrm{mix}}(0.2\delta)$. In Table 14, the prefix AA denotes answer-aware. AA-union denotes the direction component (21), AA-block denotes the block component (22), and AA-best denotes the combination (23).

Distribution-free boundaries. For bounded observations, we use per-direction betting and predictable plug-in confidence sequences (Waudby-Smith & Ramdas, 2024), together with Hoeffding and empirical-Bernstein confidence sequences (Howard et al., 2021). Their confidence levels are allocated over a generating set of decision directions. The tightest member at the prescribed split defines the distribution-free boundary.

We also evaluate a without-replacement confidence sequence (Waudby-Smith & Ramdas, 2020) for acquisition rules that reveal every recorded rating of an arriving rater. Its estimand is the mean of pairwise contrasts in the finite query pool.

Estimands. The Gaussian boundary targets population contrast means under the conditionally sub-Gaussian rating model. The distribution-free boundary targets the same population contrasts under bounded observations. The without-replacement boundary targets the finite query-pool contrast means; results for this estimand are marked by †.

Held-out raters are used only for external evaluation. Their item means provide an independent empirical estimate of population-level recommendation quality. Table 7 reports both confidence-event coverage and held-out recommendation performance.

Validity checks. Each reported boundary is evaluated through four coverage checks of the event (18). The checks comprise nominal-level Gaussian replay, replay with a target answer whose regret exceeds the tolerance by a fixed margin, replay using resampled real rating noise, and the simultaneous-binding-direction family of Proposition 4.3. The full cell-level results are included in the supplement.

The practical rule. Algorithm 1 states the practical rule. Its parameters are the hysteresis band η, the switching delay h, the confidence boundary β, and the ridge parameter λ.

Here $N_j(t)$ is the number of times query j has been selected by time t, and Π_f is the forced-exploration floor at level f.

Solver. For a fixed candidate v, the allocation problem in the check step of Algorithm 1 is

$$\min_{q\in\Delta_K}\ \max_{u\neq v}\ y_u^\top M(q)^\dagger y_u, \qquad y_u = \frac{v-u}{\gamma_u(v;\hat\theta_t)_+}. \tag{25}$$

We solve the convex program (25) by Frank–Wolfe with exact line search until the relative gap is below 10^{-7}, followed by an active-set Newton refinement until the KKT residual is below 10^{-12}. The final solution is verified against every alternative, and each update is warm-started from the preceding solution.

If the computed allocation is within a factor $1+\chi$ of optimal in (25), the finite-confidence factor $1+\alpha$ in Proposition 3.5 becomes $(1+\alpha)(1+\chi)$. The experiments use $\eta = 1/2$, the L_∞ projection, and a solver tolerance satisfying $\chi \le 10^{-7}$. Under the conditions of Corollary 3.6, the asymptotic

Algorithm 1 Sticky rule with hysteresis

Require: band η, switching delay h, boundary β, ridge λ, tolerance ϵ, level δ

Query each a_j once and work in an orthonormal basis of $\mathcal{Q}$ ▷ burn-in
$c \leftarrow 0$
for $t = K, K+1, \dots$ **do**
  Update V_t^{-1} and $\hat{\theta}_t$ by a rank-one update ▷ estimate
  if $t \le 200$ or at least $\lceil \sqrt{t} \rceil$ rounds have passed since the last check **then** ▷ check
    Form $\mathcal{V}_\epsilon(\hat{\theta}_t)$ and solve (7) for each candidate v
    if this is the first check **then**
      $v_t \leftarrow \arg\min_{v \in \mathcal{V}_\epsilon(\hat{\theta}_t)} H_v(\hat{\theta}_t)$
    else if $\eta H_{v_{t-1}}(\hat{\theta}_t) > \min_{v \in \mathcal{V}_\epsilon(\hat{\theta}_t)} H_v(\hat{\theta}_t)$ **then**
      $c \leftarrow c + 1$
      if $c \ge h$ **then**
        $v_t \leftarrow \arg\min_{v \in \mathcal{V}_\epsilon(\hat{\theta}_t)} H_v(\hat{\theta}_t)$ and $c \leftarrow 0$
      else
        $v_t \leftarrow v_{t-1}$
      end if
    else
      $v_t \leftarrow v_{t-1}$ and $c \leftarrow 0$
    end if
    $\hat{q}_t \leftarrow$ optimal allocation of (7) for v_t
  else
    $v_t \leftarrow v_{t-1}$ and $\hat{q}_t \leftarrow \hat{q}_{t-1}$
  end if
  $\tilde{q}_t \leftarrow \Pi_{f_t}(\hat{q}_t)$ with $f_t = 1/\big(2\sqrt{K^2 + t}\big)$ ▷ sample
  $J_{t+1} \in \arg\max_j \big[\sum_{s \le t} \tilde{q}_s(j) - N_j(t)\big]$ ▷ C-tracking
  $U_t(u) \leftarrow \langle \hat{\theta}_t, v_t - u \rangle + \big(\sigma\sqrt{2\beta(\delta, t, v_t - u)} + \sqrt{\lambda} S\big)\|v_t - u\|_{V_t^{-1}}$ for $u \ne v_t$
  if $\max_{u \ne v_t} U_t(u) \le \epsilon$ **then** ▷ stop
    return v_t
  end if
end for

coefficient is $(1+\chi)H_\epsilon^{\mathrm{hs}}$ on the 47 instances satisfying the separation condition at $\eta = 1/2$, and the bound is $2(1+\chi)H_\epsilon^{\mathrm{hs}}$ on the remaining instances. Table 10 compares $\eta = 1$ and $\eta = 1/2$ on the real panels.

Computational cost. All computations are performed in the quotient space of dimension $m = \dim(\mathcal{Q}) \le \min\{K, d\}$. A rank-one estimator update costs $O(m^2)$, and the stopping test costs $O(|\mathcal{V}|m)$ for an explicit answer set. Table 16 measures these costs while increasing the kernel dimension and keeping m fixed.

For product policy classes, the empirical best answer is obtained by context-wise linear minimization. The implementation uses the candidate set C_t formed by the empirical best answer and its ϵ-correct single-context deviations. With C the analogous candidate set evaluated at $\theta_\star$, the corresponding characteristic time is $H_{\epsilon,C}^{\mathrm{hs}} = \min_{v \in C \cap \mathcal{V}_\epsilon(\theta_\star)} H_v(\theta_\star) \ge H_\epsilon^{\mathrm{hs}}$. Proposition 3.5 and Corollary 3.6 apply with $H_{\epsilon,C}^{\mathrm{hs}}$ and the easiest member of C.

The stopping maximization over competitors decomposes across contexts when the design is block diagonal. We solve it exactly through the Pareto frontier of per-context gap–variance pairs. The contextual experiments contain policy classes ranging from 625 to 390,625 policies; whenever full enumeration is feasible, the Pareto-frontier computation agrees with it to numerical precision.

A.2 The rate on every family

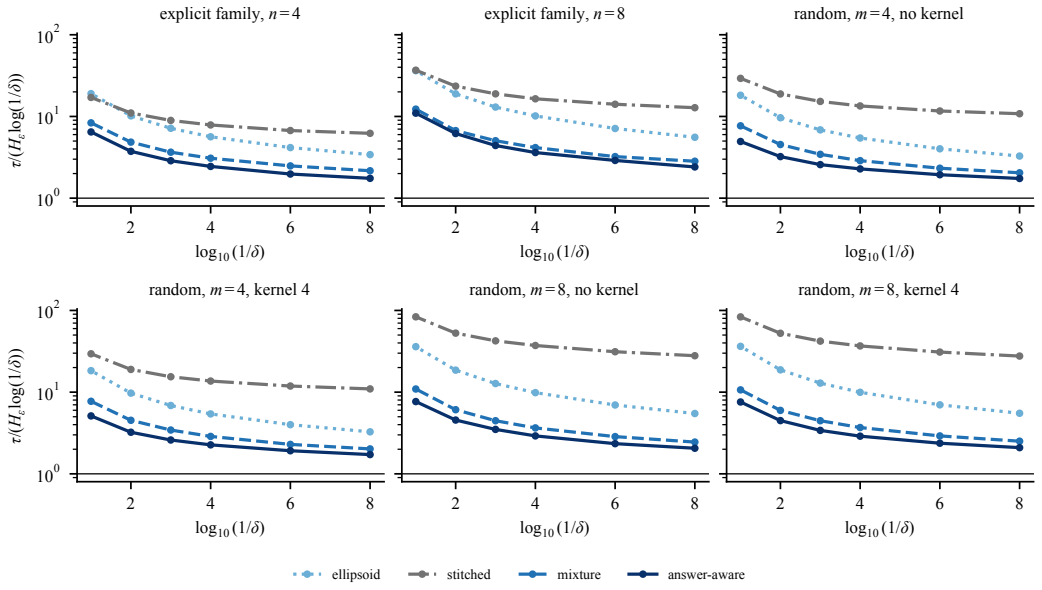
Theorem 3.4 identifies $H_\epsilon(\theta_\star)$ as the leading coefficient of the stopping time. Proposition 3.5 bounds the excess at finite δ by the boundary factor $b(\delta, t)/\log(1/\delta)$, which converges to one logarithmically,

multiplied by the forced-exploration and answer-selection slack. Figure 5 evaluates both statements on six families under four boundaries.

Each curve decreases monotonically in $\log_{10}(1/\delta)$. Across families and levels, the boundaries rank from fastest to slowest as answer-aware, mixture, ellipsoid, and stitched. The boundary radii follow the same order along the observed paths. Under the answer-aware boundary, the median ratio on the random families decreases from 4.83–7.47 at $\delta = 10^{-1}$ to 1.65–2.07 at $\delta = 10^{-8}$; under the mixture boundary it decreases to 1.95–2.49. Run by run, the remaining factor above one decomposes into a rule-and-horizon component of 1.07–1.12 at $\delta = 10^{-8}$ and a boundary component of 1.55–1.88. The sticky selection, the tracking, and the forced-exploration floor therefore cost about ten percent together, and the remaining distance to the constant is the slow convergence of $\sqrt{2\beta}$ to $\sqrt{2\log(1/\delta)}$ accounted for by Proposition 3.5.

Two further observations follow from the theory. First, the design kernel leaves the stopping time unchanged. The random families with a four-dimensional kernel (bottom row) reproduce the curves of the corresponding families without a kernel (top row) under every boundary. The kernel invariance agrees with Theorem 3.3, since every quantity computed by the rule lies in the quotient. Second, the dimension dependence of each boundary is visible. Increasing the number of decision directions from $m = 4$ to $m = 8$, or the family size from $n = 4$ to $n = 8$, raises the ellipsoid and stitched curves, whose β contains the term $\frac{m}{2}\log t$. The answer-aware curve remains nearly unchanged, since its radius on a decision direction pays a union over directions. The same ordering holds for the means in Table 1. All runs that reach the horizon use the stitched boundary, whose per-coordinate union at level $\delta/(2m)$ is the loosest of the four; no run under the other three Gaussian boundaries is censored.

Figure 5: **Normalized stopping time across six synthetic families.** Mean $\min(\tau, T)/(H_\epsilon \log(1/\delta))$ versus $\log_{10}(1/\delta)$ under four boundaries. The families comprise $\mathcal{F}_n$ at two sizes and random quotient designs with $m \in \{4, 8\}$ decision directions, with or without a four-dimensional kernel.



A.3 The configuration covered by the theory, and LεBAI

Corollary 3.6 gives the leading constant one for $\eta = 1$ with the mixing floor $\Pi_f(q) = (1 - Kf)q + f\mathbf{1}$ when the easiest answer is unique, and $H^{\mathrm{hs}}_\epsilon/\eta$ for a band $\eta < 1$ unless the separation condition holds. The experiments of Section 5 use $\eta = \frac{1}{2}$ with the L_∞ projection. Table 3 and Figure 6 evaluate the four combinations of band and floor on the 80 random instances of Figure 5. All configurations use the noise streams of the recorded runs, so every comparison is paired. Re-running the configuration of Section 5 reproduces the recorded per-instance mean stopping times on 1,419 of 1,440 matched instance–boundary–level cells, and the remaining cells differ by at most 0.4 samples in the mean.

The guaranteed configuration stops earliest. At $\delta = 10^{-8}$, in units of $H_\epsilon \log(1/\delta)$, it stops at 1.76 versus 1.90 under the answer-aware boundary, and at 2.11 versus 2.25 under the mixture boundary. The band and the floor each change the stopping time by a few percent, in the direction predicted by the corollary. The band $\eta = 1$ re-selects the easiest candidate at every check, and the mixing floor retains a larger share of the optimal allocation than the projection. The declared answers are the easiest ones. Across all boundaries and levels, every answer declared by either configuration satisfies $H_v(\theta_\star)/H^{\mathrm{hs}}_\epsilon(\theta_\star) \leq 1.093$. Hence the factor two allowed by the corollary at $\eta = \frac{1}{2}$ on the random instances, whose ϵ-correct answers are nearly tied, is never incurred. No run stops on an incorrect answer, and no run reaches the horizon.

LϵBAI (Jourdan & Degenne, 2022) is the closest prior rule. It selects the instantaneous furthest answer, defined as the ϵ-correct answer with the largest answer-conditioned likelihood ratio at the realized design. It plays the allocation of a single AdaHedge learner mixed with forced exploration $1/(tK)$, answers with the closest alternative in the $V(w_t)$ metric, feeds the learner optimistic gains with a δ-free exploration bonus, and samples by tracking. We run it on the quotient with the stopping test and the boundaries of this paper, so the comparison isolates allocation and answer selection. LϵBAI attains the same constant, with 1.77 under the answer-aware boundary and 2.14 under the mixture boundary at $\delta = 10^{-8}$, within a few percent of the practical rule at every level, with no incorrect answer. Run on the quotient, the asymptotically optimal rule of the transductive literature and the practical rule of this paper therefore stop at the same time. The practical rule additionally carries the finite-confidence bound of Proposition 3.5, and Table 3 shows that its constants are not incurred in practice.

Table 9 adds the same three rules to the allocation cells of Table 8 on the recorded noise streams, where the round-robin mean reproduces the recorded value within 2.8%. On the random families, LϵBAI and both configurations of the practical rule stop within a few percent of one another. On the panel-calibrated cells, both configurations of the practical rule stop with 0.42–0.54 of the samples of round robin, and LϵBAI with 0.59–0.66. The same direction test in latent coordinates reproduces the stopping time of the practical rule on all 520 of 520 paired runs.

Figure 6: **Band, exploration floor, and LϵBAI.** Family median of mean normalized stopping time on four random families under (a) the answer-aware and (b) the mixture boundary, using paired instances and noise streams.

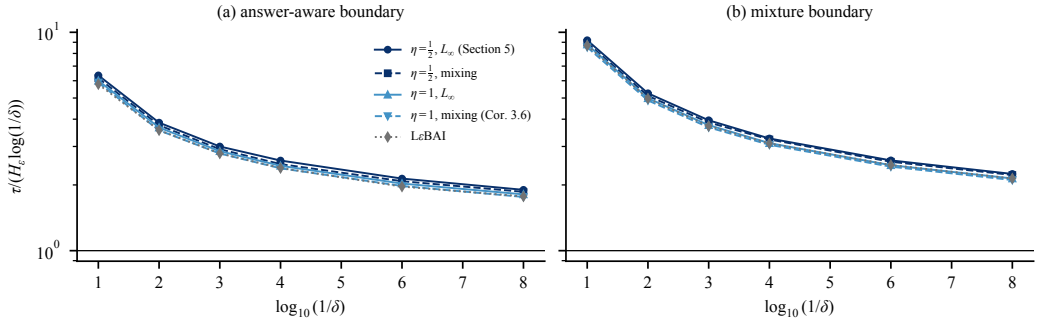


Table 3: **Normalized stopping time and answer difficulty on random families.** Family median of mean $\min(\tau, T)/(H_\epsilon \log(1/\delta))$ (200 runs per cell); the last column gives the largest $H_v(\theta_\star)/H_\epsilon^{\mathrm{hs}}(\theta_\star)$ among returned answers.

configuration	$\delta = 10^{-1}$	10^{-2}	10^{-4}	10^{-6}	10^{-8}	$\max H_{v_\tau}/H_\epsilon^{\mathrm{hs}}$
answer-aware boundary						
$\eta = \frac{1}{2}$, L_∞ projection (Section 5)	6.35	3.86	2.59	2.14	1.90	1.093
$\eta = \frac{1}{2}$, mixing floor	6.16	3.75	2.49	2.08	1.86	1.093
$\eta = 1$, L_∞ projection	6.00	3.66	2.45	2.03	1.81	1.093
$\eta = 1$, mixing floor (Corollary 3.6 with constant one)	5.89	3.55	2.38	1.98	1.76	1.093
LϵBAI (Jourdan & Degenne, 2022)	5.79	3.56	2.39	1.97	1.77	1.093
mixture boundary						
$\eta = \frac{1}{2}$, L_∞ projection (Section 5)	9.20	5.26	3.26	2.59	2.25	1.093
$\eta = \frac{1}{2}$, mixing floor	8.95	5.12	3.23	2.55	2.22	1.093
$\eta = 1$, L_∞ projection	8.79	4.99	3.11	2.46	2.14	1.093
$\eta = 1$, mixing floor (Corollary 3.6 with constant one)	8.54	4.88	3.05	2.42	2.11	1.093
LϵBAI (Jourdan & Degenne, 2022)	8.70	4.97	3.10	2.47	2.14	1.093

A.4 THE SEPARATION UNDER THREE BOUNDARIES

Theorem 4.1 shows that identifying the coordinates of $\mathcal{F}_n$ costs exactly n times the characteristic time of identifying its decision. Theorem 4.2 determines how much of this factor a stopping rule pays, as a function of the boundary. Figure 7 compares the measured ratio of stopping times under each boundary with the corresponding statement. Table 4 gives the ellipsoid local-gap-to-quotient ratios for the six sizes. Its local-box column records the coordinate-wise-to-quotient reading of the same confidence event under round robin.

Ellipsoid. The ratio divided by n lies in $[1.017, 1.135]$ over all 36 cells, so every point lies in the proved band $[n, 4n]$, close to its lower endpoint. The measured ellipsoid ratios agree with part (a) at the ridge $\lambda = 10^{-6}$. The upper endpoint $4n$ is the worst case over the ridge and is attained only when $\log(4n)/\log(\bar{\nu}_G/\lambda)$ is of order one. At this ridge, the granularity constant $c(\lambda)$ is within two percent of one, and the dimension term $\frac{n}{2}\log\bar{\varrho}$ of the ellipsoid pays for the integer cycle almost exactly.

Mixture. The ratio divided by n lies in $[0.731, 1.057]$ and decreases slowly with n. The mixture boundary grows with the sample count and is free of the dimension, so Proposition B.10(i) places the ratio between $nb/(b+n)$ and $n(1+1/b)$, which is linear in n with a slope that falls below one once n is comparable to b.

Answer-aware. The ratio divided by n decreases from 0.970 at $n = 4$ to 0.290 at $n = 32$. Part (b) attributes this behavior to a radius that stays constant as the binding support grows, so the coordinate-wise test pays only the logarithm of the number of directions it covers. Part (b) is proved for $n \geq 6.65(\log(5/\delta) + 6.23)$, which exceeds $n = 32$ at these levels. The figure therefore shows the same decrease below the proved range, and Section A.5 extends the measurement into it.

The three levels $\delta \in \{10^{-2}, 10^{-4}, 10^{-6}\}$ give curves that coincide in every panel. The measured ratio compares plug-in stopping times along random paths, whereas Theorem 4.2 bounds deterministic crossings at the true gaps under round robin; the agreement is an empirical confirmation of the prediction of the theorem. The ratio is determined by the pair (boundary, family), as the theorem states. The factor n of Theorem 4.1 is inherited in full, in part, or only logarithmically according to whether the boundary charges the dimension, the sample count, or the direction. In practice, when every coordinate is observed, the coordinate-wise test under the standard ellipsoid costs n times the quotient test, and a direction-aware boundary removes most of this cost with the test unchanged.

Figure 7: **Stopping-time ratio on $\mathcal{F}_n$ under three boundaries.** Coordinate-wise stopping time divided by quotient stopping time at three confidence levels; the dotted line marks n.

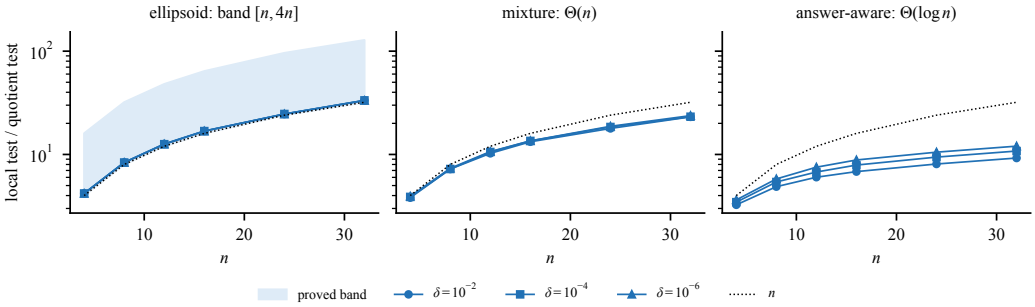


Table 4: **Stopping-time ratios on $\mathcal{F}_n$.** Coordinate-wise stopping time divided by quotient stopping time, with the comparison band of Theorem 4.2.

n	$\delta = 10^{-2}$	10^{-4}	10^{-6}	local box	band
4	4.2	4.2	4.2	2.1	$[4, 16]$
8	8.4	8.4	8.3	4.2	$[8, 32]$
12	12.6	12.6	12.6	6.4	$[12, 48]$
16	16.9	16.8	16.8	8.7	$[16, 64]$
24	24.6	24.7	24.6	13.0	$[24, 96]$
32	33.3	33.4	33.2	17.7	$[32, 128]$

A.5 The separation inside the proved range

Part (b) of Theorem 4.2 is proved for $n \geq 6.65(\log(5/\delta) + 6.23)$, that is, for $n \geq 83$ at $\delta = 10^{-2}$, $n \geq 114$ at $\delta = 10^{-4}$, and $n \geq 145$ at $\delta = 10^{-6}$. The sizes in Figure 7 lie below this range. Table 5 repeats the round-robin comparison at $n \in \{64, 128, 256\}$ under the same boundaries and levels, with 200 runs per cell. No run is censored, and no run stops on an incorrect answer.

Five of the nine answer-aware cells lie inside the proved range, and there the ratio behaves as part (b) states. As n grows fourfold, from 64 to 256, the ratio grows by a factor of 1.46 at $\delta = 10^{-2}$, 1.49 at $\delta = 10^{-4}$, and 1.55 at $\delta = 10^{-6}$. The coordinate-wise test pays a logarithm of the number of directions. The other two boundaries continue the pattern of the smaller sizes. Under the ellipsoid, all 9 cells lie inside the band $[n, 4n]$, with ratio divided by n between 1.036 and 1.050, at the lower endpoint predicted by part (a) at this ridge. Under the mixture, the ratio divided by n decreases from 0.664 to 0.493, the slope below one of Proposition B.10(i).

At $n = 256$, the coordinate-wise test needs 268.6 times the samples of the quotient test under the ellipsoid and 24.0 times under the answer-aware boundary. When every coordinate is observed, the choice of boundary alone is therefore worth a factor of ten, before any allocation is chosen.

Table 5: **Stopping-time ratio on $\mathcal{F}_n$ at $n \in \{64, 128, 256\}$.** Means under round robin (200 runs per cell). Parentheses indicate whether each answer-aware cell lies in the proved range of Theorem 4.2(b); mixture and ellipsoid columns report the ratio divided by n.

	answer-aware: ratio (inside the proved range?)			mixture	ellipsoid	
n	$\delta = 10^{-2}$	10^{-4}	10^{-6}	ratio/n	ratio/n	band
64	11.6 (no)	13.8 (no)	15.5 (no)	0.65–0.66	1.036–1.038	$[64, 256]$
128	15.2 (yes)	17.7 (yes)	19.0 (no)	0.59–0.62	1.042–1.044	$[128, 512]$
256	17.0 (yes)	20.5 (yes)	24.0 (yes)	0.49–0.55	1.049–1.050	$[256, 1024]$

A.6 STOPPING ON REAL RATING PANELS

Table 6 lists the data sources. The key column is the latent dimension. A rating panel with $K \leq 10$ items has a quotient of dimension at most $K - 1$, whereas the latent parameter contains one gauge coordinate per rater, ranging from a few dozen on MovieLens 100K to more than fifty thousand on MovieLens 10M. The span condition holds on every panel, since pairwise contrasts of item means are observed, and the kernel is the rating-scale gauge. The panels are therefore exactly the setting of Section 5. The coordinate-wise test faces the floor (15), and the quotient test is free of it. The NBA seasons and the electricity curves have a different structure. Their latent dimension is the number of teams or the 96 quarter-hours of a day, the decision directions are few, and the pool is one season or one day. They are analyzed in Section A.11.

Figure 3 (lower panel) reports, for each panel, the fraction of arrival orders on which the quotient test stops under round robin at $\delta = 0.05$. The coordinate-wise test on the same confidence events stops on none of them, as the floor requires. Three regularities structure the heat map.

The boundaries nest. On every panel and tolerance, a panel that stops under the ellipsoid also stops under the mixture, and a panel that stops under the mixture also stops under the answer-aware boundary. The nesting is empirical. The proved relations of Section A.1, $\beta_{\mathrm{mix}}(\delta) \leq \beta_{\mathrm{ell}}(\delta) + \log 2$ between (20) and (19) and $\beta^{\mathrm{aa}}(\delta, g) \leq \beta^{\mathrm{mix}}(0.2\,\delta)$ for (23), bound each boundary by the next at a shifted level.

The tolerances nest. The stopping frontier moves outward from the answer-aware column as ϵ grows. The answer-aware boundary stops in a majority of orders on 17 of the 18 panels at $\epsilon = 0.5$, on 12 at $\epsilon = 0.25$, and on the panels with the largest pools at $\epsilon = 0.1$.

Pool size governs the smallest tolerance. At $\epsilon = 0.1$, the half-width on the binding contrast must fall below a tenth of a rating point before any test can stop, and only the largest pools supply enough ratings within the recorded data. A small gap between the two best items does not by itself prevent stopping, since both items are then ϵ-correct.

The lower block of Table 2 pools over panels, tolerances, levels, and orders, with one row per source. The answer-aware boundary stops on the largest share of runs; MovieLens 1M, whose pools contain the most observations, stops most often; and no recommendation issued on any panel is incorrect on the held-out raters. The distribution-free boundary stops only on MovieLens 1M panels. Its per-direction width is built from bounded ratings, and it requires the larger pools to fall below the tolerance.

Table 6: **Data sources and query-pool sizes.**

source	tasks	K	entities	latent dimension	observations in the pool
MovieLens 100K	36	5–10	943	46–185	58–821
MovieLens 1M	36	5–10	6,040	272–1,046	477–4,933
Sushi	2	5–10	5,000	897–1,033	1,526–2,928
MovieLens 10M	2	5–10	69,878	52,407–55,355	156,636–261,658
NBA seasons	36	22–30	—	23–31	725–1,230
Electricity	120	4–16	—	96–96	17,664–17,664

A.7 CALIBRATION

A faster boundary is useful only if its coverage holds. Table 7 reports the error of every boundary against its nominal level on the three types of data used in the paper.

Gaussian replays. The target answer is invalid by a margin ζ beyond the tolerance, under an adversarial and a greedy allocation. No boundary stops on the invalid target in any run at either level, so the observed error is bounded by the Clopper–Pearson limit of an empty count, well below the nominal level.

Recorded rating noise. The residuals of six panels are resampled in three noise arms. The ellipsoid, stitched, and mixture boundaries, which pay a dimension term, miss in no cell. The answer-aware

boundary, the tightest of the four, misses in a small number of trials, and its worst cell remains several times below the nominal level at both $\delta = 0.1$ and $\delta = 0.01$.

Held-out raters. These raters enter no stopping decision. No recommendation issued on any panel–order pair is incorrect, over more than eight hundred pairs per boundary and level, with 95% upper bounds below half a percent.

Across the tested cells, observed error rates and their reported upper bounds remain below nominal levels. The ordering of the boundaries in Sections A.2 and A.6 therefore reflects their widths on the decision directions. The answer-aware boundary is faster because its radius is narrower on the directions that decide the answer.

Table 7: **Calibration of stopping errors.** Gaussian columns count acceptance of an answer invalid beyond the tolerance. Recorded-noise columns use residual resampling on six panels; the worst-cell rate ranges over 18 panel–noise-arm cells. Held-out columns count panel–order pairs with any incorrect recommendation, and the 95% upper bounds are Clopper–Pearson bounds. The stitched boundary was not replayed on panels.

		Gaussian, invalid target beyond the tolerance		recorded rating noise			held-out raters	
boundary	nominal δ	errors	95% upper	misses	worst cell	worst 95% upper	incorrect pairs	95% upper
ellipsoid	0.1	0/8,000	3.7×10^{-4}	0/27,000	0.0000	0.0025	0/837	0.0036
	0.01	0/8,000	3.7×10^{-4}	0/27,000	0.0000	0.0025	0/804	0.0037
stitched	0.1	0/8,000	3.7×10^{-4}	0/27,000	0.0000	0.0025	—	—
	0.01	0/8,000	3.7×10^{-4}	0/27,000	0.0000	0.0025	—	—
mixture	0.1	0/8,000	3.7×10^{-4}	0/27,000	0.0000	0.0025	0/1,599	0.0019
	0.01	0/8,000	3.7×10^{-4}	0/27,000	0.0000	0.0025	0/1,415	0.0021
answer-aware	0.1	0/8,000	3.7×10^{-4}	53/27,000	0.0060	0.0114	0/1,750	0.0017
	0.01	0/8,000	3.7×10^{-4}	6/27,000	0.0013	0.0048	0/1,704	0.0018

A.8 Allocation rules on the quotient

Since the quotient is a property of the problem, any transductive allocation rule can run on it. Figure 8 and Table 8 compare five rules under a common boundary and stopping test, so that only the allocation varies. The results are read against the headroom of Section 4.3, which bounds the gain of any allocation over round robin by $1/\rho_{\mathrm{adapt}}$.

On the explicit family the headroom equals one, and every rule stops with round robin, as Theorem 4.1 requires. The factor n is a property of the geometry, which no allocation changes. On the random families and on the three instances calibrated on rating panels, the adaptive rules stop with a fraction of the samples of round robin that follows the headroom of the instance (Table 12, with median 1.40 on the random families and larger values on the panel-calibrated instances). The spread among LinGame, LinGapE, and the practical rule is small relative to their common distance to round robin. The static G-optimal design remains close to round robin, within fifteen percent on the random families and within one percent on the panel-calibrated instances. The difficulty of the random and panel-calibrated instances depends on the answer, so a design fixed before the data cannot concentrate on the binding directions. RAGE realizes the headroom on the panel-calibrated instances. On the recorded ratings, where the pre-registered panels have the largest headroom, LinGapE and the practical rule again stop with a fraction of the ratings of round robin. The fraction of round-robin ratings used by the adaptive rules increases toward one as the tolerance grows, since the headroom itself decreases with ϵ.

Two latent-space alternatives are also evaluated. The first recovers the latent parameter before deciding (last row of Table 8). It costs two orders of magnitude more samples on the random families, where the design happens to have full rank, and is impossible on the panel-calibrated instances, where the design has a kernel; every rule on the quotient is unaffected by the kernel. The second tests the decision directions directly in latent coordinates, handling the gauge through the ridge on the kernel. On paired runs with the same noise streams, the latent-coordinate direction test matches the stopping time and declared answer of the quotient rule on $1{,}355$ of $1{,}355$ paired runs over 37 cells, kernel designs included, and differs only in computation (Section A.13). The gain of the quotient over this method is the computational gain measured in Section A.13, and its gain over coordinate recovery

is the sample gain measured in Table 8. The factor n of Theorem 4.1 compares two identification targets and is measured in Sections A.4 and A.5.

LinGapE is run in its greedy form, which is (ϵ, δ)-correct (Xu et al., 2018, Thm. 1); its sample-complexity bound (Xu et al., 2018, Thms. 2–3) applies to their allocation (12). LinGapE stops slightly earlier on most rows, and the practical rule stays within tens of percent of it while carrying the guarantees of Proposition 3.5 and Corollary 3.6.

Figure 8: **Allocation rules on the quotient.** Mean samples to stopping relative to round robin on explicit, random, and panel-calibrated instances under the ellipsoid boundary at $\delta = 10^{-2}$.

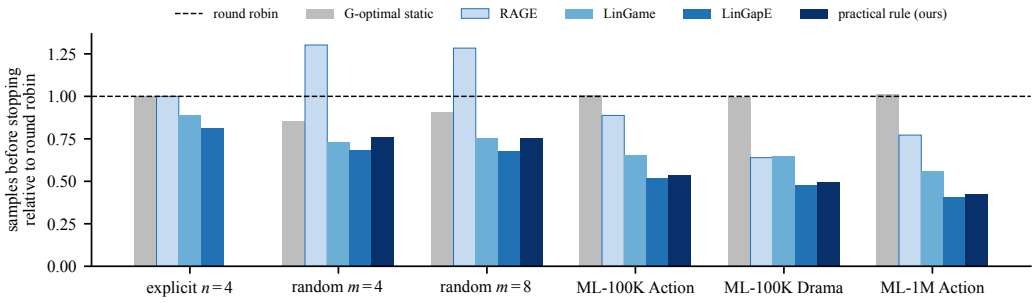


Table 8: **Samples to stopping under allocation rules, relative to round robin.** Gaussian entries are means (200 runs on random families; 37–45 on panel-calibrated instances). Rating entries are medians over pre-registered panels with at least 95% stopping among 100 paired arrival orders at $\delta = 0.05$. A dash means not evaluated; n/i means the latent parameter is unidentifiable.

	Gaussian instances, ellipsoid boundary, $\delta = 10^{-2}$					recorded ratings, answer-aware		
allocation rule	random, $m=4$	random, $m=8$	ML-100K Action	ML-100K Drama	ML-1M Action	$\epsilon = 0.1$	$\epsilon = 0.25$	$\epsilon = 0.5$
round robin	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
G-optimal static design	0.85	0.91	1.01	0.99	1.01	—	—	—
RAGE (Fiez et al., 2019)	1.30	1.28	0.89	0.64	0.77	—	—	—
LinGame (Degenne et al., 2020)	0.73	0.75	0.66	0.65	0.56	—	—	—
LinGapE (Xu et al., 2018)	0.68	0.68	0.52	0.48	0.41	0.49	0.66	0.83
practical rule (ours)	0.76	0.76	0.54	0.49	0.42	0.63	0.76	0.90
full latent recovery	166	152	n/i	n/i	n/i	n/i	n/i	n/i

Table 9: **Additional rules on five Gaussian allocation cells.** Mean samples relative to round robin at $\delta = 10^{-2}$ under the ellipsoid boundary, using the paired noise streams of Table 8.

	random, $m=4$	random, $m=8$	ML-100K Action	ML-100K Drama	ML-1M Action
round robin	1.00	1.00	1.00	1.00	1.00
practical rule, $\eta = \frac{1}{2}$, L_∞ projection	0.76	0.76	0.53	0.50	0.42
practical rule, $\eta = 1$, mixing floor	0.72	0.72	0.54	0.52	0.42
LϵBAI (Jourdan & Degenne, 2022)	0.74	0.76	0.64	0.66	0.59
the same direction test in latent coordinates	0.76	0.76	0.53	0.50	0.42

A.9 Hysteresis, headroom, and the hypotheses

The hysteresis band. The plain sticky rule keeps the first answer it selects. The practical rule keeps its answer only while its characteristic time stays within a factor $1/\eta$ of that of the easiest candidate. Figure 9(a) and Table 10 isolate the effect of the band on the pre-registered panels. Without the band ($\eta = 0$), the rule can settle on an answer that is ϵ-correct but expensive to identify. On such a panel it needs up to $7.9\times$ the ratings of $\eta = \frac{1}{2}$, with a median of $1.66\times$ across panels and tolerances, and it stops on 85.3% of its runs, compared with 89.1%. The theoretical distinction is captured by the factor $|\mathcal{V}_\epsilon| \max_v H_v / H_\epsilon^{\mathrm{hs}}$ from Section 3.3. The sticky-rule guarantee is governed by the hardest ϵ-correct answer, while the band is governed by the easiest. Any band $\eta \geq \frac{1}{4}$ removes the effect, with the worst panel at $1.4\times$. The band $\eta = 1$, which re-selects the easiest candidate at every check, performs like $\eta = \frac{1}{2}$, with the worst panel at $1.0\times$, at the cost of more answer switches per run. The last column of

Table 10 illustrates Lemma B.2. The band affects when the rule stops and leaves the set of answers it may return unchanged, and no recommendation is incorrect on the held-out raters at any η.

Pre-run headroom. Section 4.3 defines the headroom at the true parameter, which is unavailable in practice. Figure 9(b) plots the headroom computed at the training-set item means, available before any rating is acquired, against the headroom at the held-out truth, for the 18 panels and the 6 pre-registered panels at three tolerances. The points follow the diagonal. The pre-registered panels, selected on the training value alone at 3.56–5.45, have held-out headroom 2.94–6.22. The headroom can therefore serve as a screen before acquisition. When it is close to one, round robin is already near-optimal; when it is large, adaptive allocation is worth running. Section A.12 confirms this prediction on panels selected in advance.

Table 12 reports the headroom by family. It equals one on the explicit family by Theorem 4.1, is about one to two on the random families, reaches 5.42 on the rating panels, and is largest on the NBA seasons, where a fixed schedule is far from any optimal allocation over team contrasts. The headroom is thus a property of the design, and it ranks the data sources by the gain available to an active learner.

The hypotheses on the evaluated instances. Proposition 3.5 and Corollary 3.6 require $\gamma_\star > 0$, an inactive ball, a unique easiest ϵ-correct answer, and, for the coefficient at band η, the separation condition $\min_{v \neq v_\star} H_v / H_\epsilon^{\mathrm{hs}} > 1/\eta$. Table 11 checks each hypothesis on the 154 experimental instances, at the held-out truth for the real sources. The ball is inactive on all of them, so the coefficient of the corollary equals the constant H_ϵ of the lower bound. The condition $\gamma_\star > 0$ holds on every instance except the two instances of the explicit family, where $\gamma_\star = 0$ by construction and Theorem 4.2 covers them. The easiest answer is unique on every instance, and no answer lies on the $\eta = \frac{1}{2}$ threshold, so the corollary applies at $\eta = 1$ on all 152 instances with $\gamma_\star > 0$. The separation condition at $\eta = \frac{1}{2}$, which requires the second-easiest answer to be twice as hard as the easiest, holds together with $\gamma_\star > 0$ on 47 instances, mostly rating panels. On the random instances the ϵ-correct answers are nearly tied, with median separation close to one; there the corollary gives $2H_\epsilon^{\mathrm{hs}}$ at $\eta = \frac{1}{2}$ and H_ϵ^{hs} at $\eta = 1$. Table 10 shows that the two bands stop at the same time in practice, so the factor two is slack in the proof.

Figure 9: **Hysteresis and pre-run headroom.** (a) Ratings to stopping relative to $\eta = \frac{1}{2}$ across pre-registered panels and tolerances (answer-aware boundary, $\delta = 0.05$; bar denotes the median). (b) Training-set versus held-out headroom for 18 rating panels and six pre-registered panels at three tolerances.

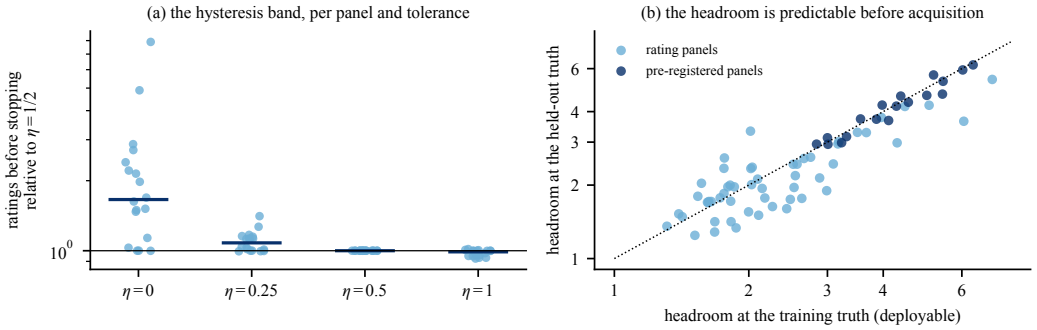


Table 10: **Effect of hysteresis on pre-registered panels.** Answer-aware boundary at $\delta = 0.05$, pooled over three tolerances and 100 paired arrival orders; runs that do not stop count at the pool size.

η	runs stopped	mean ratings before stopping	ratings relative to $\eta = \frac{1}{2}$ median panel	worst panel	answer switches per run	invalid on held-out raters
0	85.3%	4,370	1.66	7.90	0.60	0/1,535
0.25	88.9%	2,591	1.08	1.41	0.89	0/1,600
0.5	89.1%	2,444	1.00	1.00	1.12	0/1,604
1	89.2%	2,386	0.99	1.02	1.89	0/1,605

Table 11: **Conditions for the practical-rule guarantees.** Checks by family on the evaluated instances; real-source values use the held-out truth.

family	instances	ball inactive	$\gamma_\star > 0$	one ϵ-correct answer	unique easiest answer (condition at $\eta = 1$)	at $\eta = \frac{1}{2}$	off the $\eta = \frac{1}{2}$ threshold	median $\min_{v \neq v_\star} H_v / H_\epsilon^{\mathrm{hs}}$
explicit family	2	2	0	0	2	2	2	3.33
random instances	80	80	80	0	80	0	80	1.01
rating panels	54	54	54	19	54	39	54	2.65
pre-registered panels	18	18	18	2	18	8	18	1.70
all	154	154	152	21	154	49	154	1.02

Table 12: **Adaptation headroom ρ_{adapt} by family.** Real-source values use the held-out truth.

family	instances	min	median	max
explicit family	3	1.00	1.00	1.00
random, $m = 4$	20	1.15	1.35	2.03
random, $m = 8$	20	1.07	1.45	1.91
rating panels	18	1.25	1.97	5.42
pre-registered panels	6	2.94	4.23	6.22
contextual tasks	4	1.97	2.75	5.56
NBA seasons	36	3.79	8.77	25.87

A.10 THE UNION TEST

Proposition 4.3 states that, on a family with D binding directions of disjoint support, per-direction radii that skip the union over directions cannot be valid, and that the failure probability grows like the maximum of D independent normals. Table 13 and Figure 10 evaluate three choices of radii on this family, with 2000 trajectories per cell.

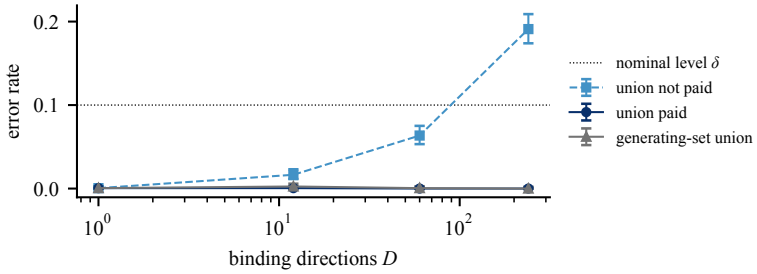
At $D = 1$ there is nothing to unite, and the three boundaries coincide. As D grows, the boundary that skips the union departs monotonically from its level and exceeds it. At $D = 240$ its error rate is 0.1910 at $\delta = 0.1$ and 0.0225 at $\delta = 0.01$. The boundary that pays the union over all of $\mathcal{G}$ errs on at most one trajectory in two thousand at every D and both levels. The boundary that pays the union over a generating set of $\mathcal{G}$, the cheaper union identified in the proof of Proposition 4.3 as attaining the price, also stays below its level and lies between the other two.

The table thus exhibits both bounds of Section 4.3. The union is necessary, since skipping it fails at the predicted rate, and a union over a generating set is sufficient, since it holds the level at a smaller cost than the union over every direction. The gap between the error of the paid union and its nominal level is the usual slack of a Bonferroni-type bound over dependent directions. It is the cost of validity under every answer selection, which Lemma B.2 requires.

Table 13: **Error rate by direction-union boundary.** Family of Proposition 4.3 with D binding directions and 2,000 trajectories per cell.

δ	D	union paid	generating-set union	union not paid
0.1	1	0.0005	0.0005	0.0005
0.1	12	0.0005	0.0025	0.0165
0.1	60	0.0000	0.0005	0.0635
0.1	240	0.0000	0.0000	0.1910
0.01	1	0.0005	0.0005	0.0005
0.01	12	0.0000	0.0000	0.0025
0.01	60	0.0000	0.0000	0.0090
0.01	240	0.0000	0.0000	0.0225

Figure 10: **Direction-union error.** Error rate by number of binding directions at $\delta = 0.1$ for radii that pay the full union, a generating-set union, or no union.



A.11 Designs of another shape, and the price by boundary

Rating panels have few decision directions inside a quotient of dimension $K - 1$. The NBA seasons and the electricity curves have a different structure and show where the choice of boundary matters most.

NBA seasons. Under the pairwise-margin design, the latent dimension of a season equals the number of teams, and the task is to identify the best team within tolerance $\epsilon = 2$ points. Since the schedule of a season is fixed, we measure how close each season comes to stopping. Figure 11(a) plots the closest approach of the stopping statistic to the boundary at the end of each of the 36 seasons. The closest season comes within 29.80 points under the ellipsoid, 9.80 under the mixture, and 3.44 under the answer-aware boundary, with its union member at 3.11. On a design with twenty-odd latent coordinates and a handful of binding contrasts, a boundary that pays the dimension is ten times wider than one that pays the directions. No season stops at this tolerance under any boundary, consistent with the large headroom of the NBA schedules in Table 12. The direction-aware boundaries reduce the distance to a few points, while the fixed schedule accounts for the remaining distance.

Electricity load curves. The curves define a task-induced quotient of daily schedules with 96 quarter-hour queries per day. Figure 11(b) reports the number of queries the practical rule needs before stopping, out of the 96 of a day, at two tolerances and $\delta = 0.05$. At $\epsilon = 0.05$, the answer-aware boundary stops after 46.10 queries, the mixture after 54.97, the coordinate box after 62.84, and the ellipsoid after 69.64. At $\epsilon = 0.01$, where a day that does not stop counts as 96, the ellipsoid leaves 68% of the day-runs without a stop and averages 83.70 queries, compared with 21% and 52.33 for the answer-aware boundary. The ordering matches every other source, and the margin between the answer-aware boundary and the ellipsoid is again the dimension term.

The price of identification by boundary. Table 14 collects the price of identification by boundary on every source. Its DF column denotes the distribution-free boundary, and DF-WoR denotes its without-replacement version. On the rating panels under the practical rule at $\epsilon = 0.25$, the answer-aware variants stop on 8 of the 18 panels in a majority of replays, compared with 4 for the mixture and 1 for the ellipsoid, and across panels the answer-aware boundary needs $1.34\times$ fewer ratings than the mixture. Under round robin at $\epsilon = 0.5$, the mixture and answer-aware boundaries stop on most panels, and the without-replacement boundary stops on all of them for its own estimand. On the contextual product classes, the mixture boundary stops on half of the runs and is the boundary of choice. Across the table, the gain of a direction-aware boundary increases as the set of decision directions becomes narrower relative to the dimension of the quotient, from a tenfold margin on the NBA design to a product class whose directions fill the quotient, where the mixture boundary is preferable.

Figure 11: **Stopping on NBA seasons and electricity load curves.** (a) Final gap to the stopping threshold on NBA seasons by boundary ($\epsilon = 2$ points, $\delta = 0.1$). (b) Queries to stopping out of 96 daily queries on electricity curves ($\delta = 0.05$); nonstopping days count as 96, and labels give their share.

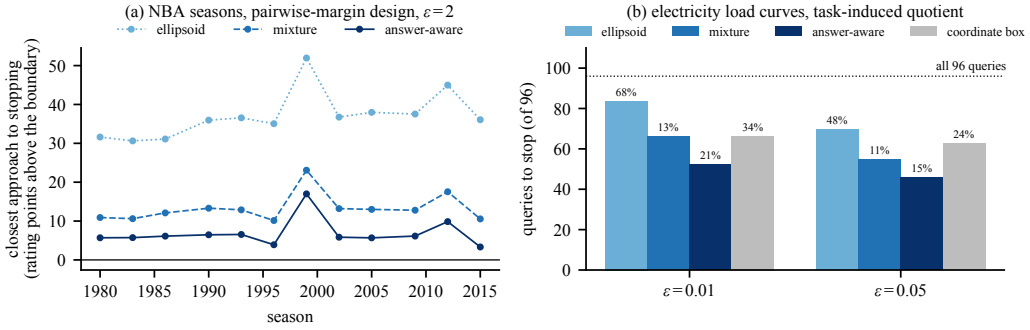


Table 14: **Stopping outcomes by boundary and data source.** Rating rows count panels stopping in a majority of 100 replays; $\delta = 0.05$ except for NBA ($\delta = 0.1$). The dagger marks the without-replacement estimand, which covers pairwise contrasts of the pool.

	ellipsoid	stitched	mixture	AA-union	AA-best	AA-block	DF	DF-WoR
Ratings, our rule, $\epsilon = 0.25$: panels stopping (of 18) ↑	1	1	4	7	**8**	**8**	—	—
Ratings, round robin, $\epsilon = 0.5$: panels stopping (of 18) ↑	8	—	15	—	17	—	12	**18**†
NBA, $\epsilon = 2$: closest approach to stopping (points) ↓	29.80	—	9.80	**3.11**	3.44	—	—	—
Electricity, $\epsilon = 0.05$: queries to stop (of 96) ↓	69.64	—	54.97	**45.81**	46.10	—	—	—
Contextual, $\epsilon = 0.5$: stops (of 400 runs) ↑	0	0	**200**	—	—	0	—	—

A.12 THE PRE-REGISTERED SEARCH

The headroom of Section 4.3 yields a prediction testable before acquisition. On panels selected for large headroom at the training truth, the adaptive rule should need fewer ratings than round robin, by a factor bounded by the headroom. Candidate panels were enumerated from MovieLens 10M and 1M (genre $\times$ $K \in \{5, 8\}$) under the feasibility rules of the released code, giving 74 candidates, of which 44 are feasible. Before any replay, we registered the rule selecting the six feasible panels with the largest training-set headroom, breaking ties by pool size, and the SHA-256 hash of the candidate table. The candidate table contains no held-out quantity. The selected panels have training headroom 3.56–5.45 and held-out headroom 2.94–6.22.

Table 15 reports the paired comparison of the practical rule against round robin over the 54 panel–tolerance–confidence combinations, with 30–100 paired arrival orders each. Under each Gaussian boundary, the adaptive rule wins the large majority of combinations, ties the rest, and loses at most three, with speed-ups on wins of $1.09\times$ to $2.43\times$ under the answer-aware boundary. The observed speed-up lies below the headroom of the same panels. Headroom bounds the asymptotic allocation gain in characteristic time (Section 4.3), while at $\delta = 0.05$ the boundary factor of Section A.2 absorbs part of the gain. The paired ratio follows the headroom across panels with Spearman correlation $+0.75$ under the answer-aware boundary. Across the 18 original panels, whose headroom is smaller, the ratios are correspondingly closer to one (correlation $+0.87$ over the clean combinations, Figure 4). Under each of the four boundaries of Table 15, no recommendation issued in the search is incorrect on the held-out raters.

Under the distribution-free boundary, round robin stops earlier. Its confidence sequences are built from the within-rater contrasts (17), which round robin supplies in full because it reveals every rating of an arriving rater, while the adaptive rule reveals single ratings. The two rules therefore differ in reveal discipline as well as in allocation, and under this boundary the comparison is between complete acquisition strategies. The headroom quantifies Gaussian information and orders allocations under the Gaussian boundaries, where the reveal discipline is the same.

Table 15: **Paired comparison on six pre-registered panels.** Wins, ties, and losses of the practical rule against round robin across 54 panel–tolerance–confidence cells; E counts cells where neither rule stops.

boundary	W/T/L	speed-up on wins	incorrect on held-out raters
answer-aware	40W/11T/3L (E=3)	1.09×–2.43×	0
mixture	42W/12T/0L (E=6)	—	0
ellipsoid	36W/15T/3L (E=15)	—	0
distribution-free	0W/6T/48L (E=6)	—	0

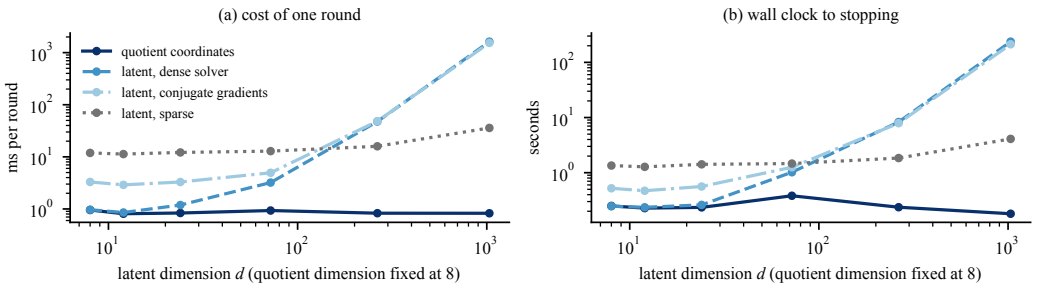
A.13 Computational cost

Theorem 3.3 shows that the quotient loses no information, and Section A.1 shows that it is also cheap to compute in. Figure 12 and Table 16 measure the computational claim on designs whose kernel grows from 0 to more than a thousand dimensions while the quotient stays eight-dimensional.

The rule in quotient coordinates costs under a millisecond per round at every latent dimension from 8 to 1,032. Its update is a rank-one step on an 8×8 matrix, and its stopping test involves $|\mathcal{V}|$ directions of length eight. The same rule in latent coordinates takes the same number of samples, by Theorem 3.3, and solves its update and design game in the latent space. The dense solver grows quadratically and reaches three orders of magnitude more time per round at $d = 1{,}032$; conjugate gradients follow it once the kernel dominates; and a sparse solver stays within an order of magnitude only by exploiting a structure absent from general designs. Panel (b) converts the per-round cost into wall-clock time to stopping. Since the number of rounds is the same in both coordinate systems, the ratio of wall-clock times equals the ratio of per-round costs. The quotient rule stops in about a second, and the latent rule needs minutes.

Table 17 splits a round into phases. In the quotient, the design-game solve for the current answer, which runs at every check, is the dominant phase at every latent dimension; the estimator update, the candidate generation, the boundary, and the stopping test together remain below a tenth of it, and none of these phases depends on d. In latent coordinates the same phases grow with d. The rank-one update and the test on the $d \times d$ inverse grow quadratically, and the design solve, which inverts $d \times d$ matrices at each Newton step, dominates by three orders of magnitude at $d = 1{,}032$. The one-off quotient preprocessing, a singular value decomposition of the query rows, takes a fraction of a millisecond. The quotient is therefore both the object identified by the theory and the natural space for computation, and the latent parameter never needs to be formed.

Figure 12: **Computation in quotient and latent coordinates.** The same rule on designs with a fixed eight-dimensional quotient and growing kernel. (a) Time per round; (b) time to stopping.



B Proofs

B.1 Setting and auxiliary results

We use the notation of Sections 3–4. The latent parameter satisfies

$$\Theta = \{\theta \in \mathbb{R}^d : \|\theta\|_2 \le S\}, \qquad \theta_\star \in \Theta, \tag{26}$$

Table 16: **Per-round computation in quotient and latent coordinates.** Median milliseconds over rounds and repetitions at fixed quotient dimension and growing kernel.

latent dimension d	kernel dimension	quotient (ms/round)	latent (ms/round)	ratio
8	0	0.95	0.97	1.0×
12	4	0.81	0.85	1.1×
24	16	0.84	1.19	1.4×
72	64	0.93	3.21	3.4×
264	256	0.83	47.57	57×
1,032	1,024	0.83	1,625.81	1,968×

Table 17: **Per-round computation by phase.** Median microseconds over rounds and repetitions; preprocessing is measured once.

d	coordinates	update	candidates	design solves	boundary	test	whole round	preprocessing (once)
8	quotient	7.34	3.67	1,066	0.53	14.8	1,143	0.42
8	latent	7.05	3.48	1,063	0.50	14.3	1,146	0.50
12	quotient	7.17	3.56	1,276	0.60	14.7	1,381	0.58
12	latent	8.23	3.74	1,170	0.65	16.5	1,267	0.54
24	quotient	7.79	3.71	811	0.64	16.1	914	0.62
24	latent	8.08	3.56	904	0.55	18.0	1,002	0.62
72	quotient	7.22	3.39	1,786	0.66	15.5	1,869	0.42
72	latent	19.8	4.07	5,049	1.05	56.7	5,270	0.50
264	quotient	7.58	3.63	1,222	0.55	15.2	1,319	0.38
264	latent	151	5.78	36,361	1.84	461	37,967	0.50
1,032	quotient	8.71	3.90	1,165	0.79	17.4	1,253	0.58
1,032	latent	3,625	14.6	1,023,416	3.85	8,070	1,053,006	0.54

for a fixed radius $S < \infty$. Context $i \in \{1, \dots, n\}$ has weight $p_i > 0$, with $\sum_i p_i = 1$, compact local decision polytope $\mathcal{Z}_i$, and known matrix A_i. A policy $\pi = (\pi_i)_i$ selects one vertex $\pi_i \in \mathrm{vert}(\mathcal{Z}_i)$ per context and induces

$$\phi(\pi) = \sum_{i=1}^{n} p_i A_i^\top \pi_i. \tag{27}$$

The finite set $\mathcal{V}$ contains the distinct aggregates $\phi(\pi)$.

The query library is $\{a_j\}_{j=1}^K$, with

$$(B\theta)_j = \langle a_j, \theta \rangle, \qquad \mathcal{Q} = \mathrm{span}\{a_j\}_{j=1}^K, \qquad \mathcal{L} = \mathrm{span}(\mathcal{V} - \mathcal{V}). \tag{28}$$

Throughout this appendix we assume

$$\mathcal{L} \subseteq \mathcal{Q}. \tag{29}$$

At round t, the learner selects a predictable query index J_t and observes

$$Y_t = \langle a_{J_t}, \theta_\star \rangle + \xi_t. \tag{30}$$

The exact characteristic-time results assume independent Gaussian noise

$$\xi_t \sim N(0, \sigma^2), \tag{31}$$

whereas the finite-confidence results require conditional σ-sub-Gaussianity. Query rows may be repeated or zero, and $M(q)$ may be singular.

Oracle allocation and candidate set. The upper bound in Theorem 3.4 applies Sticky Track-and-Stop to the observable Gaussian model

$$\mu = B\theta \in \mathcal{M}_{\mathrm{obs}} \tag{32}$$

with answer correspondence i_ϵ. The oracle rule uses exact access to the easiest-answer set

$$\arg\max_{v \in \mathcal{V}} \mathcal{D}_\epsilon(\mu, v), \tag{33}$$

an optimal allocation for each pair (μ, v), the L_∞ projection onto

$$\{q \in \Delta_K : q_j \geq f_t,\ j = 1, \ldots, K\}, \tag{34}$$

and the test of whether the current confidence region intersects the observable alternative set. Ties are resolved by a fixed order on $\mathcal{V}$. The underlying Sticky Track-and-Stop analysis follows (Degenne & Koolen, 2019).

Zero tolerance. At $\epsilon = 0$, suppose first that two distinct aggregates are optimal at $\theta_\star$ and feasible perturbations exist along their difference. Then

$$H_0(\theta_\star) = +\infty. \tag{35}$$

Indeed, an arbitrarily small perturbation along the decision difference invalidates one of the two answers at an information cost quadratic in its size under every design, giving $\Gamma_0(\theta_\star) = 0$.

Now suppose the optimal aggregate $v_\star$ is unique and has positive gap

$$g_\star = \min_{u \neq v_\star} \langle \theta_\star, u - v_\star \rangle > 0. \tag{36}$$

Choose any allocation q whose support spans $\mathcal{Q}$. Then $\rho_{M(q)}(v_\star - u) < \infty$ for every competitor u, and every alternative θ' satisfies

$$\|\theta' - \theta_\star\|^2_{M(q)} \geq \frac{g_\star^2}{\max_{u \neq v_\star} \rho_{M(q)}(v_\star - u)} > 0. \tag{37}$$

Hence $H_0(\theta_\star) < \infty$.

B.2 Proofs for Section 3.1 and Theorem 3.4

Proposition 3.1. If $B\theta = B\theta'$, then

$$\theta - \theta' \in \ker B = \mathcal{Q}^\perp \subseteq \mathcal{L}^\perp. \tag{38}$$

Therefore,

$$\langle \theta, v - u \rangle = \langle \theta', v - u \rangle \qquad \forall u, v \in \mathcal{V}. \tag{39}$$

Since regret is the maximum of these pairwise differences, $r_\theta(v) = r_{\theta'}(v)$ for every $v \in \mathcal{V}$, and consequently

$$\mathcal{V}_\epsilon(\theta) = \mathcal{V}_\epsilon(\theta'). \tag{40}$$

Proposition 3.2. Let $g = v - u$ attain

$$\max_{v,u \in \mathcal{V}} \|P(v - u)\|_2, \tag{41}$$

and define

$$z = \frac{Pg}{\|Pg\|_2}, \qquad \theta = Sz, \qquad \theta' = -Sz. \tag{42}$$

Then $z \in \ker B$, both parameters belong to Θ, and

$$B\theta = B\theta' = 0. \tag{43}$$

For any $a \in \mathcal{V}$,

$$r_\theta(a) = S\left(\langle z, a \rangle - \min_{b \in \mathcal{V}} \langle z, b \rangle\right), \qquad r_{\theta'}(a) = S\left(\max_{b \in \mathcal{V}} \langle z, b \rangle - \langle z, a \rangle\right). \tag{44}$$

Hence

$$r_\theta(a) + r_{\theta'}(a) = S \max_{b,c \in \mathcal{V}} \langle z, b - c \rangle \geq S\langle z, g \rangle = S\|Pg\|_2 = 2\epsilon_0. \tag{45}$$

For every $\epsilon < \epsilon_0$, no answer can belong to both $\mathcal{V}_\epsilon(\theta)$ and $\mathcal{V}_\epsilon(\theta')$.

The observation law is identical under θ and θ'. If a procedure stops almost surely and is (ϵ, δ)-correct at both parameters, the same probability law assigns mass at least $1 - \delta$ to each of two disjoint answer sets. Therefore

$$2(1 - \delta) \leq 1, \tag{46}$$

which is impossible for $\delta < 1/2$.

Finally,

$$\epsilon_0 = 0 \iff P(v - u) = 0\ \forall u, v \in \mathcal{V} \iff \mathcal{L} \subseteq \mathcal{Q}. \tag{47}$$

Sufficiency follows from Theorem 3.4.

Theorem 3.3. For any $\theta' \in \Theta$,

$$v \notin i_\epsilon(B\theta') \quad \Longleftrightarrow \quad r_{\theta'}(v) > \epsilon. \tag{48}$$

Thus

$$\mathrm{Alt}_\epsilon(v) = \{\theta' \in \Theta : B\theta' \in \neg_\epsilon v\}, \qquad \neg_\epsilon v = B\,\mathrm{Alt}_\epsilon(v). \tag{49}$$

For any allocation $q \in \Delta_K$ and any $\theta, \theta' \in \Theta$,

$$D(q, B\theta, B\theta') = \sum_{j=1}^{K} q_j \frac{\langle a_j, \theta - \theta' \rangle^2}{2\sigma^2} = \frac{\|\theta - \theta'\|^2_{M(q)}}{2\sigma^2}. \tag{50}$$

Taking the corresponding infima gives

$$\inf_{\mu' \in \neg_\epsilon v} D(q, B\theta, \mu') = I_q(\theta, v), \tag{51}$$

and therefore

$$\mathcal{D}_\epsilon(B\theta, v) = \max_{q \in \Delta_K} I_q(\theta, v). \tag{52}$$

The map $D(q, \mu, \mu')$ is continuous in μ', and $\mathcal{M}_{\mathrm{obs}} = B\Theta$ is compact. Replacing $\neg_\epsilon v$ by its relative closure therefore leaves the divergence infimum unchanged. If $v \notin \mathcal{V}_\epsilon(\theta)$, then $B\theta \in \neg_\epsilon v$ and

$$\mathcal{D}_\epsilon(B\theta, v) = 0. \tag{53}$$

Hence

$$\max_{v \in \mathcal{V}} \mathcal{D}_\epsilon(B\theta, v) = \Gamma_\epsilon(\theta). \tag{54}$$

All quantities depend on a policy only through its aggregate, which establishes the invariance statements.

Theorem 3.4: lower bound. A uniformly (ϵ, δ)-correct procedure on Θ induces a δ-correct multiple-answer identification procedure on $(\mathcal{M}_{\mathrm{obs}}, i_\epsilon)$. Proposition 3.1 transfers correctness to the observable model. The standard multiple-answer lower bound (Degenne & Koolen, 2019) yields

$$\liminf_{\delta \downarrow 0} \frac{\mathbb{E}_{\theta_\star}[\tau_\delta]}{\log(1/\delta)} \geq \left(\max_{v \in \mathcal{V}} \mathcal{D}_\epsilon(B\theta_\star, v) \right)^{-1}. \tag{55}$$

Theorem 3.3 identifies the right-hand side with $H_\epsilon(\theta_\star)$.

If $\mathrm{Alt}_\epsilon(v) = \varnothing$ for some answer v, then v is valid on all of Θ and can be returned immediately. Otherwise every correct answer has a nonempty alternative set, and any full-support allocation gives

$$0 < \Gamma_\epsilon(\theta_\star) < \infty. \tag{56}$$

Theorem 3.4: upper bound. We apply Sticky Track-and-Stop on the observable problem $(\mathcal{M}_{\mathrm{obs}}, i_\epsilon)$ (Degenne & Koolen, 2019). The required regularity conditions are as follows.

First, $\mathcal{M}_{\mathrm{obs}}$ and the relative closures of the observable alternative sets are compact. Second,

$$(q, \mu) \mapsto \inf_{\mu' \in \overline{\neg_\epsilon v}} D(q, \mu, \mu') \tag{57}$$

is continuous. Berge's maximum theorem then gives continuity of $\mu \mapsto \mathcal{D}_\epsilon(\mu, v)$ and upper hemicontinuity of the optimal-allocation and easiest-answer correspondences.

Third, forced exploration ensures that each informative query is sampled at least at order $\sqrt{t}$, giving

$$\hat{\mu}_t \to B\theta_\star \qquad \text{almost surely.} \tag{58}$$

Zero query rows contribute no divergence.

Fourth, the all-time Gaussian confidence event with threshold proportional to

$$\log \frac{Ct^2}{\delta} \tag{59}$$

ensures δ-correctness of the strict stopping rule.

These properties imply that the sticky answer stabilizes, the empirical allocation approaches the optimal-allocation set of that answer, and the stopping statistic grows at rate $\mathcal{D}_\epsilon(B\theta_\star)$. The standard tail-sum argument then gives

$$\limsup_{\delta \downarrow 0} \frac{\mathbb{E}_{\theta_\star}[\tau_\delta]}{\log(1/\delta)} \leq H_\epsilon(\theta_\star). \tag{60}$$

B.3 Proof of Theorem 4.1

For $\epsilon = \omega\Delta$, the aggregate answer $v_A = \mathbf{1}_A/n$ has regret

$$r_\theta(v_A) = \frac{1}{n}\left(\sum_{i\in A}\theta_i - \sum_{i=1}^{n}\min\{\theta_i, 0\}\right). \tag{61}$$

At $\theta_\star = \Delta\mathbf{1}$,

$$v_A \in \mathcal{V}_\epsilon(\theta_\star) \quad \Longleftrightarrow \quad |A| \le \omega n. \tag{62}$$

For coordinate-wise identification,

$$\rho_{M(q)}(e_i) = \frac{1}{q_i}, \qquad \min_{q\in\Delta_n}\max_i \frac{1}{q_i} = n, \tag{63}$$

with the minimum attained by the uniform allocation. Hence

$$H_{\mathrm{loc}}^{(n)} = \frac{2n\sigma^2}{\Delta^2}. \tag{64}$$

For the lower bound on Γ_ϵ, take $v_\varnothing$ and the uniform allocation q_{u}. Any strict alternative must make some set B of $s \ge 1$ coordinates sufficiently negative, satisfying

$$\sum_{i\in B}\theta_i' < -n\omega\Delta. \tag{65}$$

With $h = \theta' - \theta_\star$,

$$\sum_{i\in B} h_i < -\Delta(\omega n + s). \tag{66}$$

Cauchy–Schwarz gives

$$\frac{\|h\|^2_{M(q_{\mathrm{u}})}}{2\sigma^2} > \frac{\Delta^2}{2\sigma^2}\frac{(\omega n + s)^2}{ns} \ge \frac{2\omega\Delta^2}{\sigma^2}, \tag{67}$$

with equality in the final inequality at $s = \omega n$.

For the upper bound, fix an ϵ-correct answer v_A and write $a = |A|$. Let

$$k = \lceil \omega n \rceil - a, \tag{68}$$

and choose $J \subseteq A^c$ as the k coordinates with smallest allocation mass. Averaging over k-subsets gives

$$q(J) \le \frac{k}{n-a}. \tag{69}$$

Define θ^J by setting the coordinates in J to $-\Delta$ and all others to Δ. Since $S \ge \Delta\sqrt{n}$, $\theta^J \in \Theta$, and

$$r_{\theta^J}(v_A) = \frac{(a+k)\Delta}{n} \ge \epsilon. \tag{70}$$

Its information cost is

$$\frac{\|\theta^J - \theta_\star\|^2_{M(q)}}{2\sigma^2} = \frac{2\Delta^2 q(J)}{\sigma^2}. \tag{71}$$

If $a + k = \omega n$, an arbitrarily small feasible perturbation moves θ^J into the strict alternative set with the same limiting information cost. If $k = 0$, then $a = \omega n$ and v_A lies exactly on the tolerance boundary at $\theta_\star$; an arbitrarily small feasible perturbation gives

$$I_q(\theta_\star, v_A) = 0. \tag{72}$$

Consequently,

$$I_q(\theta_\star, v_A) \le \frac{2\Delta^2}{\sigma^2}\frac{\lceil\omega n\rceil - a}{n-a} \le \frac{2\Delta^2}{\sigma^2}\frac{\lceil\omega n\rceil}{n}. \tag{73}$$

When $\omega n \in \mathbb{N}$, the lower and upper bounds coincide and give

$$\Gamma_\epsilon(\theta_\star) = \frac{2\omega\Delta^2}{\sigma^2}. \tag{74}$$

Therefore

$$\frac{H_{\mathrm{loc}}^{(n)}}{H_\epsilon(\theta_\star)} = 4\omega n. \tag{75}$$

Setting $\omega = 1/4$ gives the exact factor n when $n \in 4\mathbb{N}$, and the general bound

$$n \le \frac{H_{\mathrm{loc}}^{(n)}}{H_\epsilon(\theta_\star)} \le 4\left\lceil \frac{n}{4} \right\rceil \tag{76}$$

for every $n \ge 4$.

B.4 Proof of Proposition 3.5 and Corollary 3.6

Conditions and constants of Proposition 3.5. Let $g_{\max} = \max_{g\in\mathcal{G}} \|g\|_2$, fix an analysis radius $0 < r < \gamma_\star/(2g_{\max})$, and set $\kappa = r g_{\max}/\gamma_\star$. The separation condition of the proposition is

$$H_v(\theta_\star) > \left(\frac{1+\kappa}{1-\kappa}\right)^2 \frac{H_\epsilon^{\mathrm{hs}}(\theta_\star)}{\eta} \qquad \forall v \in \mathcal{V}_\epsilon(\theta_\star) \setminus \{v_\star\}. \tag{77}$$

For a confidence level $x \in (0,1)$, the boundary $\beta(x,t,g) \ge 0$ induces the radius $R_t(x,g) = \sigma\sqrt{2\beta(x,t,g)} + \sqrt{\lambda}S$ and satisfies the anytime coverage

$$\Pr\left(\forall t \ge 1,\ \forall g \in \mathcal{G} :\ |\langle \hat{\theta}_t - \theta_\star, g\rangle| \le R_t(x,g)\|g\|_{V_t^{-1}}\right) \ \ge\ 1 - x. \tag{78}$$

The constants in (8) are $1 + \alpha = ((1+\kappa)/(1-\kappa))^2$, the information-retention factor c_{fl} of the forced-exploration map, with $c_{\mathrm{fl}} \ge \frac{1}{2}$ for the mixing floor and $c_{\mathrm{fl}} \ge 1/K$ for the L_∞ projection, and the freezing time T_1 of Lemma B.5. For the self-normalized ellipsoid, with $A = \max_j \|a_j\|_2$ and $m = \dim \mathcal{Q}$, the envelope $b(x,t) = \log(1/x) + \frac{m}{2}\log(1 + tA^2/(m\lambda))$ gives $b_1 = 1$ and $b_2 = m/2$; the mixture and answer-aware boundaries admit envelopes of the same logarithmic order. The radius r and the floor enter the analysis and leave the rule unchanged, and their slack vanishes asymptotically. At $\eta = 1$ the separation condition reduces to uniqueness of the easiest answer as $\kappa \downarrow 0$.

Let

$$M_A = \sum_{j=1}^{K} a_j a_j^\top\Big|_{\mathcal{Q}}, \qquad \lambda_A = \lambda_{\min}(M_A), \tag{79}$$

and let $E(x)$ denote the coverage event in (78) at confidence level x. Define

$$\bar{R}_t(x) = \sigma\sqrt{2b(x,t)} + \sqrt{\lambda}\, S. \tag{80}$$

On $E(x)$,

$$R_t(x,g) \le \bar{R}_t(x) \qquad \forall g \in \mathcal{G}. \tag{81}$$

Let

$$\tilde{q}_s = \Pi_{f_s}(\hat{q}_s) \tag{82}$$

be the floored target tracked by C-tracking, where $\hat{q}_s$ is an exact optimal allocation for the current sticky answer.

The practical rule stops at the first time t such that

$$\max_{u \ne v_t}\left[\langle \hat{\theta}_t, v_t - u\rangle + R_t(\delta, v_t - u)\|v_t - u\|_{V_t^{-1}}\right] \le \epsilon. \tag{83}$$

Throughout this appendix, $H_v(\theta,q)$, $H_v(\theta)$, and $H_\epsilon^{\mathrm{hs}}(\theta)$ are defined as in (7), with alternatives taken over $\mathbb{R}^d$. Equivalently,

$$H_v(\theta,q)^{-1} = \min_{u\ne v} \frac{\gamma_u(v;\theta)_+^2}{2\sigma^2\rho_{M(q)}(v-u)}. \tag{84}$$

For a fixed competitor u, this quantity is the squared $M(q)$-distance from θ to the half-space

$$\left\{\theta' \in \mathbb{R}^d : \langle \theta', v - u\rangle \ge \epsilon\right\}, \tag{85}$$

scaled by $1/(2\sigma^2)$. Hence $H_v(\theta,q)^{-1}$ is the distance to the union of the competitor half-spaces.

Lemma B.1 (The ball is inactive)**.** *For every $\theta \in \Theta$,*

$$H_\epsilon(\theta) \le H_\epsilon^{\mathrm{hs}}(\theta). \tag{86}$$

Assume $\gamma_\star > 0$. For each $v \in \mathcal{V}_\epsilon(\theta)$, let $q^v \in \Delta_K$ attain $H_v(\theta)$ with

$$\operatorname{range} M(q^v) = \mathcal{Q}. \tag{87}$$

Call a competitor u active if it attains the maximum in $H_v(\theta, q^v)$, and define

$$\theta_u = \theta + \frac{\gamma_u(v;\theta)}{\rho_{M(q^v)}(v-u)} M(q^v)^\dagger (v-u). \tag{88}$$

Then θ_u is the closest point to θ, in the $M(q^v)$ metric, on the boundary

$$\langle \theta', v-u \rangle = \epsilon. \tag{89}$$

If every such θ_u lies in the interior of Θ, then

$$H_\epsilon(\theta) = H_\epsilon^{\mathrm{hs}}(\theta). \tag{90}$$

Proof. For any $v \in \mathcal{V}_\epsilon(\theta)$, the constrained alternative set satisfies

$$\mathrm{Alt}_\epsilon(v) \subseteq \bigcup_{u \neq v} \left\{ \theta' \in \mathbb{R}^d : \langle \theta', v-u \rangle > \epsilon \right\}. \tag{91}$$

Therefore, for every allocation $q \in \Delta_K$,

$$I_q(\theta, v) \ge H_v(\theta, q)^{-1}. \tag{92}$$

Maximizing over q and minimizing over $v \in \mathcal{V}_\epsilon(\theta)$ gives

$$H_\epsilon(\theta) \le H_\epsilon^{\mathrm{hs}}(\theta). \tag{93}$$

Now fix $v \in \mathcal{V}_\epsilon(\theta)$. Since $\gamma_\star > 0$, every competitor margin satisfies

$$\gamma_u(v;\theta) > 0. \tag{94}$$

Because $\operatorname{range} M(q^v) = \mathcal{Q}$ and $v - u \in \mathcal{Q}$, we also have

$$\rho_{M(q^v)}(v-u) < \infty \tag{95}$$

for every competitor u.

Define

$$F_u(q) = \frac{\gamma_u(v;\theta)^2}{2\sigma^2 \rho_{M(q)}(v-u)}. \tag{96}$$

Then

$$H_v(\theta)^{-1} = \max_{q \in \Delta_K} \min_{u \neq v} F_u(q). \tag{97}$$

Each $F_u(q)$ is concave in q. At the optimal allocation q^v, the envelope theorem gives

$$\frac{\partial F_u(q^v)}{\partial q_j} = \frac{\langle a_j, \theta_u - \theta \rangle^2}{2\sigma^2}. \tag{98}$$

By optimality of q^v for the max–min problem, there exist weights $\nu_u \ge 0$ on the active competitors, with

$$\sum_u \nu_u = 1, \tag{99}$$

such that for every query j,

$$\sum_u \nu_u \frac{\langle a_j, \theta_u - \theta \rangle^2}{2\sigma^2} \le H_v(\theta)^{-1}. \tag{100}$$

Multiplying by q_j and summing over j yields, for every $q \in \Delta_K$,

$$\sum_u \nu_u \frac{\|\theta_u - \theta\|_{M(q)}^2}{2\sigma^2} \le H_v(\theta)^{-1}. \tag{101}$$

Hence at least one active competitor u satisfies

$$\frac{\|\theta_u - \theta\|^2_{M(q)}}{2\sigma^2} \le H_v(\theta)^{-1}. \tag{102}$$

By assumption, each active θ_u lies in the interior of Θ and on the boundary of the corresponding competitor half-space. Therefore θ_u belongs to the closure of $\mathrm{Alt}_\epsilon(v)$. Theorem 3.3 shows that passing to the closure leaves the divergence infimum unchanged, so

$$I_q(\theta, v) \le H_v(\theta)^{-1} \qquad \forall q \in \Delta_K. \tag{103}$$

Combining this with the earlier inequality gives

$$\max_{q \in \Delta_K} I_q(\theta, v) = H_v(\theta)^{-1}. \tag{104}$$

Finally, minimizing over $v \in \mathcal{V}_\epsilon(\theta)$ yields

$$H_\epsilon(\theta) = H^{\mathrm{hs}}_\epsilon(\theta). \tag{105}$$

□

Lemma B.2 (Correctness)**.** *On $E(\delta)$, every answer that passes the stopping test is ϵ-correct.*

Proof. If v passes the test, then for every $u \in \mathcal{V}$,

$$\langle \theta_\star, v - u\rangle \le \langle \hat{\theta}_t, v - u\rangle + R_t(\delta, v-u)\|v-u\|_{V_t^{-1}} \le \epsilon. \tag{106}$$

Hence $r_{\theta_\star}(v) \le \epsilon$. □

Lemma B.3 (Tracking and exploration)**.** *C-tracking satisfies*

$$N_j(t) \ge 1 + \sum_{s \le t} \tilde{q}_s(j) - K. \tag{107}$$

Consequently,

$$V_t \succeq \lambda I + \left(\sqrt{K^2 + t} - 2K\right)_+ M_A. \tag{108}$$

For $T_1 \ge 4K^2$ and $t > T_1$,

$$V_t \succeq \sum_{s=T_1+1}^{t} M(\tilde{q}_s). \tag{109}$$

The flooring map satisfies

$$M(\Pi_f(q)) \succeq c_{\mathrm{fl}} M(q), \qquad c_{\mathrm{fl}} \ge \frac{1}{2} \tag{110}$$

for the mixing floor, and

$$c_{\mathrm{fl}} \ge \frac{1}{K} \tag{111}$$

for the L_∞ projection.

Proof. The C-tracking inequality follows from (Garivier & Kaufmann, 2016). Since

$$f_s = \frac{1}{2\sqrt{K^2 + s}}, \tag{112}$$

we have

$$\sum_{s \le t} f_s \ge \sqrt{K^2 + t} - K. \tag{113}$$

Combining this with the tracking deficit K gives the first matrix bound. The second follows by removing the samples collected before T_1. For the mixing floor,

$$(1 - Kf)q + f\mathbf{1} \ge \frac{1}{2}q \tag{114}$$

whenever $f \le 1/(2K)$. The L_∞ projection gives the stated factor by its coordinate-wise floor. □

Lemma B.4 (Concentration and stability)**.** *Let $T_0(x)$ be the smallest $T \geq 4K^2$ such that, for every $s \geq T$,*

$$4\sigma^2 b(x,s) + 2\lambda S^2 \leq r^2 \left[\lambda + \left(\sqrt{K^2+s} - 2K\right)\lambda_A\right]. \tag{115}$$

Then $T_0(x) < \infty$, and

$$T_0(t^{-2}) = O(\log^2 t). \tag{116}$$

On $E(x)$, for every $t \geq T_0(x)$ and $g \in \mathcal{G}$,

$$|\langle \hat{\theta}_t - \theta_\star, g\rangle| \leq r g_{\max}. \tag{117}$$

If $\gamma_\star > 0$ and $\kappa = r g_{\max}/\gamma_\star < 1$, then

$$\mathcal{V}_\epsilon(\hat{\theta}_t) = \mathcal{V}_\epsilon(\theta_\star), \tag{118}$$

and for every candidate v and allocation q,

$$(1+\kappa)^{-2} H_v(\theta_\star, q) \leq H_v(\hat{\theta}_t, q) \leq (1-\kappa)^{-2} H_v(\theta_\star, q). \tag{119}$$

Thus an allocation optimal at $\hat{\theta}_t$ is $(1+\alpha)$-optimal at $\theta_\star$, with

$$1 + \alpha = \left(\frac{1+\kappa}{1-\kappa}\right)^2. \tag{120}$$

Proof. On $E(x)$,

$$|\langle \hat{\theta}_t - \theta_\star, g\rangle| \leq \frac{\bar{R}_t(x)\|g\|_2}{\sqrt{\lambda_{\min}(V_t)}}. \tag{121}$$

Lemma B.3 gives the required lower bound on $\lambda_{\min}(V_t)$. The decision margins change by at most $r g_{\max}$, which is at most $\kappa|\gamma_u(v;\theta_\star)|$. The candidate-set identity and the multiplicative bounds on H_v follow directly. □

Lemma B.5 (Freezing)**.** *Assume the hypotheses of Lemma B.4, the separation condition (77), and $\eta \in (0,1]$. After time T_0, the unique empirically easiest candidate is $v_\star$. The hysteresis trigger fires whenever the current answer differs from $v_\star$. The rule therefore fixes $v_\star$ by time T_1, where T_1 is the smallest integer satisfying*

$$T_1 \geq T_0 + (h+1)\left(12 + 2\sqrt{T_1}\right). \tag{122}$$

Proof. For every ϵ-correct $v \neq v_\star$, Lemma B.4 and (77) give

$$\eta H_v(\hat{\theta}_t) > H_{v_\star}(\hat{\theta}_t). \tag{123}$$

For $v = v_\star$,

$$\eta H_{v_\star}(\hat{\theta}_t) \leq \min_u H_u(\hat{\theta}_t) \tag{124}$$

because $\eta \leq 1$. Checks occur at most $12 + 2\sqrt{t}$ rounds apart, so h consecutive trigger events imply the stated freezing time. □

Lemma B.6 (Stopping condition)**.** *On $E(x)$, suppose $t > T_1 \geq 4K^2$ and*

$$t - T_1 \geq \frac{1+\alpha}{c_{\mathrm{fl}}} H_\epsilon^{\mathrm{hs}}(\theta_\star)\left(\sqrt{b(\delta,t)} + \sqrt{b(x,t)} + \frac{\sqrt{2\lambda}S}{\sigma}\right)^2. \tag{125}$$

Then $\tau \leq t$.

Proof. Let

$$\bar{q} = \frac{1}{t-T_1}\sum_{s=T_1+1}^{t} \tilde{q}_s. \tag{126}$$

Lemmas B.3 and B.4 imply

$$\|v_\star - u\|^2_{V_t^{-1}} \leq \frac{\rho_{M(\bar{q})}(v_\star - u)}{t - T_1}, \tag{127}$$

and

$$\max_u \frac{\rho_{M(\bar{q})}(v_\star - u)}{\gamma_u(v_\star;\theta_\star)^2} \le \frac{(1+\alpha)H_\epsilon^{\mathrm{hs}}(\theta_\star)}{2\sigma^2 c_{\mathrm{fl}}}. \tag{128}$$

On $E(x)$, the stopping statistic for $v_\star$ is therefore at most ϵ under the displayed condition. □

Assembly. Set

$$x = t^{-2}. \tag{129}$$

The logarithmic envelope gives

$$b(t^{-2}, t) \le b_0 + (2b_1 + b_2)\log(1+t), \qquad T_1(t^{-2}) = O(\log^2 t). \tag{130}$$

Hence the right-hand side of (8) is $O(\log^2 t)$, and $T^{\star\star} < \infty$. For every $t \ge T^{\star\star}$,

$$\Pr(\tau > t) \le \Pr(E(t^{-2})^c) \le t^{-2}. \tag{131}$$

Summing the tail probabilities gives

$$\mathbb{E}[\tau] \le T^{\star\star} + 1. \tag{132}$$

Correctness follows from Lemma B.2.

For the ellipsoid boundary,

$$\det(V_t/\lambda I) \le \left(\frac{\operatorname{tr} V_t}{m\lambda}\right)^m, \qquad \operatorname{tr} V_t \le m\lambda + tA^2. \tag{133}$$

The determinant bound gives the logarithmic envelope used in Section 3.3. The mixture boundary satisfies the same order up to an additive constant, and the answer-aware boundary inherits the same order through its mixture component.

Corollary 3.6. Fix an admissible radius r and an integer $s_0 \ge 4K^2$. Since the freezing property persists after T_1, the stopping argument can begin at

$$T_1' = \max\{T_1, s_0\}. \tag{134}$$

For the mixing floor,

$$M(\Pi_{f_s}(q)) \succeq (1 - Kf_s)M(q) \succeq (1 - Kf_{s_0})M(q) \tag{135}$$

for every $s > T_1'$. Hence the stopping proof may use

$$c_{\mathrm{fl}} = 1 - Kf_{s_0}. \tag{136}$$

Along the fixed point,

$$t = O(\log(1/\delta)), \qquad \bar{\beta}(t^{-2}, t) = O(\log t) = o(\log(1/\delta)), \tag{137}$$

and

$$T_1(t^{-2}) = O(\log^2 t) = o(\log(1/\delta)). \tag{138}$$

Therefore

$$\limsup_{\delta \downarrow 0} \frac{\mathbb{E}[\tau_\delta]}{\log(1/\delta)} \le \frac{1+\alpha(r)}{1 - Kf_{s_0}} H_\epsilon^{\mathrm{hs}}(\theta_\star). \tag{139}$$

Letting $r \downarrow 0$ and $s_0 \to \infty$ gives the coefficient $H_\epsilon^{\mathrm{hs}}(\theta_\star)$ whenever the separation condition holds.

For a hysteresis band $\eta \in (0, 1]$, the threshold condition divides the competing answers according to whether

$$\eta H_v(\theta_\star) > H_\epsilon^{\mathrm{hs}}(\theta_\star) \tag{140}$$

or

$$\eta H_v(\theta_\star) < H_\epsilon^{\mathrm{hs}}(\theta_\star). \tag{141}$$

For sufficiently small r, these inequalities remain unchanged after estimation error is introduced. The rule eventually freezes at an answer satisfying

$$H_v(\theta_\star) \le \frac{H_\epsilon^{\mathrm{hs}}(\theta_\star)}{\eta}. \tag{142}$$

Applying Lemma B.6 to that answer yields

$$\limsup_{\delta \downarrow 0} \frac{\mathbb{E}[\tau_\delta]}{\log(1/\delta)} \le \frac{H_\epsilon^{\mathrm{hs}}(\theta_\star)}{\eta}. \tag{143}$$

An allocation with objective value within a factor $1+\chi$ of optimal multiplies the right-hand side by $1+\chi$.

For the L_∞ projection, let $q_{\min}$ denote the smallest positive coordinate of the optimal allocation in a neighborhood of $\theta_\star$. Then

$$c_{\mathrm{fl}}(q, f) \ge 1 - \frac{(K-1)f}{q_{\min}}. \tag{144}$$

If $q_{\min} > 0$ uniformly on that neighborhood, the flooring factor converges to one on late windows.

Without the separation condition (77), applying the same stopping argument to every candidate gives the bound

$$|\mathcal{V}_\epsilon(\theta_\star)| \max_{v \in \mathcal{V}_\epsilon(\theta_\star)} H_v(\theta_\star). \tag{145}$$

B.5 Proofs for Sections 4.2 and 4.3

Under round robin on $\mathcal{F}_n$, replace the plug-in gaps by their population values. After r cycles,

$$\nu = r + \lambda, \qquad \|e_i\|^2_{V_t^{-1}} = \frac{1}{\nu}. \tag{146}$$

The quotient and coordinate-wise stopping conditions become

$$\Psi_n(r) = \max_{1 \le s \le n} \left[-s + \frac{R(r)}{\Delta} \sqrt{\frac{s}{\nu}} \right] \le \frac{n}{4}, \qquad R(r) \le \Delta\sqrt{\nu}. \tag{147}$$

Lemma B.7 (Budget identity)**.** *For a radius R independent of s,*

$$\max_{s>0} \left[-s + \frac{R}{\Delta} \sqrt{\frac{s}{\nu}} \right] = \frac{R^2}{4\Delta^2 \nu}. \tag{148}$$

Hence the quotient test passes whenever

$$R^2 \le n\Delta^2 \nu. \tag{149}$$

For $n \in 4\mathbb{N}$, this condition is exact at the crossing.

Lemma B.8 (Crossings under the ellipsoid)**.** *Set*

$$\sigma = \Delta = 1, \qquad w = 4\sqrt{\lambda n}, \qquad \beta(r) = \log \frac{1}{\delta} + \frac{n}{2} \log \frac{\nu}{\lambda}, \tag{150}$$

and define

$$\varphi(r) = \left(\sqrt{2\beta(r)} + w \right)^2. \tag{151}$$

For $c \le n$, let $\bar{r}_c$ be the continuous crossing of

$$\varphi(r) \le c\nu. \tag{152}$$

Write

$$\bar{\nu}_G = \bar{r}_G + \lambda, \qquad \bar{\nu}_L = \bar{r}_L + \lambda, \qquad \bar{\varrho} = \frac{\bar{\nu}_L}{\bar{\nu}_G}, \tag{153}$$

and

$$u_G = 4\sqrt{\frac{\lambda}{\bar{\nu}_G}}, \qquad u_L = 4\sqrt{\frac{\lambda n}{\bar{\nu}_L}}. \tag{154}$$

Then

$$\bar{\varrho} \ge n, \tag{155}$$

and

$$\bar{\nu}_L (1-u_L)^2 - n\bar{\nu}_G (1-u_G)^2 = n \log \bar{\varrho}. \tag{156}$$

Define

$$\Lambda = \bar{\nu}_G \left[1 - (1 - u_G)^2\right] = 8\sqrt{\lambda \bar{\nu}_G} - 16\lambda. \tag{157}$$

Furthermore,

$$\bar{\nu}_G \geq \log \frac{\log 16}{\lambda}, \tag{158}$$

and for $\lambda \leq 10^{-4}$*,*

$$\bar{\nu}_G \leq 3M, \qquad M = \frac{2}{n} \log \frac{1}{\delta} + \log \frac{1}{\lambda}. \tag{159}$$

Proof. The function $\varphi - c\nu$ is concave and crosses zero once. Evaluating the quotient crossing at $\nu = n\bar{\nu}_G$ gives the lower bound $\bar{\varrho} \geq n$. At the two crossings,

$$2\beta(\bar{r}_G) = n\bar{\nu}_G(1 - u_G)^2, \qquad 2\beta(\bar{r}_L) = \bar{\nu}_L(1 - u_L)^2. \tag{160}$$

Subtracting the two expressions yields the logarithmic identity. The remaining inequalities follow from the crossing equation and elementary bounds on $\bar{\nu}_G$. □

Proposition B.9 (Both endpoints)**.** *If*

$$\Lambda \leq \log n - 1, \tag{161}$$

then

$$r_L \geq n r_G + \lceil n(\log n - 1 - \Lambda) \rceil, \tag{162}$$

and therefore $\varrho_{\mathrm{gap}} \geq n$*. For* $\lambda \leq 10^{-4}$*, a sufficient condition is*

$$192\lambda \left(\frac{2}{n} \log \frac{1}{\delta} + \log \frac{1}{\lambda} \right) \leq (\log n - 1)^2. \tag{163}$$

For $\lambda \in (0, 1)$ *and* $4 \leq n \leq \lambda^{-3}/4$*, define*

$$L = \log \frac{\bar{\nu}_G}{\lambda}. \tag{164}$$

Then

$$\bar{\varrho} \leq n \left(1 + \frac{\log(4n)}{L} \right) \leq 4n, \tag{165}$$

and

$$\varrho_{\mathrm{gap}} \leq c(\lambda)\bar{\varrho}, \tag{166}$$

where

$$c(\lambda) = 1 + \frac{1}{4(\log(2.77/\lambda) - \lambda)} + \frac{\lambda}{\bar{\nu}_G - \lambda}. \tag{167}$$

Proof. The lower endpoint follows from the crossing identity

$$\bar{\nu}_L \geq n\bar{\nu}_G + n(\log n - \Lambda). \tag{168}$$

Converting continuous crossings to integer cycles gives the stated bound.

For the upper endpoint, divide the crossing identity by

$$n\bar{\nu}_G(1 - u_G)^2 = 2\beta(\bar{r}_G) \geq nL. \tag{169}$$

The crossing identity therefore yields

$$\bar{\varrho} \leq n \left(1 + \frac{\log \bar{\varrho}}{L} \right). \tag{170}$$

The resulting fixed-point map is concave on the relevant range and gives $\bar{\varrho} \leq 4n$, followed by the sharper displayed bound. Integer-cycle granularity contributes the factor $c(\lambda)$. □

Proposition B.10 (Further boundaries on $\mathcal{F}_n$)**.** *In the setting of Theorem 4.2, (i) if* β *does not grow with the cycle count, then with* $b = 2\sigma^2\beta/\Delta^2$*,*

$$\frac{b}{b+n} \leq \frac{\varrho_{\mathrm{gap}}}{n} \leq 1 + \frac{1}{b}; \tag{171}$$

(ii) at $\lambda = 0$*, a fully direction-separable boundary makes the quotient test coincide with the coordinate-wise box test, and the stopping times coincide.*

Proof of Proposition B.10 and part (b) of Theorem 4.2. If β is independent of r, the budget identity gives

$$r_G = \left\lceil \frac{b}{n} - \lambda \right\rceil_+, \qquad r_L = \lceil b - \lambda \rceil, \tag{172}$$

up to the ridge correction. Hence

$$\frac{b}{b+n} \le \frac{\varrho_{\mathrm{gap}}}{n} \le 1 + \frac{1}{b}. \tag{173}$$

At $\lambda = 0$, a direction-separable boundary places a scalar confidence sequence at level δ/n on each coordinate. The quotient stopping condition then decomposes across coordinates and agrees with the coordinate-wise box test.

For the answer-aware boundary,

$$r_L = 2 \log \frac{n}{0.4\delta} + O\Big(\log\log \frac{n}{\delta}\Big), \tag{174}$$

while the quotient cycle count remains bounded at the corresponding order. Therefore

$$\varrho_{\mathrm{gap}} = \Theta(\log n). \tag{175}$$

Transfer. For an arbitrary count vector N, define

$$u_i = (N_i + \lambda)^{-1}, \tag{176}$$

and let

$$u_{(1)} \ge u_{(2)} \ge \cdots \ge u_{(n)}, \qquad U_s = \sum_{j=1}^{s} u_{(j)}. \tag{177}$$

For a direction-independent radius, define the quotient and coordinate-wise stopping budgets

$$\mathcal{B}_G = \Delta^2 \min_s \frac{(n/4 + s)^2}{U_s}, \qquad \mathcal{B}_L = \frac{\Delta^2}{u_{(1)}}. \tag{178}$$

Since

$$U_s \le s u_{(1)}, \qquad \Big(\frac{n}{4} + s\Big)^2 \ge ns, \tag{179}$$

we obtain

$$n \le \frac{\mathcal{B}_G}{\mathcal{B}_L} \le \Big(\frac{n}{4} + 1\Big)^2. \tag{180}$$

Proposition B.11 (Adaptivity headroom)**.** *Assume*

$$0 < H_\epsilon^{\mathrm{hs}}(\theta_\star) < \infty. \tag{181}$$

Then

$$1 \le \rho_{\mathrm{adapt}} \le K. \tag{182}$$

Equality $\rho_{\mathrm{adapt}} = 1$ holds exactly when the uniform allocation q_{u} solves the characteristic-time problem.

Round robin with the stop-on-any rule is (ϵ, δ)-correct, and its stopping bound follows from (8) after replacing

$$\frac{1+\alpha}{c_{\mathrm{fl}}} H_\epsilon^{\mathrm{hs}} \tag{183}$$

with H_{unif} and replacing T_1 with $K - 1$.

On $\mathcal{F}_n$, if $\omega n \in \mathbb{N}$, then

$$\rho_{\mathrm{adapt}} = 1. \tag{184}$$

On the two-gap family with K unit queries, one gap g, and $K - 2$ gaps γg,

$$\rho_{\mathrm{adapt}} \ge \frac{K}{3} \tag{185}$$

whenever

$$2\gamma^2 - 1 \ge K - 2. \tag{186}$$

Proof. For every answer v,

$$H_v(\theta, q_{\mathrm{u}}) \geq \min_{q \in \Delta_K} H_v(\theta, q), \tag{187}$$

which gives the lower bound. Since

$$q_{\mathrm{u}} \geq \frac{q}{K} \tag{188}$$

entrywise for every $q \in \Delta_K$,

$$M(q_{\mathrm{u}}) \succeq \frac{1}{K} M(q), \qquad \rho_{M(q_{\mathrm{u}})}(g) \leq K \rho_{M(q)}(g), \tag{189}$$

which gives the upper bound.

The round-robin stopping statement follows from the proof of Proposition 3.5 with a fixed allocation. On $\mathcal{F}_n$, the uniform design attains the lower bound established in Theorem 4.1. The two-gap bound follows by optimizing

$$I_q(\theta_\star, e_1) = \min_{j \geq 2} \frac{\tilde{\Delta}_j^2}{2\sigma^2(1/q_1 + 1/q_j)} \tag{190}$$

over q and comparing the optimum with q_{u}. □

B.6 Proof of Proposition 4.3 and the information floor

We prove two results used in Section 5. The first establishes a lower bound on the confidence radius required across multiple independent decision directions. The second gives an information lower bound for distribution-free stopping rules in an i.i.d. per-direction observation model. The latter result applies to repeated i.i.d. sampling from per-direction populations. Finite recorded pools, whose means become known after complete enumeration, fall outside this model.

Proposition 4.3 is stated in Section 4.3. Its construction contains D coordinate pairs, with one decision direction g_d supported on each pair. The true parameter is chosen so that every competitor v_d of a reference answer v_0 has regret

$$r_{\theta_\star}(v_d) = \epsilon + \zeta, \qquad d = 1, \ldots, D, \tag{191}$$

for a margin $\zeta > 0$. Sampling follows a deterministic allocation that is symmetric across the D pairs, with ridge $\lambda = 0$.

Let $\mathcal{T}_{\mathrm{bal}}$ be the ends of complete balanced cycles, when the restricted designs are identical across pairs up to relabeling. Let T_{live} be the first balanced checkpoint after which every cross-direction confidence half-width is at most ϵ at all subsequent balanced checkpoints. The *live window* is $\mathcal{T}_{\mathrm{live}} = \{t \in \mathcal{T}_{\mathrm{bal}} : t \geq T_{\mathrm{live}}\}$. At every $t \in \mathcal{T}_{\mathrm{bal}}$, $\|g_d\|_{V_t^{-1}}$ is independent of d; this equality need not hold between checkpoints.

Proof of Proposition 4.3. For each direction d, define the standardized estimation error

$$W_d(t) = \frac{\langle \theta_\star, g_d \rangle - \langle \hat{\theta}_t, g_d \rangle}{\sigma \|g_d\|_{V_t^{-1}}}. \tag{192}$$

At balanced checkpoints with $\lambda = 0$, each $W_d(t)$ has a standard normal marginal distribution. The disjoint supports of the g_d and the symmetric deterministic allocation make the D checkpoint processes independent and identically distributed.

For $t \in \mathcal{T}_{\mathrm{bal}}$, define the shifted boundary

$$b_{t,d} = \frac{R_t(\delta, g_d)}{\sigma} + \frac{\zeta}{\sigma \|g_1\|_{V_t^{-1}}}, \qquad \bar{b}_t = \max_{1 \leq d \leq D} b_{t,d}. \tag{193}$$

Step 1: crossing the shifted boundary produces an invalid answer. Fix a balanced live checkpoint $t \in \mathcal{T}_{\mathrm{live}}$ and suppose

$$\max_{1 \leq d \leq D} W_d(t) \geq \bar{b}_t. \tag{194}$$

Let $d^\star$ attain the maximum. Then

$$W_{d^\star}(t) \geq b_{t,d^\star}, \tag{195}$$

which implies that $v_{d^\star}$ passes its comparison against the reference answer v_0.

For any other competitor v_e, both directions have the same confidence norm and the same true loss relative to v_0. Moreover,

$$W_{d^\star}(t) \geq W_e(t). \tag{196}$$

Hence

$$\langle \hat{\theta}_t, v_{d^\star} - v_e \rangle \leq 0. \tag{197}$$

During the live window, the corresponding cross half-width is at most ϵ, so $v_{d^\star}$ also passes against every v_e. Thus the stopping test accepts $v_{d^\star}$, whose true regret is $\epsilon + \zeta$.

Step 2: lower bound on the largest radius. Define

$$\bar{q} = \Pr\left(\exists t \in \mathcal{T}_{\text{live}} : W_1(t) \geq \bar{b}_t\right). \tag{198}$$

By Step 1 and independence, the probability that at least one invalid answer passes is at least

$$1 - (1 - \bar{q})^D. \tag{199}$$

Validity therefore requires

$$1 - (1 - \bar{q})^D \leq \delta. \tag{200}$$

Hence

$$\bar{q} \leq 1 - (1 - \delta)^{1/D} \leq \frac{\log(1/(1-\delta))}{D}. \tag{201}$$

At any fixed $t \in \mathcal{T}_{\text{live}}$,

$$\bar{q} \geq \bar{\Phi}(\bar{b}_t), \tag{202}$$

where

$$\bar{\Phi}(x) = 1 - \Phi(x) \tag{203}$$

is the standard normal survival function. Consequently,

$$\bar{b}_t \geq \Phi^{-1}\left(1 - \frac{\log(1/(1-\delta))}{D}\right). \tag{204}$$

Substituting the definition of $\bar{b}_t$ gives

$$\max_{1 \leq d \leq D} \frac{R_t(\delta, g_d)}{\sigma} \geq \Phi^{-1}\left(1 - \frac{\log(1/(1-\delta))}{D}\right) - \frac{\zeta}{\sigma \|g_1\|_{V_t^{-1}}}. \tag{205}$$

Using the standard Gaussian tail asymptotic

$$\Phi^{-1}(1-y) = \sqrt{2\log(1/y)}\,(1 + o(1)), \qquad y \downarrow 0, \tag{206}$$

and

$$\frac{\log(1/(1-\delta))}{D} \leq \frac{2\delta}{D} \qquad \text{for } \delta \leq \frac{1}{2}, \tag{207}$$

the leading term becomes

$$\sqrt{2\log(D/\delta)}\,(1 + o(1)) \tag{208}$$

as $D/\delta \to \infty$.

Step 3: symmetric radii. Suppose the radii are identical across directions. Then

$$b_{t,d} = \bar{b}_t \qquad \forall d. \tag{209}$$

Let

$$q_\zeta = \Pr\left(\exists t \in \mathcal{T}_{\text{live}} : W_d(t) \geq b_{t,d}\right) \tag{210}$$

denote the per-direction probability of crossing the shifted boundary. Step 1 lower-bounds the failure probability by

$$1 - (1 - q_\zeta)^D. \tag{211}$$

Define the live-window crossing probability for the unshifted boundary by

$$p_{\text{live}} = \Pr\left(\exists t \in \mathcal{T}_{\text{live}} : W_d(t) \geq \frac{R_t(\delta, g_d)}{\sigma}\right). \tag{212}$$

Continuity of the Gaussian law gives

$$q_\zeta \uparrow p_{\text{live}} \qquad \text{as } \zeta \downarrow 0. \tag{213}$$

If $p_{\text{live}} = \delta$, the lower bound on the probability that some invalid answer passes converges to

$$1 - (1-\delta)^D > \delta \qquad \forall D \geq 2. \tag{214}$$

The construction establishes the claimed necessity of the direction-union penalty. □

Proposition B.12 (Information floor for the i.i.d. per-direction model)**.** *Consider a collection of directions (b, j), each associated with a probability law P_{bj} supported on $[-B, B]$. Observations from each direction are i.i.d. draws from its law and are independent across directions.*

Suppose a stopping rule uses only these per-direction observations and is valid at level δ for every such product distribution. Let b be an item that the rule declares with probability at least $\beta > \delta$. For every direction (b, j) incident to b,

$$\mathbb{E}\left[n_{bj}(\tau)\right] \geq \frac{\mathrm{kl}(\beta, \delta)}{\mathrm{KL}_j^\star(\epsilon)}, \tag{215}$$

where $n_{bj}(\tau)$ is the number of observations collected from direction (b, j) before stopping,

$$\mathrm{kl}(p, q) = p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q} \tag{216}$$

is the binary relative entropy, and

$$\mathrm{KL}_j^\star(\epsilon) = \inf_{Q:\, \mathrm{supp}(Q) \subseteq [-B,B],\, \mathbb{E}_Q[X] > \epsilon} \mathrm{KL}(P_j \| Q). \tag{217}$$

Proof of Proposition B.12. Fix a direction (b, j). Let P denote the original product distribution over all directions, and choose any law Q on $[-B, B]$ satisfying

$$\mathbb{E}_Q[X] > \epsilon. \tag{218}$$

Construct an alternative product distribution P' by replacing the law of direction (b, j) with Q and leaving every other direction unchanged.

Under P', item b has regret greater than ϵ. Level-δ validity therefore implies

$$\Pr_{P'}\left(\hat{a} = b,\, \tau < \infty\right) \leq \delta. \tag{219}$$

By hypothesis,

$$\Pr_{P}\left(\hat{a} = b,\, \tau < \infty\right) \geq \beta. \tag{220}$$

Let LLR_τ denote the log-likelihood ratio between P and P' evaluated at the stopping time. Since the two models differ only on direction (b, j), and query selection is predictable, Wald's identity gives

$$\mathbb{E}_P[\mathrm{LLR}_\tau] = \mathbb{E}_P[n_{bj}(\tau)]\, \mathrm{KL}(P_j \| Q). \tag{221}$$

Applying data processing to the event

$$\{\hat{a} = b,\, \tau < \infty\} \tag{222}$$

gives

$$\mathbb{E}_P[\mathrm{LLR}_\tau] \geq \mathrm{kl}(\beta, \delta). \tag{223}$$

Combining the two inequalities yields

$$\mathbb{E}_P[n_{bj}(\tau)] \geq \frac{\mathrm{kl}(\beta, \delta)}{\mathrm{KL}(P_j \| Q)}. \tag{224}$$

Taking the infimum over all admissible alternatives Q gives

$$\mathbb{E}_P[n_{bj}(\tau)] \geq \frac{\mathrm{kl}(\beta, \delta)}{\mathrm{KL}_j^\star(\epsilon)}. \tag{225}$$

The change-of-measure step is the standard transportation argument (Kaufmann et al., 2016, Lemma 1). □