REPLANNING PRECISION FOR PARETO IDENTIFICATION WITH ACCUMULATED EVIDENCE

Anonymous authors
Paper under double-blind review

ABSTRACT

How should existing measurements guide the next experiment in Pareto set identification? We distinguish the remaining cost of completing a sufficient certificate from the information needed to exclude incorrect answers. A three-candidate construction reverses their preferred history allocation: concentration completes a rectangular certificate, while balanced history requires less additional Gaussian information cost. We explain this reversal through convexity and symmetry: averaging histories over an information-preserving symmetry cannot increase residual information cost, whereas alternative certificate routes can favor concentration. For fixed routes, we establish complementary bottlenecks and a suffix identity under which frozen progress scores tie across required actions. We implement certificate and information replanning with primal–dual allocation bounds and an anytime-valid Gaussian stopping rule. Across 1,068 synthetic trajectories, certificate replanning has lower observed mean cost on two targeted historical suites; an information planner's initial gain disappears on fresh seeds, and native PSIPS uses fewer capped measurements on cold-start suites. The results separate history value, certificate geometry, and sequential execution as distinct design problems.

1 INTRODUCTION

Sequential experiments often begin with uneven historical measurements. With several objectives and no fixed scalar utility, the task is to identify the Pareto set while minimizing the cost of additional measurements.

History placement matters because Pareto decisions have different logical requirements. Exclusion needs one verified dominator; acceptance must rule out every possible dominator. A measurement can support several comparisons, and several certificate routes can establish the same answer. Planning the cheapest remaining route is therefore natural. Whether that route captures the statistical value of history is the question studied here.

Our motivating construction has three candidates and two objectives. Concentrating history on one witness completes a rectangular exclusion certificate. Splitting the same budget leaves both routes incomplete, yet increases information against the closest answer-changing Gaussian alternative. A symmetry argument explains this contrast: information allocation admits beneficial averaging, while certificate routing takes a minimum over competing proofs.

We connect this geometry to computation and execution. A cutting-plane solver allocates residual information with computable lower and upper bounds. A common Gaussian stopping certificate separates allocation effects from stopping calibration. Fresh-seed comparisons then locate where certificate replanning helps and where information planning fails to improve stopping cost.

APE studies adaptive Pareto identification (Kone et al., 2023); Pareto Track-and-Stop solves information allocation (Crépon et al., 2024); PSIPS uses posterior sampling (Kone et al., 2025a). Offline-data best-arm identification formalizes additional sampling cost (Agrawal et al., 2023). Building on these formulations, we study how certificate choice changes the value assigned to accumulated evidence.

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.

2 PROBLEM AND TWO MEASURES OF REMAINING EVIDENCE

There are K candidates and m objectives to maximize. Querying candidate i costs $c_i > 0$ and reveals an independent Gaussian vector with mean μ_i and known diagonal covariance. For the information oracle and the new experiments, the covariance is common across candidates, $\Sigma = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_m^2)$. The historical counts n^0 are deterministic, and historical and future observations follow the same stationary distributions. No target labels or future noise are available to the policy.

Candidate j dominates i when $\mu_{j\ell} \geq \mu_{i\ell}$ for every ℓ, with at least one strict inequality. Write $\mathcal{P}(\mu)$ for the complete nondominated set. A procedure is δ-correct if its returned set equals $\mathcal{P}(\mu)$ with probability at least $1 - \delta$, jointly over historical and online observations. The experiments cap additional cost and record censoring separately. General termination statements use an identifiable class with strict certificate margins: each excluded candidate has a strictly dominating witness, and every accepted candidate has a strictly favorable coordinate against each other candidate.

Sufficient certificates. For confidence bounds L, U, exclude i if some j has $L_{j\ell} > U_{i\ell}$ for all ℓ; accept i if every $j \neq i$ has some ℓ with $L_{i\ell} > U_{j\ell}$. A fixed sufficient route can be represented by edges E and positive comparison gaps h_e. Given decreasing convex radii r_i, its remaining continuous cost is

$$V_E(n) = \min_{N \geq n} c^\top (N - n) \quad \text{s.t.} \quad r_i(N_i) + r_j(N_j) \leq h_{ij}, \quad (i,j) \in E. \tag{1}$$

For finitely many valid routes $\mathcal{C}$, let $W(n) = \min_{C \in \mathcal{C}} V_C(n)$. These fixed-gap planning objectives charge additional measurements. Online policies substitute empirical gaps and update their plans as observations arrive.

Information against incorrect answers. Define $\mathrm{Alt}(\mu) = \{\nu : \mathcal{P}(\nu) \neq \mathcal{P}(\mu)\}$ and

$$I_\mu(N) = \inf_{\nu \in \mathrm{Alt}(\mu)} \sum_i N_i D_i(\mu, \nu), \qquad D_i(\mu, \nu) = \tfrac{1}{2}(\mu_i - \nu_i)^\top \Sigma^{-1} (\mu_i - \nu_i). \tag{2}$$

Boundary alternatives are understood through the closure when the infimum is unattained. For a fixed evidence threshold $\beta > 0$, define

$$C_\beta(n) = \min_{x \geq 0} \{c^\top x : I_\mu(n + x) \geq \beta\}. \tag{3}$$

The function I_μ is concave, nondecreasing, and positively homogeneous in counts, since it is an infimum of nonnegative linear functions. Its superlevel set is convex. Accordingly, (3) is a convex allocation problem even when W is not.

Proposition 1 (Historical information lower bound)**.** *For an almost surely terminating, jointly δ-correct procedure with finite expected additional counts T_i, $0 < \delta < 1/2$,*

$$\mathbb{E}_\mu \sum_i c_i T_i \;\geq\; C_{\mathrm{kl}(1-\delta,\delta)}(n^0).$$

The proof includes historical likelihood terms in the change-of-measure argument of Kaufmann et al. (2016); Agrawal et al. (2023); see Appendix A. The expectation is joint over history and future observations, and the oracle program uses population means.

3 THE GEOMETRY CAN REVERSE THE VALUE OF HISTORY

Theorem 2 (Fixed-route geometry)**.** *Suppose $n_i > 0$, $c_i > 0$, $h_e > 0$, and each radius is continuous, decreasing, convex, and tends to zero. Then V_E is convex and nonincreasing. If $n' \geq n$,*

$$0 \leq V_E(n) - V_E(n') \leq c^\top (n' - n). \tag{4}$$

If N^* is an optimal target and $n \leq n' \leq N^*$, then

$$V_E(n') = V_E(n) - c^\top (n' - n). \tag{5}$$

Every action inside a frozen optimal target reduces predicted remaining cost by exactly its price. Thus progress per paid unit equals one for all such actions: the frozen value alone supplies no execution order.

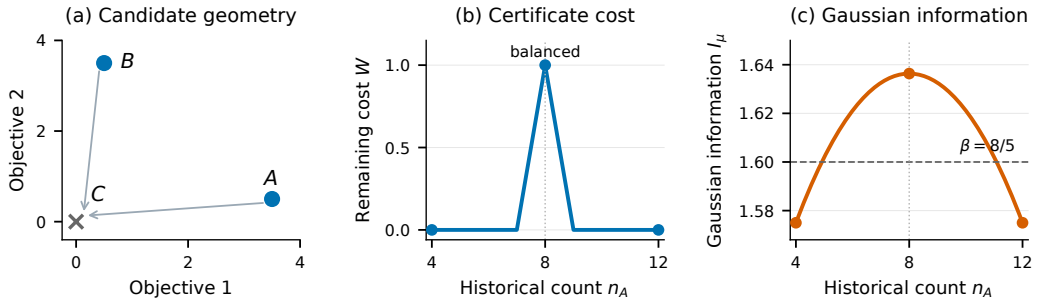


Figure 1: Two measures of historical evidence in Theorem 3. (a) Either A or B can certify exclusion of C. With continuous planning counts $n_A + n_B = 16$ and $n_C = 36$, (b) concentration completes a rectangular route, whereas (c) balance maximizes Gaussian information and exceeds the threshold $\beta = 8/5$. The curves use the stated planning radius $r(n) = n^{-1/2}$ and exact Gaussian information, respectively.

Complementary requirements. For a star with edges $(1, 2), (2, 3)$, radii $r(N) = N^{-1/2}$, unit costs and gaps, histories $(a, 1, b)$ with $a, b \geq 4$ have

$$V(a, 1, b) = \left(1 - \min(a, b)^{-1/2}\right)^{-2} - 1.$$

The four histories $(9, 1, 9), (16, 1, 9), (9, 1, 16), (16, 1, 16)$ have remaining costs $5/4, 5/4, 5/4, 7/9$. Either seven-measurement leaf bundle alone has no value; together they save $17/36$. Integer optima are $2, 2, 2, 1$. Thus armwise diminishing returns fails despite convexity. These are concrete certificate examples of established complementarity concepts (Golovin & Krause, 2011; Chen & Waggoner, 2016).

Theorem 3 (Certificate–information preference reversal)**.** *Consider unit costs, independent unit-variance Gaussian objectives, means*

$$\mu_A = (3.5, 0.5), \qquad \mu_B = (0.5, 3.5), \qquad \mu_C = (0, 0),$$

and the stylized rectangular planning radius $r(N) = N^{-1/2}$. Let

$$n^a = (12, 4, 36), \quad n^b = (4, 12, 36), \quad \bar{n} = (8, 8, 36).$$

The weak rectangular planning costs are $W(n^a) = W(n^b) = 0$ and $W(\bar{n}) = 1$. However,

$$I_\mu(n^a) = I_\mu(n^b) = \frac{63}{40} < I_\mu(\bar{n}) = \frac{18}{11}.$$

At $\beta = 8/5$, the information allocation costs reverse the ranking:

$$C_\beta(n^a) = C_\beta(n^b) = \frac{40}{161} > C_\beta(\bar{n}) = 0.$$

Both A and B belong to the true front. A rectangular exclusion of C can use either witness, whose smaller gap is 0.5. Concentrating history finishes one route. Information testing, however, can combine evidence from both witnesses against an alternative that makes C nondominated. The threshold and radius in this construction specify two optimization models; they are not asserted to represent identical confidence calibrations. Appendix C gives an exact rational calculation and optimality certificate.

Figure 1 varies the placement of the same historical budget, keeping the means and observation model fixed.

The first part also proves nonconvexity of W: both endpoint costs vanish while the midpoint cost is positive. Strict integer certification at the midpoint needs two measurements rather than one. The nonconvexity persists for any continuous strictly decreasing common radius by choosing a gap strictly between concentrated and balanced radius sums. The information-feasible region remains convex. Therefore, nonconvex certificate routing is not an intrinsic nonconvex information lower bound.

Proposition 4 (Symmetrizing accumulated information)**.** *Let G be a finite group of candidate permutations such that $I_\mu(P_g N) = I_\mu(N)$ and $c^\top P_g = c^\top$ for every $g \in G$. For $\bar{n} = |G|^{-1} \sum_g P_g n$,*

$$I_\mu(\bar{n}) \geq I_\mu(n), \qquad C_\beta(\bar{n}) \leq C_\beta(n).$$

Concavity proves the information inequality. Averaging permuted feasible targets proves the cost inequality (Appendix D). In Theorem 3, swapping A, B and the two objectives preserves the Gaussian instance. Balancing the two witness counts therefore improves the information objective, while the certificate objective favors completing either witness. This consequence of convex averaging explains the reversal beyond the particular numerical threshold.

Total cost can conceal residual inefficiency. Let a cold optimum N^{cold} minimize total cost subject to a fixed route. Clipping it to history, $\max(n, N^{\mathrm{cold}})$, gives a total cost at most twice the history-aware optimum, but its remaining-cost ratio can be arbitrarily large. For a positive integer q, radius $r(N) = \sqrt{q/N}$, and unit gap and costs, the cold pair target is $(4q, 4q)$. With integer history (M, q) and

$$M > q \left(1 - \sqrt{q/(q+1)}\right)^{-2},$$

one new sample at the second endpoint suffices, while cold-target clipping requests $3q$. The count-floor ablation in Section 5 tests the online effect of this deterministic distinction.

4 Residual allocation and valid sequential execution

4.1 Information allocation by constraint generation

We independently implement the two-objective Gaussian alternative geometry of Crépon et al. (2024). Standardizing each objective reduces to unit variance. The oracle considers removing a front candidate or adding a non-front candidate. In two dimensions, sorting the front makes the addition subproblem a sequence of threshold partitions; Appendix E describes the independent implementation.

At counts N, the oracle returns $I_\mu(N)$ and a vector $g_i = D_i(\mu, \nu^*)$ for a minimizing alternative. Any feasible allocation must satisfy $g^\top(n + x) \geq \beta$. We accumulate these cuts and solve an LP. If A contains rows $g^\top/\beta$, the relaxation is

$$\min_{x \geq 0} c^\top x \quad \text{s.t.} \quad Ax \geq \mathbf{1} - An.$$

A feasible dual vector $\lambda \geq 0$, $A^\top \lambda \leq c$, gives the lower bound $\lambda^\top(\mathbf{1} - An)$. Homogeneity gives a feasible upper target by scaling a positive trial N by $\max\{1, \beta/I_\mu(N)\}$. Their relative gap controls optimization accuracy for the fixed-mean program.

In 120 independently generated states, the oracle's maximum relative discrepancy against the published-code oracle is 6.42×10^{-14}. All 24 allocation checks reach relative primal–dual gap 10^{-4}. Additional checks cover supporting-function inequalities, homogeneity, and finite differences.

4.2 One stopping rule shared by the allocation comparisons

For the new experiments, split error budget δ equally between a rectangular certificate and a full-vector ellipsoid. The rectangular radius is

$$r_{i\ell}(n) = \sigma_\ell \frac{\sqrt{(n+1)\{2\log(2Km/\delta) + \log(n+1)\}}}{n}. \tag{6}$$

For the ellipsoid, set $w_i = n_i^2/(n_i + 1)$ and

$$I_{\widehat{\mu}}(w) > \log(2/\delta) + \frac{m}{2} \sum_i \log(1 + n_i). \tag{7}$$

If the rectangular certificate completes, return its accepted set. If (7) holds, return $\mathcal{P}(\widehat{\mu})$. Otherwise continue.

Proposition 5 (Anytime correctness)**.** *Under the independent known-Gaussian observation model, either stopping route above returns an incorrect Pareto set with combined probability at most δ, for any predictable sampling policy. Historical data must be included in the counts and sums. If sampling forces every candidate to be observed infinitely often on nontermination, strict positive certificate margins imply almost-sure finite termination.*

Appendix F derives the rule using Gaussian mixture martingales and Ville's inequality (Kaufmann & Koolen, 2021). The theorem uses an exact information oracle; the implemented oracle handles $m = 2$.

4.3 Policies and a stopping-alignment follow-up

Certificate replanning. The full planner constructs unresolved comparison clauses, chooses sufficient witness routes, and solves (1) with empirical gaps and the actual mixture radius. It selects the largest relative count deficit $1 - n_i/N_i$. Failed numerical solves fall back to a feasible target and remain counted. The implementation label `exact` denotes the radius model, not exact attainment on every solve.

Information replanning. Every 16 queries, the information planner solves (3) using current empirical means, count floors $n_i - 1$, and the current threshold in (7). Planning counts M correspond to actual targets $N = M + 1$; $N_i - 1 \le N_i^2/(N_i + 1)$ makes the linear planning weights conservative at a fixed threshold. We allow 20 cutting-plane iterations and a 2% relative gap, retaining a feasible upper target if convergence is incomplete. The policy follows relative deficits, with a local information-gradient fallback after the frozen target is reached. Actual stopping always recomputes (7) or the rectangular certificate from observations.

Competing plans. After the initial results, we test whether choosing the cheaper of a certificate plan and an information plan improves execution. The information target is scaled until it satisfies its *future* frozen-mean threshold, so both candidates are feasible in their own deterministic planning models. The control recomputes only the certificate plan at the same 16-query cadence. This isolates competition from update frequency. All common-stopping policies use the same minimum-count exploration floor $\max\{2, \lfloor\sqrt{(t+1)/K}\rfloor\}$, where t counts new queries.

5 Experiments

5.1 Design and baseline fidelity

We use synthetic, independent Gaussian vector observations with known noise. Costs are one in the new online comparisons. Historical OR tasks have $K \in \{3, 8\}$, gaps $g \in \{0.35, 0.40\}$, means $(3, g), (g, 3), (0, 0)$, and, when present, clearly dominated additional candidates. Each configuration compares histories $(168, 8, 64)$ and $(88, 88, 64)$, with eight observations for every added candidate. Corresponding methods receive identical history and future per-arm streams.

Random cold tasks use eight candidates in hub, front, and alternative-witness geometries, with respective common standard deviations $0.4, 0.15, 0.25$. The cap is 2,000 added measurements for historical tasks and 4,000 for cold tasks. Censored trajectories are charged the full cap; their mean is a restricted cost, not an uncensored mean stopping time. The first new stage uses 12 seed blocks and 468 trajectories. A separately frozen follow-up uses 12 different blocks and 600 trajectories. Development seeds are disjoint. Intervals use 10,000 paired seed-block bootstrap resamples and are descriptive, pointwise, and unadjusted for multiple comparisons.

APE-common retains APE-style pivot/challenger selection with the shared certificate. APE-native is an independent reimplementation of the published exact-set equations with the Gaussian calibration of Kaufmann & Koolen (2021), rather than the approximate experimental calibration. PSIPS-native runs the authors' C++ `peps_psi` implementation at commit `0b78a2f`, retaining its empirical stopping calibration. We add a replay environment and an independent query counter. The source initializes its reported counts to one per arm without those corresponding sample calls, so the reported count exceeds actual calls by K; both quantities are retained. Header preincludes resolve compilation

Table 1: First new stage. Mean capped added measurements; all methods within the first three rows share stopping and exploration. Native rows use their own stopping rules.

Method	Historical	Cold	Historical complete	Cold complete
APE-common	47.36	689.31	96/96	34/36
Certificate planner	43.34	672.72	96/96	34/36
Information planner	42.00	700.69	96/96	34/36
APE-native	—	713.83	—	34/36
PSIPS-native	—	559.53	—	34/36

Table 2: Separately seeded follow-up. The primary competition comparison holds the update cadence at 16 queries. All historical runs (96 per method) and cold runs (36 per method) complete.

Method	Historical	Cold
APE-common	43.61	460.89
Certificate planner, every 16 queries	37.96	454.86
Competing plans, every 16 queries	42.31	454.86
Information planner, every 16 queries	43.83	475.36
Certificate planner, every query	—	441.44
PSIPS-native	—	336.89

dependencies without changing policy equations. Native results are complete-algorithm comparisons with different stopping rules.

5.2 Certificate ablations

The preceding investigation contains 7,120 trajectories across historical placement, count-floor, action-branch, and witness-selection studies. They are not 7,120 independent tasks. An equal-history-budget experiment found lower additional cost from jointly strengthening two required comparisons. A count-floor ablation, however, produced nearly equal stopping costs with and without the history-aware optimization floor. Consequently, targeted gains cannot be attributed to that floor alone.

A subsequent witness-focused APE adaptation reduced mean additional cost from 54.78 to 37.48 on designed warm OR tasks, but increased cost from 446.75 to 449.55 on randomized cold controls. A simpler margin witness rule was cheaper than its cost-based repair on the targeted suite. These stages motivate the new information comparison and are retained in the evidence archive. Appendix G gives the core table and the failed causal-attribution tests.

5.3 Sequential cost and replication

In Table 1, information planning reduces historical-task mean cost by 11.33% relative to APE-common (95% interval, 1.66–20.55% reduction), but its 3.10% advantage over the certificate planner is unresolved (signed change interval $[-8.45, 4.41]\%$). On cold tasks it increases cost by 1.65% relative to APE-common and by 4.16% relative to the certificate planner. PSIPS-native stops with fewer capped measurements, although its stopping calibration differs. No erroneous completed answers occur; ten trajectories are censored, two per cold method.

The follow-up in Table 2 gives historical means of 43.83 for information planning and 43.61 for APE-common, so the initial advantage does not replicate. Competing plans increase cost by 11.47% over certificate-only planning at the same cadence (interval, 0.69–24.38% increase). On cold tasks these two policies have identical per-task costs, and native PSIPS again uses fewer measurements. All runs complete with correct outputs.

Where certificate replanning helps. A post-hoc split by history layout locates the follow-up advantage. With concentrated witness history, certificate replanning costs 32.77 measurements versus 45.13 for APE-common; with balanced witness history, the means are 43.15 and 42.10. Information

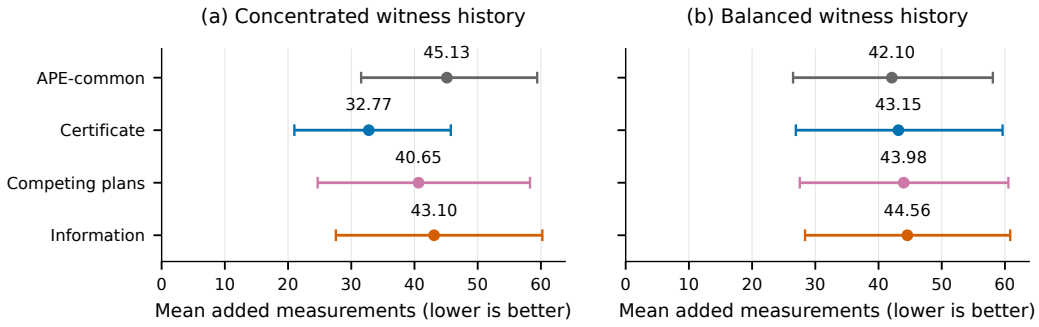


Figure 2: Follow-up costs by history layout, using all four common-stopping methods. Points show mean additional measurements; bars are pointwise 95% seed-block bootstrap intervals with 12 seed blocks and 48 trajectories per method and layout. All methods use paired resampling indices. Certificate replanning has the lowest observed mean under concentrated history; balanced history yields similar means. This is the post-hoc split of Table 2.

planning costs 43.10 and 44.56 in the same strata. Thus the aggregate certificate gain is concentrated in a specific starting layout. These descriptive comparisons reuse the same evaluation data; all strata and paired seed-block intervals are supplied in the supplement.

Figure 2 shows the stratum means and their sampling variability.

Numerical accounting. In the first stage, information planning makes 1,836 successful optimization calls, records five degenerate-geometry fallbacks, and leaves two calls above the requested 2% gap after the iteration cap. Certificate planning records 28 fallbacks. In the follow-up, competing plans record four fallbacks and no above-tolerance calls; certificate-only planning at the same cadence records one fallback. These events are retained in the run records. Concurrent CPU timings are descriptive and are not used as evidence of implementation-independent runtime superiority.

6 Related Work and Implications

The closest algorithmic literature is pure-exploration Pareto identification. APE (Kone et al., 2023), Pareto Track-and-Stop (Crépon et al., 2024), PSIPS (Kone et al., 2025a), and constrained PSI (Kone et al., 2025b) already study adaptive evidence collection and information-theoretic difficulty. The recent anytime Bayesian approach of Saerens et al. (2026) studies continuing recommendations and molecular exploration; its anytime setting differs from our fixed-confidence historical-count question.

Offline best-arm identification (Agrawal et al., 2023) already studies additional samples, information allocation, and conditional versus joint historical correctness. Our historical KL program is an application of this foundation to the Pareto alternative set. Logical evidence acquisition and complementarity likewise have established theories (Deshpande et al., 2014; Chen & Waggoner, 2016); our examples specify how they appear in repeated noisy vector measurements. Generic convex sensitivity and cutting-plane optimization are used as tools, with their specializations proved for completeness.

The results separate three questions: which history is informative, which certificate is cheap, and which experiment advances the stopping decision. Convex averaging addresses the first; route selection addresses the second; the suffix identity shows why the first two do not determine execution order. The history-stratified results motivate policies that explicitly model changes in the active comparison as new observations arrive.

Scope. The experiments cover synthetic, stationary Gaussian tasks with known noise and two objectives. Their 12 seed blocks per stage support local comparisons; cross-task transfer and real scientific assays remain untested. The stopping guarantee controls error and establishes almost-sure termination, while expected-cost optimality remains open. Native-baseline comparisons include differences in stopping calibration.

7 CONCLUSION

The value of accumulated evidence depends on how a Pareto decision is certified. We characterize a concentration–balancing reversal, explain its information side by symmetry, and derive the execution-order ambiguity of frozen cost plans. Sequential comparisons reveal layout-specific certificate gains and unstable information-planning gains. These findings make stopping-cost alignment the central issue in turning historical information into efficient experiments.

AI USE STATEMENT

The submitting author takes final responsibility for this research, including research ideation, mathematical derivations and proofs, code generation and execution, synthetic experiments, and result analysis, and is responsible for verifying the paper's claims, references, code, results, and final text. An OpenAI assistant supported these research activities as well as literature retrieval and manuscript drafting, editing, and translation.

REPRODUCIBILITY STATEMENT

The supplement contains frozen evaluation protocols, all run records, numerical checks, code hashes, and the post-hoc stratum analysis. Appendix H specifies the environment, commands, and baseline provenance; Appendices A–F give the proofs and oracle derivation.

REFERENCES

Shubhada Agrawal, Sandeep Juneja, Karthikeyan Shanmugam, and Arun Sai Suggala. Optimal best-arm identification in bandits with access to offline data, 2023. URL https://arxiv.org/abs/2306.09048.

Yiling Chen and Bo Waggoner. Informational substitutes. In *IEEE Symposium on Foundations of Computer Science*, 2016. URL https://arxiv.org/abs/1703.08636.

Élise Crépon, Aurélien Garivier, and Wouter M. Koolen. Sequential learning of the pareto front for multi-objective bandits. In *International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pp. 3583–3591, 2024. URL https://proceedings.mlr.press/v238/crepon24a.html.

Amol Deshpande, Lisa Hellerstein, and Devorah Kletenik. Approximation algorithms for stochastic boolean function evaluation and stochastic submodular set cover. In *ACM-SIAM Symposium on Discrete Algorithms*, 2014. URL https://arxiv.org/abs/1303.0726.

Daniel Golovin and Andreas Krause. Adaptive submodularity: Theory and applications in active learning and stochastic optimization. *Journal of Artificial Intelligence Research*, 42:427–486, 2011. URL https://arxiv.org/abs/1003.3967.

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.

Cyrille Kone, Emilie Kaufmann, and Laura Richert. Adaptive algorithms for relaxed pareto set identification. In *Advances in Neural Information Processing Systems*, 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/6e976e7930460b5c3167a104ba8cc39c-Abstract-Conference.html.

Cyrille Kone, Marc Jourdan, and Emilie Kaufmann. Pareto set identification with posterior sampling. In *International Conference on Artificial Intelligence and Statistics*, volume 258 of *Proceedings*

of Machine Learning Research, pp. 1000–1008, 2025a. URL https://proceedings.mlr.press/v258/kone25a.html.

Cyrille Kone, Emilie Kaufmann, and Laura Richert. Constrained pareto set identification with bandit feedback. In *International Conference on Machine Learning*, 2025b. URL https://proceedings.mlr.press/v267/kone25a.html.

Lennert Saerens, Bram Silue, Eleni Litsa, Peter Vrancx, and Pieter Libin. Bayesian anytime pareto set identification for multi-objective multi-armed bandits, 2026. URL https://arxiv.org/abs/2606.18785.

A Proof of the historical information bound

Let μ and ν be identifiable instances with different Pareto sets. The log likelihood ratio of the complete stopped transcript is the sum of historical likelihood ratios and likelihood ratios of subsequently queried observations. Conditional action probabilities cancel: the same predictable policy is used under both instances. The standard stopped change-of-measure identity, under the finite expected-count assumption, gives

$$\mathrm{KL}(\mathbb{P}_\mu^{\mathrm{transcript}}, \mathbb{P}_\nu^{\mathrm{transcript}}) = \sum_i (n_i^0 + \mathbb{E}_\mu T_i) D_i(\mu, \nu).$$

For the event $E = \{\widehat{\mathcal{P}} = \mathcal{P}(\mu)\}$, correctness gives $\mathbb{P}_\mu(E) \geq 1 - \delta$ and $\mathbb{P}_\nu(E) \leq \delta$. Data processing to $\mathbf{1}_E$ and monotonicity of binary relative entropy on these ranges imply

$$\sum_i (n_i^0 + \mathbb{E}_\mu T_i) D_i(\mu, \nu) \geq \mathrm{kl}(1 - \delta, \delta).$$

This holds for every answer-changing alternative. Therefore $x_i = \mathbb{E}_\mu T_i$ is feasible in (3), proving Proposition 1. If expected additional cost is infinite, the finite lower bound is immediate.

The probability statement is joint over historical and future data. Requiring correctness separately for each fixed historical realization would be a different model: supplying the same fixed history under both instances contributes no additional conditional KL. With adaptively logged history under a common predictable logging rule, expected historical counts enter the corresponding unconditional identity. Shifted historical distributions require separate likelihood terms and cannot be represented by pooled target counts without assumptions.

B Fixed-certificate value and complementary history

Let $F = \{N > 0 : r_i(N_i) + r_j(N_j) \leq h_{ij},\ (i,j) \in E\}$. It is convex and upward closed. A sufficiently large target is feasible. On $N \geq n > 0$, the positive-cost objective is coercive, so a minimum exists.

For histories n^a, n^b and optimal targets N^a, N^b, the convex combination $N^\theta = \theta N^a + (1-\theta) N^b$ is feasible and dominates n^θ. Its objective is $\theta V_E(n^a) + (1-\theta) V_E(n^b)$, proving convexity. For $n' \geq n$, a feasible target at n' is feasible at n, so $V_E(n') \geq V_E(n) - c^\top (n' - n)$. Conversely, $\max\{N^*, n'\}$ is feasible at n' and costs no more than $V_E(n)$, proving monotonicity. If $n' \leq N^*$, using N^* itself attains the opposite inequality with equality and proves the suffix identity. That identity only requires a fixed feasible set, not convexity.

For differentiable radii, let $\lambda_e \geq 0$ and $\eta_i \geq 0$ be multipliers for the edge and count-floor constraints. The stationarity equations are

$$c_i + r_i'(N_i^*) \sum_{e \ni i} \lambda_e - \eta_i = 0, \quad \eta_i (n_i - N_i^*) = 0, \quad \lambda_e \{r_i(N_i^*) + r_j(N_j^*) - h_e\} = 0.$$

Slater's condition holds. An optimal dual pair yields the value subgradient $\eta - c$. At differentiability points, the marginal historical saving is $c_i - \eta_i \in [0, c_i]$. At a nondifferentiable point, a single selected multiplier does not encode every directional derivative.

For the star example, put $u = \min(a, b)$ and $T(u) = (1 - u^{-1/2})^{-2}$. If $a \leq b$, the feasible target is $(a, T(a), b)$. Choose $\lambda_{12} = 2T(a)^{3/2}$, $\lambda_{23} = 0$, $\eta_1 = 1 - [T(a)/a]^{3/2}$, $\eta_2 = 0$, and $\eta_3 = 1$. Because $a \geq 4$, all multipliers are nonnegative. These values satisfy stationarity and complementary slackness, establishing global optimality. Interchange leaves when $b < a$. Substituting $a, b \in \{9, 16\}$ gives the displayed rational values. Enumerating increments of total cost at most two gives integer costs $2, 2, 2, 1$.

The total-cost factor and residual gap. Let $T(n) = \min_{N \in F,\ N \geq n} c^\top N$. Both $c^\top n$ and $c^\top N^{\mathrm{cold}}$ are at most $T(n)$, and upward closure makes the clipped target feasible. Consequently,

$$c^\top \max\{n, N^{\mathrm{cold}}\} \leq c^\top n + c^\top N^{\mathrm{cold}} \leq 2T(n).$$

For the two-candidate construction, convexity and symmetry give the cold target $(4q, 4q)$. The displayed condition on M is precisely $\sqrt{q/M} + \sqrt{q/(q+1)} < 1$. Thus one new measurement suffices for strict certification. At the original history (M, q) the second radius equals one, so zero new measurements cannot suffice. The optimal integer residual cost is therefore one, while clipping costs $3q$.

Convexity of the actual mixture radius. Write $r(n) = \sigma\sqrt{q(n)}$, $q(n) = (n+1)L/n^2$, and $L = A + \log(n+1)$, where $A = 2\log(2Km/\delta)$. Then

$$q'(n) = \frac{n - (n+2)L}{n^3}, \qquad q''(n) = \frac{2(n+3)L - 3n - n/(n+1)}{n^4}.$$

The numerator of $2qq'' - (q')^2$, after multiplication by n^6, is

$$(3n^2 + 12n + 8)L^2 - 4n(n+1)L - n^2.$$

For $L \geq 2$ this is positive: it is increasing in L there, and at $L = 2$ equals $3n^2 + 40n + 32$. Hence $r'' > 0$. Also $q' < 0$ for $L \geq 2$, so the radius is decreasing. All experiment parameters satisfy this condition.

C Exact certificate–information preference reversal

The true front is $\{A, B\}$. With all counts at least one, the three-unit A–B trade-off and each front point's strong 3.5-unit comparison against C are already certified by the stylized radius. Only exclusion of C remains. The two possible routes require

$$N_A^{-1/2} + N_C^{-1/2} \leq \tfrac{1}{2} \quad \text{or} \quad N_B^{-1/2} + N_C^{-1/2} \leq \tfrac{1}{2}.$$

At either endpoint history a route holds strictly, since $12^{-1/2} + 1/6 < 1/2$. At the midpoint neither route holds. The weak optimum changes one witness count from eight to nine, at cost one. For example, at $(9, 8, 36)$ the active-edge multiplier is 54, the target-arm floor multiplier is $7/8$, and the unused-arm floor multiplier is one; the witness is above its floor. These satisfy KKT. Strict integer certification requires two increments, for example $(9, 8, 37)$; all one-increment choices fail the strict test.

For the Gaussian information calculation, removing A or B costs

$$\frac{9}{2}\frac{N_A N_B}{N_A + N_B}.$$

The relevant cheapest addition alternative promotes C by using its first coordinate against B and its second against A. Minimizing the two weighted quadratic terms gives

$$J(N) = \frac{1}{8}\left(\frac{N_C N_A}{N_C + N_A} + \frac{N_C N_B}{N_C + N_B}\right). \tag{8}$$

The other two sorted-front addition partitions require resolving a 3.5-unit gap in one coordinate. Their cost is at least $(3.5)^2 N_C N_j / [2(N_C + N_j)]$ for the corresponding strong witness, exceeding J at all displayed counts. Removal costs also exceed J. Thus $I_\mu = J$ at $n^a, n^b, \bar{n}$, and at the target constructed below. Substitution gives $63/40, 63/40, 18/11$.

At $n^a = (12, 4, 36)$, the target

$$N^* = (12, 684/161, 36)$$

has $J(N^*) = 8/5$. The two means in the minimizing alternative tie in the small-gap coordinates with $z_1 = 19/360$, $z_2 = 1/8$ for C. The corresponding supporting coefficients are

$$g_A = \frac{9}{128}, \qquad g_B = \frac{25921}{259200}, \qquad g_C = \frac{361}{259200} + \frac{1}{128}.$$

Here g_B is largest. Every feasible target $N \geq n^a$ must satisfy the necessary linear alternative constraint $g^\top N \geq 8/5 = g^\top N^*$. Since $g_i \leq g_B$ and all increments are nonnegative,

$$\sum_i (N_i - n_i^a) \geq \frac{8/5 - g^\top n^a}{g_B} = \frac{40}{161}.$$

The target attains this bound, proving optimality. Symmetry gives the result at n^b. At $\bar{n}$, $18/11 > 8/5$, so no added count is needed. This proves Theorem 3.

The stylized radius and information threshold specify separate planning models. For the actual mixture radius, route nonconvexity follows by taking $a > b \geq 1$ and $d \geq 1$, then choosing h strictly between $r(a)+r(d)$ and $r((a+b)/2)+r(d)$. Means $(M+h, h), (h, M+h), (0, 0)$ with sufficiently large M isolate the same alternative-route structure.

The independent numerical program approximates $40/161$ to within 2×10^{-7} in absolute objective value. At the deliberately stringent 10^{-9} requested gap, two runs hit the 100-iteration cap; their nonconverged status is retained. The exact rational proof establishes the result independently of that tolerance.

D SYMMETRY AND RESIDUAL INFORMATION COST

We prove Proposition 4 for continuous planning counts. Since I_μ is concave,

$$I_\mu\left(\frac{1}{|G|}\sum_g P_g n\right) \geq \frac{1}{|G|}\sum_g I_\mu(P_g n) = I_\mu(n).$$

Take any feasible target $N \geq n$ with $I_\mu(N) \geq \beta$, and define $\bar{N} = |G|^{-1}\sum_g P_g N$. Coordinatewise averaging gives $\bar{N} \geq \bar{n}$. The same concavity argument gives $I_\mu(\bar{N}) \geq \beta$, while cost invariance gives

$$c^\top(\bar{N} - \bar{n}) = \frac{1}{|G|}\sum_g c^\top P_g(N - n) = c^\top(N - n).$$

Thus averaging converts every feasible target into a target for balanced history with the same added cost. Taking the infimum proves the cost inequality, including the case where the infimum is approached rather than attained. More generally, averaging feasible targets for two histories proves convexity of C_β in its history argument.

In the three-candidate construction, exchange of A, B together with exchange of the two objective coordinates preserves the mean matrix, covariance, cost, and Pareto alternatives. Hence G has two elements and the stated invariance holds. On the active-alternative region with means $(L, d), (d, L), (0, 0)$, counts $(s + h, s - h, t)$, and $|h| < s$, direct substitution into the counterpart of (8) gives the balancing gain

$$J(s, s, t) - J(s + h, s - h, t) = \frac{d^2 t^2 h^2}{(t + s)\{(t + s)^2 - h^2\}} > 0 \quad (h \neq 0).$$

At the displayed instance this equals $27/440$. The closed form applies where that addition alternative is minimal; the group-averaging inequality holds across changes of the active alternative.

A separate diagnostic checks information invariance and the averaging inequality on 120 symmetric Gaussian instances. Sixteen additional checks average feasible allocation targets and verify unchanged added cost and retained information feasibility. These checks use the same two-dimensional oracle and supplement the convexity proof.

E TWO-DIMENSIONAL INFORMATION ORACLE AND ALLOCATION BOUNDS

This section explains the independent implementation of the established alternative decomposition of Crépon et al. (2024). Standardize objective ℓ by σ_ℓ. Let P be the empirical front. A removal alternative between $i, j \in P$ has value

$$\frac{1}{2}\frac{N_i N_j}{N_i + N_j}\sum_\ell [\mu_{i\ell} - \mu_{j\ell}]_+^2.$$

For a non-front candidate k, the addition part has the form

$$\inf_{z \in \mathbb{R}^2}\left\{\frac{N_k}{2}\|z - \mu_k\|^2 + \frac{1}{2}\sum_{j \in P} N_j \min_{\ell=1,2}[\mu_{j\ell} - z_\ell]_+^2\right\}.$$

Only $z \geq \mu_k$ need be considered: decreasing a coordinate below its original value raises its own quadratic cost and cannot help any positive-part penalty. For a fixed z, the objective attaining the smaller penalty switches at the threshold $\mu_{j1} - \mu_{j2} = z_1 - z_2$. Along a two-dimensional front sorted by the first coordinate, this difference is nondecreasing. Hence there is an optimal assignment in which a prefix uses coordinate one and the suffix uses coordinate two. Enumerating the $|P| + 1$ splits is sufficient.

For each assigned coordinate, solve

$$\min_z \ \frac{w_0}{2}(z-a)^2 + \frac{1}{2}\sum_j w_j[b_j - z]_+^2.$$

Sort b_j in decreasing order. Start at $z = a$; include successive $b_j > z$ and update z to their weighted average together with a. When the next $b_j \leq z$, the active set is complete. The squared displacements divided by two give the supporting coefficient for each arm. Taking the minimum over removals and additions yields the oracle and one global supporting linear function. Degenerate tied means can produce zero information; the online implementation records an explicit fallback for nearly zero values.

Cutting-plane lower and upper bounds. For rows $A_s = g_s^\top/\beta$, every information-feasible x obeys $Ax \geq \mathbf{1} - An$. Thus a feasible LP dual vector gives a lower bound for the full program. We clip tiny negative multipliers to zero and scale them down if $A^\top\lambda \leq c$ is violated numerically. A positive trial target N with $I_\mu(N) > 0$ becomes feasible after scaling by $\max\{1, \beta/I_\mu(N)\}$. The best such upper target is retained. The reported relative gap is $(U - L)/\max\{1, U\}$.

These bounds hold at the fixed mean matrix, threshold, and history floor supplied to the solver. The follow-up additionally scales its information target to satisfy the future frozen-mean threshold. Execution updates empirical means, and stopping reevaluates the certificate using the observed data.

F MIXTURE STOPPING PROOF AND TERMINATION

For arm i and objective ℓ, write

$$S_{i\ell}(n_i) = \sum_{s=1}^{n_i} \frac{Y_{is\ell} - \mu_{i\ell}}{\sigma_\ell}.$$

For any fixed vector $\theta \in \mathbb{R}^{Km}$, predictability of sampling and Gaussian conditional moment generating functions make

$$\exp\left\{\sum_{i,\ell}\theta_{i\ell}S_{i\ell} - \frac{1}{2}\sum_{i,\ell} n_i\theta_{i\ell}^2\right\}$$

a mean-one nonnegative martingale. Mixing independently with standard normal densities and integrating gives

$$M_t = \prod_i (1+n_i)^{-m/2} \exp\left\{\frac{1}{2}\sum_{i,\ell}\frac{S_{i\ell}^2}{1+n_i}\right\}.$$

By Ville's inequality, with probability at least $1 - \delta/2$, at all times

$$\frac{1}{2}\sum_{i,\ell}\frac{n_i^2}{1+n_i}\frac{(\widehat{\mu}_{i\ell} - \mu_{i\ell})^2}{\sigma_\ell^2} \leq \log(2/\delta) + \frac{m}{2}\sum_i \log(1+n_i).$$

If the returned empirical Pareto set were wrong, the true mean matrix would be an alternative to $\widehat{\mu}$. Its quadratic distance would then upper-bound $I_{\widehat{\mu}}(w)$. Inequality (7) would contradict the displayed coverage event. This proves the ellipsoid route's error bound.

For a single coordinate, the same mixture is $(n+1)^{-1/2}\exp\{S_n^2/[2(n+1)]\}$. Use error $\delta/(2Km)$ and union-bound over coordinates to obtain (6). On simultaneous coverage, every rectangular acceptance and exclusion is correct. A union bound over the two routes proves total error at most δ; the routes need not be independent.

Historical samples are the initial prefix of this martingale. This does not give a uniform conditional guarantee at each realized history. The lower bound and correctness theorem both use the same joint interpretation.

Under the forced minimum-count floor, nontermination implies every count tends to infinity. Independent armwise strong laws give $\widehat{\mu}_i \to \mu_i$ almost surely, and rectangular radii tend to zero. Strict positive certificate margins then make all finitely many required tests true after a finite random time, proving almost-sure finite termination. Tied front points can prevent strict certification; a theorem covering them requires an approximate-output definition.

G EARLIER MECHANISM EVIDENCE

The preceding stages have the following accounting. The table counts evaluation trajectories, including censored trajectories; it is not a count of statistically independent tasks.

Stage	Trajectories	Censored	Erroneous completed
Fixed-history first-action branches	768	0	0
Equal-budget history placement	1920	0	0
Count-floor ablation	1536	0	0
Margin witness selection	1736	10	0
Completion-cost witness follow-up	1160	0	0

The first-action study uses eight fixed histories and 32 continuation streams per history; sixteen select an action and sixteen evaluate it. The selected action coincides with the original planner at every state, giving no held-out gain. This is not evidence that all possible actions have identical value. A future-informed prefix analysis is diagnostic only and is not a deployed policy.

The count-floor ablation retains empirical means, route selection, tracking, and stopping, but solves at count floor one before clipping to actual history. Mean costs are 129.24 versus 129.21 for cold-floor versus history-floor planning on AND tasks, and 38.71 versus 38.46 on OR tasks. The deterministic unbounded-ratio construction therefore does not describe the average behavior of this suite.

For the completion-cost witness follow-up, common-stopping mean costs are:

Method	Targeted warm OR	Random cold
APE rectangular adaptation	54.7813	446.7500
Full certificate replanning	36.5104	416.4000
Margin witness rule	35.6823	449.5500
Completion-cost witness rule	37.4844	449.5500

There are 192 targeted tasks per method and 40 random tasks per method. The cost-based witness rule repairs a fixed-state margin-ranking failure but does not improve mean cost over the margin rule here. These staged findings explain the subsequent information and competing-plan experiments. The archive preserves protocols and outcomes for favorable and unfavorable settings alike.

H REPRODUCTION AND BASELINE PROVENANCE

The Python environment is 3.12 with NumPy 2.3.5, SciPy 1.17.0, pandas 2.2.3, and matplotlib 3.10.8. Experiments use CPU execution with four worker processes and one BLAS thread per worker. Paths resolve from the project directory. No GPU, model API key, or paid service is needed for the supplied experiments.

The principal commands, run from the project root, are:

```
python code/check_information.py --reference PATH_TO_PSIPS
python code/check_reversal_v14.py
python code/run_information_v14.py --native-binary PATH_TO_RUNNER
python code/run_competition_v14.py --native-binary PATH_TO_RUNNER
```

```
python code/analyze_information_v14.py
python code/analyze_competition_v14.py
```

The run scripts refuse to overwrite existing stage protocols. To reproduce a full stage, use a fresh copy or move the corresponding output directory. The independent source oracle is used only for cross-checking; it is not required by the new allocation implementation.

The PSIPS source URL is `https://github.com/cyrille-kone/psips`, pinned to commit `0b78a2fb7c30a997eb6c76b0ff0290274856e681`. The separate build script fetches the source and fixed header dependencies, then compiles a replay/counting harness. The original policy source is not redistributed in the supplementary archive. The replay environment supplies independent per-arm Gaussian streams; it changes random-number coupling relative to the original environment, not the prescribed observation law. We do not claim bitwise parity with the authors' original environment. Correctness is evaluated against the population front only after stopping.

Uncertainty intervals use pointwise seed-block bootstraps with twelve blocks per stage. Zero observed errors describe the recorded runs; the statistical error guarantee follows from the mixture proof for common-stopping policies. Native PSIPS retains its empirical calibration.

The added symmetry diagnostics and history-stratum analysis are run with

```
python code/check_symmetry_v15.py
python code/analyze_history_strata_v15.py
```

The latter reuses all warm trajectories in the two existing stages, reports both history layouts, and retains all pointwise bootstrap intervals. It is an exploratory analysis of existing observations, not an additional independently seeded replication.