ACTIVE DISCOVERY OF PREFERENCE CYCLES WITH ANYTIME-VALID INFERENCE

Anonymous authors
Paper under double-blind review

ABSTRACT

Preference-based A/B comparisons guide model updates, yet a sequence of locally winning updates can conceal declining underlying strength. Preference cycles expose matchup advantages that no single ranking preserves, but noisy comparisons require evidence for every direction. Uniform sampling can leave a promising cycle waiting on one unresolved pair. We develop an active procedure that targets cycles by estimated remaining evidence cost and shares observations across overlapping and later-proposed candidates. Under stationary independent Bernoulli streams, confidence sequences control the probability of any false report throughout the adaptive search. A fixed-order analysis charges the first positive cycle and the evidence needed to refute earlier candidates, counting shared pairs once. With identical confidence bounds and all initial labels charged, targeting reduces mean first-report cost by 15%-88% across four cyclic environments relative to a representative adaptive transitivity-testing method that we adapt to continued discovery. Updating priorities concentrates queries on cycles that can be completed, while shared evidence supports subsequent discoveries within the same budget.

1 INTRODUCTION

Preference-based A/B comparisons help decide whether a new model should replace the current version. Figure 1 illustrates a possible failure: a favorable style change offsets a loss of strength at every update, while average performance against a fixed reference population declines.

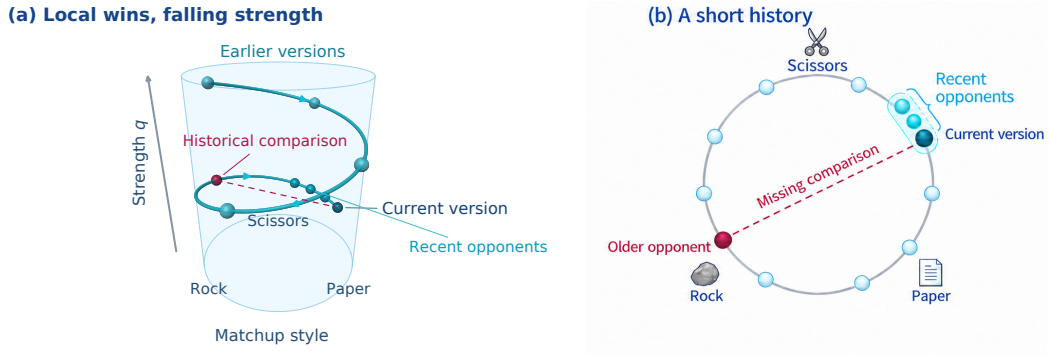

Figure 1: Schematic of locally winning updates that lose strength. (a) Style changes offset strength losses against the predecessor. Height is scalar strength; angle is matchup style. (b) A short history leaves older matchups unobserved (dashed links). Appendix D.1 specifies this constructed comparison model.

Published Go results exhibit cyclic preferences: Balduzzi et al. (2018) report that AlphaGo's value-network variant beats its policy-network variant, the policy-network variant beats the Go program Zen, and Zen beats the value-network variant. These preferences form a cycle that no single ranking can preserve. Nontransitive language-model judgments similarly make rankings depend on the baseline (Xu et al., 2025).

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.

In self-play, Hernandez et al. (2022) found that policies trained against the most recent half of their history tended to beat opponents inside that window but could lose to older policies just outside it, in repeated rock-paper-scissors and Connect Four. Sampling the full history improved temporal progress. Incomplete populations can hide cyclic interactions (Balduzzi et al., 2019); covering relevant styles is important to population methods (Czarnecki et al., 2020).

We seek cyclic preferences among saved versions under a fixed task and judging population: report a first cycle quickly, then reveal more within the remaining budget. Exhaustive comparison spends labels on pairs that may not help complete a report. A promising cycle can instead be one uncertain direction away from confirmation, and an observation on a shared pair can advance several candidates. Active discovery must decide which incomplete cycle is worth pursuing and which comparison will supply its missing evidence.

Our procedure keeps one confidence sequence per physical pair in a shared record, reusing observations across candidates. Reports require all directed lower bounds to exceed a prespecified margin above one half. An empirical guide estimates each direction's remaining evidence cost; we select the lowest-total-cost candidate and sample its costliest unresolved direction, revising priorities as evidence accumulates. Simultaneous coverage keeps reporting valid as the guide, candidate set and query priorities change.

The analysis distinguishes first discovery from family completion. For an order fixed by preliminary observations, it charges each failed candidate only through its first negative direction, counting shared pairs once. Under common confidence bounds and fully charged initial sampling costs, the experiments show lower first-report costs than the query rule of Haddenhorst et al. (2021) adapted to continued discovery. Shared-information comparisons show that revising priorities can substantially increase later output even when first-report costs are nearly identical.

Section 2 positions the work, and Section 3 defines the discovery problem. Sections 4–5 develop the reporting guarantee, query policy and cost analysis; Section 6 evaluates first discovery and continued reporting.

2 Related work

Ranking and cyclic preferences. Paired-comparison ranking (Bradley & Terry, 1952) summarizes systems by scalar scores, while ranking decompositions separate ordered and cyclic components (Jiang et al., 2011). Interaction models represent cyclic matchups (Chen & Joachims, 2016; Balduzzi et al., 2018), including richer preference models for language-model alignment (Zhang et al., 2025; Huang et al., 2026). Prediction-assisted evaluation estimates ranking scores and guides data collection (Frauen et al., 2026). These models can guide queries; reports require evidence for each directed population preference.

Sequential preference queries and transitivity testing. Dueling-bandit methods include finding tournament winners (Zoghi et al., 2015; Ramamohan et al., 2016). For nontransitivity, Haddenhorst et al. (2021) resolve majority directions until a cycle appears or all graph completions are transitive. Their query rule π^* prunes pairs whose endpoints have the same known relation to every other vertex, then samples unresolved pairs uniformly. We compare its query rule under our continued-discovery task (Section 6.1) and, separately, its original global test with its stopping decision and known-gap input (Section 6.3).

Anytime inference and discovery costs. Confidence sequences give bounds valid across observation times (Darling & Robbins, 1967; Howard et al., 2021); we use simultaneous edge coverage for overlapping and later-proposed cycles. Multiple-answer exploration distinguishes finding any acceptable answer from producing prescribed answers (Degenne & Koolen, 2019). Our cost analysis makes this distinction for cycles, using change of measure for prescribed discoveries (Kaufmann et al., 2016).

Our contribution is to make cycle discovery an ongoing, adaptive search with shared statistical evidence. Candidate-directed sampling pursues reports that are close to completion, while simultaneous coverage allows overlapping and later-proposed cycles to reuse earlier observations. For a fixed

candidate order, the cost analysis identifies how failed candidate prefixes and shared pairs determine the work before a first discovery.

3 THE SEQUENTIAL DISCOVERY TASK

A pure scalar-strength model provides a useful reference. In the Bradley-Terry model, win probability increases with the difference in two systems' strengths (Bradley & Terry, 1952): if A is stronger than B and B than C, their population majority preferences follow the same order. Matchup interactions can favor different opponents in different pairings, producing cycles even when systems also have scalar strengths (Chen & Joachims, 2016; Balduzzi et al., 2018). The strength-style model in Appendix D.1 illustrates this distinction. We formulate discovery directly in terms of pairwise win probabilities.

For n fixed systems, a candidate is a directed simple cycle: it visits distinct systems before returning to its start. We seek cycles whose every direction has population win probability above a prespecified threshold. Let $\mathcal{E}$ be the unordered pairs, called *physical edges*. Store each edge e once in a fixed orientation, with win probability $p_e \in (0, 1)$. Reversal exchanges wins and losses. A directed edge $a = (e, s)$, with $s \in \{+1, -1\}$, has probability

$$\mu_a = \begin{cases} p_e, & s = +1, \\ 1 - p_e, & s = -1, \end{cases} \qquad \theta_0 = \tfrac{1}{2} + \varepsilon, \quad 0 \le \varepsilon < \tfrac{1}{2}.$$

A candidate is *positive* when $\mu_a > \theta_0$ for every a in the cycle. The margin ε is fixed before assessing discovery cost. Write $p = (p_e)_{e \in \mathcal{E}}$ for the environment and $\mathcal{R}_t$ for the cycles reported through round t. We require

$$\sup_p \mathbb{P}_p\{\exists t \ge 1,\ C \in \mathcal{R}_t : \min_{a \in C} \mu_a \le \theta_0\} \le \alpha, \qquad 0 < \alpha < 1/2. \tag{1}$$

The error event is any incorrect report anywhere in a run. As an illustration, simultaneous lower bounds $(0.61, 0.59, 0.57)$ on $A \to B$, $B \to C$, $C \to A$ support a cycle at $\theta_0 = 0.55$, with minimum advantage at least 0.07 above one half. If the first two directions remain supported while the third has interval $[0.52, 0.80]$, the candidate is unresolved: all three upper bounds exceed θ_0, so completing the cycle remains possible, but the third lower bound has yet to cross the reporting threshold. An upper bound at most θ_0 on any direction instead refutes the candidate. A budget stop preserves these three evidence states.

The comparison population fixes model versions, task distribution, response sampling, judge mixture and display-order protocol. Evaluator inconsistency makes this specification consequential (Wang et al., 2026). Fresh task-response-judge units are drawn independently for each pair; an independent fair coin can resolve ties, contributing one half to win probability on tied judgments. Reusing tasks or responses across pairs can introduce dependence outside the model below. Appendix D gives the application protocol.

Initial observations fit a predictor before confirmation begins; $\mathcal{F}_0$ contains these data, the predictor and independent randomization.

Assumption 3.1 (Stationary comparison streams)**.** Conditionally on $\mathcal{F}_0$, each physical edge has an independent infinite i.i.d. Bernoulli(p_e) confirmation stream. A query reveals the next unobserved outcomes. Edge, batch size and any new error allocation are chosen before those outcomes are observed. Initial data remain separate from these confirmation streams.

Fresh confirmation gives a guarantee conditional on the realized initial sample, allowing the guide and edge errors to depend on it. Alternatively, registering all edge processes and errors before collecting any labels permits reuse of initial observations in those same streams, with an unconditional guarantee from the start (Appendix E.3).

4 INFERENCE FROM SHARED SEQUENTIAL EVIDENCE

To reuse observations across adaptively chosen cycles, we attach a confidence sequence to each physical pair: an interval covering its probability simultaneously at every sample size. A shared

registry stores each pair's observations, bounds and error. Registering an edge assigns its error before its first confirmation outcome. For the jth new edge, choose δ_j from the available history, with $\sum_j \delta_j \le \alpha$ almost surely. The experiments fix $\delta_e = \alpha/R$ on a known graph of R edges. For edges introduced during the search, one possible allocation is

$$\delta_j = \frac{6\alpha}{\pi^2 j^2}. \tag{2}$$

Revisiting a pair retains its observations and error allocation, so one update can advance every candidate containing it. Registration-rank spending makes an edge's error depend on when it enters; fixed allocations keep this statistical choice constant across the query-policy comparisons. Appendix E.1 describes allocation costs.

At round t, let $[L_e(t), U_e(t)]$ be the intersection of all confidence intervals obtained for that edge. Initially it is $[0, 1]$; reversal maps it to $[1 - U_e(t), 1 - L_e(t)]$. Write $[L_a(t), U_a(t)]$ for the interval in direction a. A direction is unresolved when $L_a \le \theta_0 < U_a$. Report a candidate once all its lower bounds exceed θ_0; exclude it from further sampling once any upper bound is at most θ_0. With nonempty intersections, every remaining candidate has an unresolved direction. Directions whose lower bounds exceed θ_0 form the supported graph.

Allocating Hoeffding's fixed-time error bounds over all sample sizes yields a familiar time-uniform confidence sequence (Appendix F). We use the Jeffreys beta-mixture sequence (Howard et al., 2021), which requires fewer labels in the fixed-path comparison of Appendix C.5. For m outcomes on one edge with S_m wins, a symmetric beta mixing distribution with shape $\beta > 0$ gives

$$M_m(p) = \frac{\mathrm{B}(S_m + \beta, m - S_m + \beta)}{\mathrm{B}(\beta, \beta) p^{S_m} (1 - p)^{m - S_m}}, \qquad M_0(p) = 1,$$

where B is the beta function. Under Bernoulli(p), this is a nonnegative mean-one martingale. Ville's inequality gives the raw interval

$$\mathcal{I}_m^{\mathrm{mix}} = \{p \in [0, 1] : M_m(p) \le 1/\delta\}. \tag{3}$$

We set $\beta = 1/2$. Both constructions cover each edge uniformly over time. The next result gives simultaneous coverage through adaptive queries and proposals; its general form applies to any conditionally valid confidence sequence.

Theorem 4.1 (Simultaneous validity of the shared registry)**.** *Under Assumption 3.1, either confidence-sequence construction with the reporting rule satisfies* (1)*, conditionally on the initial data. More generally, this holds whenever each process is valid conditional on its registration information and the predictable allocations satisfy $\sum_j \delta_j \le \alpha$ almost surely. Candidate proposals, comparison choices, and report times can be arbitrary functions of previously observed information.*

Proof sketch. Conditionally on registration, each edge process fails with probability at most its assigned δ_j. The tower property and a union bound control the event that any registered process ever fails by α. On its complement, all directed lower bounds lie below the true probabilities at every time, so every reported conjunction is correct. Appendix G gives the filtration argument. □

The common coverage event permits candidate proposals after inspecting data, priority updates after each batch, and reports whenever the evidence suffices. Existing pairs retain their observations and error allocations; new pairs receive predictable allocations within the same total budget. Algorithm 1 combines this registry with adaptive targeting and suspends reporting if an intersection becomes empty.

Each report can quantify its weakest advantage, $\gamma_C = \min_{a \in C} \mu_a - 1/2$, with interval $[\min_{a \in C} L_a(t) - 1/2, \min_{a \in C} U_a(t) - 1/2]$. Taking minima preserves the simultaneous bounds, so these intervals accompany adaptive reports under the same error budget. Appendix E.5 gives this result and the corresponding log-odds summary.

5 Targeting cycles and understanding discovery cost

A query can complete a cycle, refute an expensive candidate or advance several candidates through a shared pair. We give the adaptive rule, then use a initially fixed order to isolate failed candidates' contribution to first-discovery cost.

5.1 ADAPTIVE TARGETING

The experiments store both orientations of every triangle. The policy estimates what remains to support each direction, then combines these estimates at the candidate level. For W_e^0 wins in n_e^0 initial observations, the empirical guide is $\pi_e = (W_e^0 + 1/2)/(n_e^0 + 1)$, giving $1/2$ on an unobserved pair. After N_e confirmation labels and S_e wins, blend the guide with the observed frequency. Let $\widetilde{\mu}_a$ be $\widetilde{p}_e$ in the stored direction and $1 - \widetilde{p}_e$ in reverse, and define

$$\widetilde{p}_e = \frac{w\pi_e + S_e}{w + N_e}, \qquad \widetilde{d}_a = \max\{\widetilde{\mu}_a - \theta_0, d_{\min}\}, \qquad c_a = \left\lceil \frac{2\log(c_0 + N_e)}{\widetilde{d}_a^2} - N_e \right\rceil_+ . \quad (4)$$

Here $[x]_+ = \max(x, 0)$, and supported directions have score zero. The inverse squared predicted gap approximates the evidence needed to establish a direction; subtracting the observed count estimates what remains. The weight w controls the persistence of initial guidance, the positive floor $d_{\min}$ keeps the inverse-gap term finite, and c_0 sets the initial logarithmic scale. We fix $w = 20$, $d_{\min} = 0.015$ and $c_0 = 200$ throughout the sampling comparisons.

The candidate sum favors an inexpensive complete report; its largest unresolved score identifies the next bottleneck. A queried direction can also refute the candidate, removing work on its other directions. Shared observations remain useful when another cycle becomes more promising.

Algorithm 1 starts each new physical-edge record at $N_e = S_e = 0$ and $[L_e, U_e] = [0, 1]$. Later proposals retain existing counts, bounds and errors. The total label count z charges the initial sample once; initial outcomes do not enter the confirmation counts.

Algorithm 1 Empirical targeting with shared evidence and continuing reports

Require: Initial candidates $\mathcal{C}$ and proposer; initial guide (π_e, n_e^0); threshold θ_0; error allocation with $\sum_e \delta_e \le \alpha$; batch b, total budget $B \ge \sum_e n_e^0$

$z \leftarrow \sum_e n_e^0$, $\mathcal{R} \leftarrow \varnothing$, failure $\leftarrow$ false; initialize edge records
while true **do**
  if $z < B$ **then** accept available proposals into $\mathcal{C}$, sharing existing records
  Report unreported $C \in \mathcal{C}$ with $\min_{a \in C} L_a > \theta_0$; add them to $\mathcal{R}$
  Refute C with $\min_{a \in C} U_a \le \theta_0$; let $\mathcal{U}$ be the unresolved candidates
  if $z = B$ or ($\mathcal{U} = \varnothing$ and proposer finished) **then break**
  if $\mathcal{U} = \varnothing$ **then** request the next proposal and **continue**
  Recompute c_a by (4); $C^* \leftarrow \arg\min_{C \in \mathcal{U}} \sum_{a \in C} c_a$
  $a^* = (e, s) \leftarrow \arg\max_{a \in C^* : L_a \le \theta_0 < U_a} c_a$
  If this maximum is zero, prefer smallest N_e; all remaining ties use fixed index order
  If e is unregistered, assign its error δ_e predictably; otherwise keep its allocation
  $m \leftarrow \min\{b, B - z\}$; query m fresh outcomes $Y_1, \ldots, Y_m$ in stored orientation
  $N_e \leftarrow N_e + m$, $S_e \leftarrow S_e + \sum_{k=1}^m Y_k$, $z \leftarrow z + m$
  $[L_e, U_e] \leftarrow [L_e, U_e] \cap \mathcal{I}_{N_e}^{\mathrm{mix}}$ from (3), using S_e, δ_e
  if $L_e > U_e$ **then** failure $\leftarrow$ true; suspend reporting and **break**
end while
return $\mathcal{R}$, supported graph, unresolved candidates $\mathcal{U}$, failure flag

We evaluate targeting against uniform sampling of unresolved pairs and Greedy-margin, which selects a candidate by its minimum predicted advantage. Appendix A defines these controls and the implementation.

5.2 WHAT A FIXED PRIORITY ORDER SPENDS BEFORE ITS FIRST REPORT

What evidence can an unsuccessful candidate consume before the search reaches a positive one? To isolate this cost, fix candidate and within-candidate direction orders using the initial sample. *Priority auditing* follows this order, resolving directions until the current candidate is refuted or supported and reusing settled shared directions. Write $U(F)$ for the physical edges in an ordered direction sequence F.

The confidence boundary determines how many observations settle a direction. For a construction B, a *resolution profile* $h_{\mathsf{B}}(d, \delta)$ is a finite positive integer, nonincreasing in the gap d, such that every raw

interval has diameter strictly less than d at every count $m \geq h_{\mathsf{B}}(d, \delta)$, for every binary data prefix. On the coverage event this settles any direction at distance at least d from θ_0. Both Hoeffding and the Jeffreys mixture admit the conservative profile

$$h(d, \delta) = \left\lceil 16 d^{-2} \log \frac{16}{\delta d^2} \right\rceil . \tag{5}$$

Appendix H proves these profiles. Before the first positive candidate, a failed candidate needs at most the prefix ending at its first true negative direction. This separates the pairs selected by the order from the evidence required on each pair.

Theorem 5.1 (Evidence prefixes before first discovery)**.** *Fix a finite candidate list and its candidate and direction orders given the initial sample, positive edge errors and a conditionally valid construction* B *with resolution profile* h_{B} *and batches of size at most* b*. Start with zero confirmation counts and run until the first report without an earlier budget stop. Suppose a positive candidate exists and every relevant directional gap is nonzero. Let* C_{j_*} *be the first positive candidate in that order. For each* $j < j_*$ *let* F_j *be the ordered prefix of* C_j *ending at its first true negative direction, and put* $F_{j_*} = C_{j_*}$*. Let* $E_* = \bigcup_{j \leq j_*} U(F_j)$ *and* $d_e^* = \min_{a \in \cup_{j \leq j_*} F_j : a = (e,s)} |\mu_a - \theta_0|$*. On the simultaneous coverage event, priority auditing reports a first cycle within*

$$T_{\mathrm{first}} \leq \sum_{e \in E_*} \{ h_{\mathsf{B}}(d_e^*, \delta_e) + b - 1 \}, \tag{6}$$

where T_{first} *counts confirmation labels.*

A favorable order reaches a positive cycle early or quickly exposes negative directions in earlier candidates. Failed prefixes enlarge E_*, with each pair charged once at its smallest relevant gap; when the first candidate is positive, only its edges enter. These sets describe the environment in the bound, not information given to the policy. The theorem concerns the initially fixed order (proof in Appendix I); adaptive reordering is evaluated in Section 6.

5.3 Family completion and prescribed discoveries

We next give sufficient evidence costs for resolving a finite candidate family and necessary evidence costs for prescribed reports. For a fixed family $\mathcal{C}$, let $U(\mathcal{C})$ be its distinct physical edges and set

$$d_e = \min_{a \in \cup_{C \in \mathcal{C}} C :\, a = (e,s)} |\mu_a - \theta_0| > 0.$$

Theorem 5.2 (Resolution of a finite candidate family)**.** *Use a conditionally valid construction* B *with resolution profile* h_{B} *and batches of size at most* b*. Start with zero confirmation counts and query only this fixed family. Suppose the sampling rule always queries an unresolved directed edge in an unresolved candidate and continues until all candidates are resolved. On the simultaneous coverage event, the number of confirmation labels is at most*

$$\sum_{e \in U(\mathcal{C})} \{ h_{\mathsf{B}}(d_e, \delta_e) + b - 1 \}. \tag{7}$$

Unqueried edges can receive unused registration ranks when evaluating this bound.

The bound charges each physical edge once at the precision needed for its hardest relevant direction. Positive gaps ensure termination; a direction at the threshold may remain unresolved. Appendix H proves the per-edge sample bounds and termination. For $R = |U(\mathcal{C})|$, profile (5) gives the sufficient cost

$$O\left(\sum_{e \in U(\mathcal{C})} d_e^{-2} \{ \log(R/\alpha) + \log(1/d_e) \} + Rb \right). \tag{8}$$

Here the errors are either fixed at α/R or assigned by (2) with no outside registrations. For a family formed after other registrations, use its actual errors in (7).

Standard polynomial stitching (Howard et al., 2021) yields a resolution profile of order $d^{-2}\{\log(1/\delta) + \log\log(\mathrm{e}/d)\}$. It replaces $\log(1/d_e)$ in (8) by $\log\log(\mathrm{e}/d_e)$ under the same allocations, and also sharpens the prefix sum. Appendix F.4 gives the boundary and Appendix H.1 its

profile. Candidate order determines which pairs incur cost; the boundary determines their resolution cost.

Prescribing a positive family requires evidence on every supporting pair: changing its required direction to θ_0 invalidates a specified cycle. Under unrestricted edge probabilities and unchanged non-label side information, Theorem J.1 gives an expected-cost lower bound that includes informative initial labels (Appendix J; Kaufmann et al., 2016). Near an interior threshold, each required pair contributes an inverse squared gap. Finding any acceptable cycle can choose an easier answer; resolving a family also spends evidence on negative candidates. The upper bounds above hold on the coverage event and count fresh confirmation labels.

Shared evidence can support far more cycles than pairs: three cyclic groups of g systems use $3g^2$ pairs to support g^3 triangles. The supported graph represents them without enumeration. Depth-first search finds a cycle or establishes acyclicity of this graph in $O(n + R)$ time and storage (Appendix E.2); its acyclicity does not establish population transitivity. Explicit candidate scoring has a separate computational cost (Appendix A).

6 Numerical experiments

The synthetic matrices have known population truth. We measure labels to the first true report (lower is better) and true reports by a budget (higher is better). Direct and independent-matrix studies allow 12000 *total* labels, including each method's actual initial sample. First cost is capped at 12000 without a true report; success and actual stopping cost are recorded separately. Paired policies reveal prefixes of common per-edge Bernoulli streams. Targeting's independent initial sample guides queries but supplies no confirmation evidence.

6.1 First discovery under matched confidence bounds

This comparison asks which complete sampling policy discovers cycles more efficiently under the same evidence requirements. We adapt π^* from Haddenhorst et al. (2021) by retaining its uniform unresolved-pair queries and published graph pruning, replacing its original tests with our confidence sequences, and continuing to report supported triangles. Both policies use Jeffreys mixtures, directional threshold $\theta_0 = 0.55$, global error $\alpha = 0.05$, fixed edge errors α/R and batches of at most 25. Targeting uses the same score constants throughout and pays 720 initial labels at nine objects or 294 at 18; adapted π^* has no initial sample.

Four cyclic matrices have 60 paired repetitions each. Sparse rock-paper-scissors (RPS) embeds one strong triangle in a scalar preference background; Block RPS has three cyclic groups and 27 positive triangles. Both have nine objects. Low-rank has 18 objects with scalar strengths and a rank-two cyclic interaction, yielding 216 positive triangles. Ordered-cycle has nine objects and a moderately strong triangle in an otherwise strongly ordered population, giving pruning an opportunity to remove pairs while that triangle remains unresolved. A fifth matrix is transitive. Appendix B specifies the matrices and seeds.

Table 1: First-discovery cost under matched confidence bounds: $\theta_0 = 0.55$, identical mixture boundaries and edge errors, 60 paired repetitions per row. Costs are mean total labels, including initial labels; lower is better. Saving is adapted π^* minus targeting, with a 95% paired Monte Carlo interval. All policies find a cycle in all 60 repetitions.

Setting	Targeting	Adapted π^*	Saving [95% interval]
Sparse RPS (9)	957	4099	3142 [2733, 3552]
Block RPS (9)	1106	2549	1443 [1304, 1581]
Low-rank (18)	553	4682	4129 [3764, 4495]
Ordered-cycle (9)	2946	3470	524 [389, 659]

Targeting reduces mean first-report cost by 15%-88% across the four cyclic matrices; every paired saving interval is positive (Table 1). The largest reduction is in Low-rank, where many unresolved pairs compete with an inexpensive cycle. In Ordered-cycle, pruning activates in 26 of 60 adapted-π^*

runs: mean first cost is 3470 with pruning, 3488 without it and 2946 for targeting. Pruning removes irrelevant pairs, while candidate priority concentrates evidence on a report that can be completed.

On the transitive control, adapted π^* instead ends the search sooner: mean actual stopping costs are 2769 versus 4861 for targeting, with no reports. Here stopping settles the margin-positive triangle search. It is different from the original method's global transitivity decision below.

6.2 Continued discovery and priority updates

The first report leaves evidence that later candidates can reuse. At the total budget in Low-rank, targeting reports all 216 true triangles on average, compared with 53.4 for adapted π^* (Figure 2a). These triangles overlap in vertices and pairs: their count measures supported cyclic structure, not independent discoveries.

With shared initial observations, fixed priority and adaptive targeting have similar first costs in 18-object Low-rank, 563.4 and 572.8 labels, but report 120.9 and 216.0 triangles. Under misspecification, their counts are 11.4 and 46.4. An initially useful order can stall on a difficult later candidate; adaptation redirects batches toward cheaper reports. In Block RPS, fixed priority instead finds the first cycle sooner, at 1080.9 versus 1106.6 labels, and both report all 27 triangles. This separate study uses 12000 *confirmation* labels at 9 and 18 objects, adding the common initial sample to total cost (Appendix C.2).

Targeting also saves labels against uniform-unresolved sampling across five cyclic settings with 9-30 objects. Greedy-margin captures much of the early gain; under misspecification, however, it reports 13.7 triangles versus 46.4 despite nearly equal first costs. A further shared-information comparison on Misspecified matrix 1 uses a 12000-total-label budget. Targeting yields 52.7 [48.4,57.0] more reports than Greedy-margin and 50.0 [43.7,56.3] more than fixed priority (95% paired intervals; Appendix C.3), isolating continued-output gains beyond choosing an initially favorable candidate.

To examine variation across environments, we generate four independent matrices from each of the 18-object Low-rank and Misspecified families. Low-rank varies phases of the cyclic interaction; Misspecified additionally perturbs each edge logit. Each matrix has 20 paired label repetitions, independent of the fixed-matrix comparison. Confidence bounds, score constants and the 12000-total-label budget remain unchanged; targeting pays its 294-label initial sample and adapted π^* has none (Appendix B.4).

Table 2: Independent matrices, 20 paired repetitions per row. T is empirical targeting and π^* is the continued-discovery adaptation, with the same confidence bounds and reporting rule. First cost is mean total labels, including initial labels; lower is better. Saving is π^* minus T with a within-matrix 95% Monte Carlo interval. Reports are mean true oriented triangles by total budget 12000; higher is better. Every policy finds a first cycle in all 20 runs. Means and paired savings are rounded separately.

	First cost (labels)			Reports	
Matrix	T	π^*	Saving [95% interval]	T	π^*
Low-rank 1	562	5019	4457 [3965, 4949]	216.0	53.5
Low-rank 2	552	4571	4020 [3385, 4655]	216.0	57.0
Low-rank 3	537	5045	4509 [4018, 4999]	216.0	50.4
Low-rank 4	579	4774	4195 [3746, 4644]	212.7	50.5
Misspecified 1	478	2941	2464 [2065, 2862]	66.9	33.9
Misspecified 2	502	2566	2065 [1528, 2601]	64.9	46.0
Misspecified 3	565	3199	2634 [2225, 3042]	49.1	36.3
Misspecified 4	390	2974	2584 [2132, 3035]	68.4	37.1

Mean first-report cost falls on every matrix: by 88-89% in Low-rank and 80-87% under misspecification (Table 2). All eight paired saving intervals are positive. Targeting also produces more mean reports: 213-216 of 216 true triangles in Low-rank versus 50-57 for adapted π^*, and 49-68 versus 34-46 under misspecification, where the populations contain 161-184 true triangles. All eight paired output-gain intervals are positive (Figure 2b). The intervals describe label randomness within each matrix, while rows vary the environment. Joint relabelings of vertices, initial datasets and

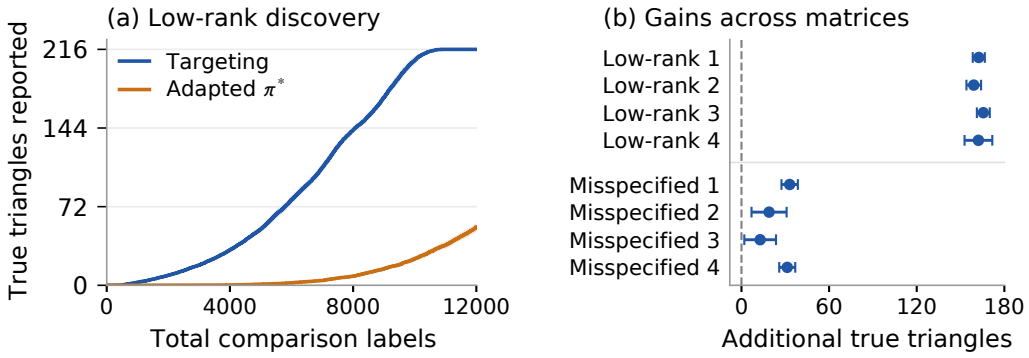

Figure 2: Continued discovery under matched confidence bounds. (a) Mean curves in 18-object Low-rank, 60 paired runs, with pointwise 95% Monte Carlo intervals. (b) Targeting minus adapted π^* output at budget 12000 on eight independent matrices, with within-matrix 95% intervals from 20 paired runs each. Costs include each policy's initial sample; output counts distinct true triangles that may share edges.

outcomes retain the continued-output gains over shared-information controls in the reported check (Appendix C.3).

6.3 The original global test with its gap input

The original test answers a different question: stop on a majority cycle of any length or accept global transitivity. We retain Algorithm 2 of Haddenhorst et al. (2021), its π^* query rule and sequential probability ratio test (SPRT), with both global error levels set to 0.05. The matrices are separated to have majority gaps at least 0.025. Targeting uses $\theta_0 = 0.5$ on those same matrices, still seeking triangles without a gap input or a global acceptance decision. The original test receives either fixed $h_{\mathrm{HB}} = 0.02$ or oracle $0.99 \min_e |p_e - 0.5|$. Its initial observation on every pair is charged (Appendix B).

With the fixed gap, targeting has lower first cost in all four cyclic settings. With the oracle gap, the original test remains slower in Sparse and Block RPS and is censored without detection in every Low-rank run, but wins in Ordered-cycle: 695 versus 1804 labels. Its large minimum majority gap makes the SPRT efficient. On the transitive control, the oracle test accepts transitivity in all 60 runs at mean cost 4048. These results concern the original terminal decision with a supplied separation bound; Table 1 concerns continued margin-positive discovery with common bounds.

Run-level calibration, a longer-cycle example and computational costs appear in Appendices C.4 and B.5.

7 Discussion

Candidate targeting allocates queries toward completing cycles; shared confidence sequences make reports valid across overlaps and later proposals. The fixed-order bound explains the cost of failed priorities, while the experiments measure adaptive targeting's first-discovery and continued-output benefits. A cycle and its weakest-margin interval identify matchup advantages omitted by a scalar ranking. For model evaluation, this output locates which saved versions defeat which others and supplies matchups for follow-up evaluation, while later candidates reuse the accumulated evidence.

Reducing the cubic search cost of enumerated triangles is a useful next step. Joining supported paths through unresolved edges could extend discovery to longer cycles within the same inference framework, with each new proposal checked immediately against the existing pairwise confidence bounds. Evaluation on fresh model judgments would connect discovery efficiency with model development, keeping model versions, tasks and judging protocols fixed within each run while candidate choices evolve.

AI USE STATEMENT

Generative AI tools were used to revise wording and improve the clarity and readability of the manuscript; to identify potentially relevant literature, references, and search directions, and to assist with research execution, including drafting, debugging, and refining code. For mathematical proofs, AI was used primarily to review and check the authors' arguments for potential errors, inconsistencies, or gaps. AI-assisted manuscript revisions, literature-search suggestions, feedback on research ideas and proofs, and generated or modified code were reviewed and verified by the authors before incorporation into the final work. The authors independently assessed the accuracy and appropriateness of these outputs and retain full responsibility for the methodology, proofs, implementation, analysis, and conclusions presented in the paper.

REPRODUCIBILITY STATEMENT

The code and an accompanying `README` file are submitted with the manuscript. Scripts and numerical records cover each experimental study, with separate entry points for the main figures and tables. The `README` provides instructions for regenerating and summarizing the results.

REFERENCES

David Balduzzi, Karl Tuyls, Julien Perolat, and Thore Graepel. Re-evaluating evaluation. In *Advances in Neural Information Processing Systems*, volume 31, 2018. URL https://arxiv.org/abs/1806.02643.

David Balduzzi, Marta Garnelo, Yoram Bachrach, Wojciech Czarnecki, Julien Perolat, Max Jaderberg, and Thore Graepel. Open-ended learning in symmetric zero-sum games. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 434–443, 2019. URL https://proceedings.mlr.press/v97/balduzzi19a.html.

Ralph Allan Bradley and Milton E. Terry. Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, 39(3/4):324–345, 1952. doi: 10.2307/2334029.

Shuo Chen and Thorsten Joachims. Modeling intransitivity in matchup and comparison data. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*, pp. 227–236, 2016.

Wojciech M. Czarnecki, Gauthier Gidel, Brendan Tracey, Karl Tuyls, Shayegan Omidshafiei, David Balduzzi, and Max Jaderberg. Real world games look like spinning tops. In *Advances in Neural Information Processing Systems*, volume 33, pp. 17443–17454, 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/hash/ca172e964907a97d5ebd876bfdd4adbd-Abstract.html.

D. A. Darling and Herbert Robbins. Confidence sequences for mean, variance, and median. *Proceedings of the National Academy of Sciences*, 58(1):66–68, 1967. doi: 10.1073/pnas.58.1.66.

Rémy Degenne and Wouter M. Koolen. Pure exploration with multiple correct answers. In *Advances in Neural Information Processing Systems*, volume 32, 2019. URL https://arxiv.org/abs/1902.03475.

Dennis Frauen, Athiya Deviyani, Mihaela van der Schaar, and Stefan Feuerriegel. Nonparametric LLM evaluation from preference data. *arXiv preprint arXiv:2601.21816*, 2026. URL https://arxiv.org/abs/2601.21816. DMLRank, version 2; accepted at ICML 2026.

Björn Haddenhorst, Viktor Bengs, and Eyke Hüllermeier. On testing transitivity in online preference learning. *Machine Learning*, 110:2063–2084, 2021. doi: 10.1007/s10994-021-06026-2.

Daniel Hernandez, Kevin Denamganai, Sam Devlin, Spyridon Samothrakis, and James Alfred Walker. A comparison of self-play algorithms under a generalized framework. *IEEE Transactions on Games*, 14(2):221–231, 2022. doi: 10.1109/TG.2021.3058898.

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.

Yucong Huang, Xiucheng Li, Kaiqi Zhao, and Jing Li. Transitivity meets cyclicity: Explicit preference decomposition for dynamic large language model alignment. *arXiv preprint arXiv:2605.17342*, 2026. URL https://arxiv.org/abs/2605.17342. Accepted at ICML 2026.

Xiaoye Jiang, Lek-Heng Lim, Yuan Yao, and Yinyu Ye. Statistical ranking and combinatorial Hodge theory. *Mathematical Programming*, 127:203–244, 2011. doi: 10.1007/s10107-010-0419-x.

Émilie 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.

Siddartha Ramamohan, Arun Rajkumar, and Shivani Agarwal. Dueling bandits: Beyond Condorcet winners to general tournament solutions. In *Advances in Neural Information Processing Systems*, volume 29, 2016. URL https://papers.nips.cc/paper/2016/file/fccb3cdc9acc14a6e70a12f74560c026-Paper.pdf.

Yidong Wang, Yunze Song, Tingyuan Zhu, Xuanwang Zhang, Zhuohao Yu, Hao Chen, Chiyu Song, Qiufeng Wang, Zhen Wu, Xin-Yu Dai, Yue Zhang, Cunxiang Wang, Wei Ye, and Shikun Zhang. TrustJudge: Inconsistencies of LLM-as-a-judge and how to alleviate them. In *International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/ecb87120e1e9c68ad186385e146094f4-Abstract-Conference.html.

Yi Xu, Laura Ruis, Tim Rocktäschel, and Robert Kirk. Investigating non-transitivity in LLM-as-a-judge. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 69583–69612, 2025. URL https://proceedings.mlr.press/v267/xu25w.html.

Yifan Zhang, Ge Zhang, Yue Wu, Kangping Xu, and Quanquan Gu. Beyond Bradley–Terry models: A general preference model for language model alignment. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 76939–76965, 2025. URL https://proceedings.mlr.press/v267/zhang25de.html.

Masrour Zoghi, Zohar Karnin, Shimon Whiteson, and Maarten de Rijke. Copeland dueling bandits. In *Advances in Neural Information Processing Systems*, volume 28, 2015. URL https://proceedings.neurips.cc/paper_files/paper/2015/hash/9872ed9fc22fc182d371c3e9ed316094-Abstract.html.

Implementation and experimental designs appear first (Appendices A-D), followed by additional consequences of the common coverage event. The confidence-sequence construction and all proofs are grouped thereafter (Appendices F-K).

A Implementation and sampling designs

A.1 Candidate representation and sampling rules

For each $i < j < k$, the forward triangle uses stored edges $(i, j), (j, k), (i, k)$ with signs $(+1, +1, -1)$, and the reverse negates those signs. Fixed index order resolves ties. Candidate targeting excludes supported and refuted candidates. All-edge uniform can query any physical edge. Graph-unresolved samples uniformly among edges whose interval intersects either θ_0 or $1 - \theta_0$; it does not prune an edge merely because all candidates containing it are already resolved. Uniform-unresolved samples uniformly over the *unique physical edges* of unresolved directions in unresolved candidates, rather than sampling a candidate-edge incidence. Greedy-margin selects the unresolved cycle with largest minimum blended predicted probability and then its unresolved direction with largest interval width. These controls use the same reporting rule.

The default empirical guide and remaining-cost score are given in Equation (4). The sampling comparisons change the query rule while keeping the evidence process fixed. Fixed priority orders candidates by the initial-data inverse-gap score and directions by increasing initial estimated win probability, with stable canonical ties. It advances after refutation, skips already resolved shared directions, and reports other supported cycles immediately. Adaptive targeting recomputes priorities after each batch. These rules distinguish an initially useful order from one that continues to adapt to the accumulated evidence.

A.2 Registry and proposal handling

Each edge record contains its fixed storage orientation, wins, own-edge count, registration rank, assigned error and nested endpoints. A new edge must be registered before its first update; repeated registration returns the original allocation. Confirmation never backfills earlier unregistered outcomes. A new candidate is immediately checked using all current evidence. A confidence intersection with lower endpoint above upper endpoint sets a failure flag and terminates further reporting. Batch size is capped by remaining budget. Output distinguishes reports, unresolved candidates, supported graph and failure status; a budget stop is not a statistical acceptance decision.

The general-cycle reporter stores explicit simple vertex traversals, canonically rotates them without reversing orientation, and maintains their physical-edge incidences. The primary sampling simulator enumerates triangles and rescans candidate scores at each batch. The generic registry theorem allows growing registration; the tested implementation has a prespecified physical edge universe and dynamically registers or proposes within it. An incidence index can reduce update work, but the reported runtime is for the actual dense candidate scan. With L candidate-edge incidences, storage is $O(R + L)$ and a full score scan costs $O(L)$ per batch. For triangles, $L = O(n^3)$.

A.3 Probability matrices and initial sampling designs

Write p_{ij} for the stored probability on pair (i, j) with $i < j$, and index vertices from zero. Write $\sigma(z) = (1 + e^{-z})^{-1}$ for the logistic function. The following models and initial sampling designs supply the additional sampling comparisons and the corresponding environments in the direct literature comparison.

The additional sampling experiment uses seed 59061000. Nine-object Sparse RPS starts with strengths equally spaced in $[-0.5, 0.5]$ and probabilities $p_{ij} = \sigma(s_i - s_j)$, then sets $p_{01} = p_{12} = 0.8$ and $p_{02} = 0.2$. Block RPS assigns groups by $i \bmod 3$, probability 0.75 along the group cycle $0 \to 1 \to 2 \to 0$ and 0.5 within groups. The Bradley-Terry (BT) null keeps the unmodified scalar logits on the same strength range. The first two have one and 27 positive oriented triangles, respectively.

For the settings with sparse initial observations, let $\phi_i = 2\pi(i \bmod 3)/3 + \xi_i$, $\xi_i \sim \mathcal{N}(0, 0.12^2)$, and strengths equally spaced in $[-0.2, 0.2]$. Write $\mathrm{logit}(p) = \log\{p/(1-p)\}$. Low-rank logits are

$$\mathrm{logit}\, p_{ij} = s_i - s_j + 1.7 \sin(\phi_j - \phi_i).$$

The Misspecified setting adds independent $\mathcal{N}(0, 1.3^2)$ edge perturbations to these logits. One matrix is generated per setting, with seed equal to the main seed plus its setting index, and stays fixed across repetitions. The fixed-matrix sampling study has six environments, including its null control; uncertainty describes repeated label sampling conditional on each matrix.

The dense initial design contains 20 labels on each of 36 edges. The sparse initial design takes a connected ring and the union with $2n$ physical edges selected without replacement using seed $705+n$. Duplicates with ring edges are retained only once. There are six labels on every selected edge and none on others, totaling 294 at $n = 18$ and 492 at $n = 30$. Unobserved pairs therefore receive the empirical guide's $1/2$ fallback.

For setting indices 0–5 in the order Sparse, Block, BT null, Low-rank (18), Misspecified (18), Low-rank (30), the confirmation-stream seed is $202609260000 + 100000 \cdot \mathrm{setting} + \mathrm{replicate}$. Initial counts use an independent seed offset 70000. Each replication generates independent initial counts and complete Bernoulli potential streams per edge. Policies reveal prefixes of these streams, so changing a batch size preserves individual outcome identity rather than changing a Binomial batch draw. Random queries use the confirmation-stream seed plus 444. Budgets are 12000 confirmation labels for $n = 9, 18$ and 18000 for $n = 30$. Every primary policy has errors $\alpha/\binom{n}{2}$ registered before its first fresh outcome. Initial observations are used for prediction; only fresh outcomes enter confirmation.

Each policy reveals prefixes of the same potential streams within a repetition. First-discovery cost is capped at the confirmation budget before initial sampling cost is added; success and final true-report counts are recorded separately. A missing report is not a declaration of transitivity. Paired Monte Carlo intervals are the mean within-repetition cost difference plus or minus 1.96 standard errors. Repetitions regenerate the initial sample and confirmation outcomes rather than the fixed matrix. The independent-matrix design in Appendix B.4 varies both sources separately.

B Direct comparison with the 2021 transitivity test

The comparisons use independent implementations of the empirical policy in (4) and Sections 7-8 of Haddenhorst et al. (2021), with 60 paired repetitions per fixed matrix. The matched-evidence comparison adapts the published query rule to continued triangle discovery at threshold 0.55. The global-test comparison retains the original tests and terminal decision on separated matrices at threshold 0.5, with fixed or oracle gap input. Empirical score constants stay unchanged. The two comparisons are reported separately below.

B.1 Matrices and total comparison cost

Write p_{ij} for the stored probability of edge (i, j), $i < j$. Sparse RPS and Block RPS use the nine-object formulas in Appendix A.3. Low-rank uses the 18-object formula there with matrix seed 59061003. Ordered-cycle sets every $p_{ij} = 0.85$ for $i < j$, except $p_{67} = p_{78} = 0.65$ and $p_{68} = 0.35$, with vertices indexed from zero. Its only positive triangle is $6 \to 7 \to 8 \to 6$; all other vertices have consistently ordered preferences. The transitive control sets every $p_{ij} = 0.8$ for $i < j$, except $p_{34} = 0.525$. This creates a difficult adjacent direction that need not be resolved to establish transitivity. The last two structures were chosen to expose the role of graph pruning, before their outcomes were generated.

For setting index $s \in \{0, 1, 2, 3, 4\}$ in the order Sparse, Block, Low-rank, Ordered-cycle and Transitive, repetition $r \in \{0, \ldots, 59\}$ uses seed $202609121500 + 100000s + r$. Each physical pair receives an independent Bernoulli potential stream; policies reveal its next unused observations. Initial counts use an independent seed offset 70000, and randomized queries use offset 444. Targeting collects 20 observations on every pair for nine objects, and six observations on the 49 pairs of the fixed sparse design for 18 objects. Initial outcomes guide targeting and are excluded from its confirmation sequences. The literature policies collect no initial sample. Every method pays for all

labels it uses within the same total budget 12000. The additional comparisons in Appendix C specify a confirmation budget before adding their shared initial sampling cost.

Each cell of Table 1 summarizes repeated label sampling at one fixed matrix. Approximate 95% Monte Carlo intervals are the mean within-repetition cost difference plus or minus 1.96 standard errors. For a run without a true discovery, first-discovery cost is capped at 12000 even if the method stops sooner; actual stopping cost and decisions are retained separately. A reported triangle is counted once per orientation, and distinct triangles may share edges or vertices. No false reports, erroneous global decisions, or empty confidence intersections occurred in the 1500 principal trajectories or the 300 oracle-gap trajectories.

B.2 Adapting the query rule under matched confidence bounds

Both policies use the Jeffreys beta-mixture confidence sequence at error $0.05/R$ per pair and the same 0.55 directional threshold. At each batch endpoint, the adapted π^* builds a majority-sign graph by placing $i \to j$ when the lower bound for p_{ij} exceeds 0.5, or the reverse arc when its upper bound is below 0.5. A pair $\{i, j\}$ is negligible if, for every other vertex k, the graph contains both $i \to k$ and $j \to k$, or both $k \to i$ and $k \to j$ (Definition 8.2 of Haddenhorst et al., 2021). Such a pair cannot belong to a positive triangle on the simultaneous coverage event. It is excluded from the uniform query pool, which otherwise contains pairs with an unresolved direction at the reporting threshold. A third policy uses the same pool without this graph pruning. All three continue reporting supported triangles until the budget or their stopping rule ends the run.

Table 1 gives the core first-discovery comparison. The pruning rule never activates in the first three cyclic settings, so its paths coincide with graph-unresolved sampling there. It activates in 26/60 Ordered-cycle runs and 36/60 transitive runs. Targeting wins the paired first-report comparison in 60/60 Sparse, 59/60 Block, 60/60 Low-rank and 54/60 Ordered-cycle runs. On the transitive control, mean actual stopping costs are 4861 for targeting, 2786 without pruning and 2769 for adapted π^*. No policy reports a triangle. These costs quantify ending the margin-positive triangle search, not accepting global transitivity.

B.3 Original global test and its known-gap input

For the original test, probabilities with $|p_e - 0.5| < 0.025$ are moved to 0.525 or 0.475 according to their original side; an exact tie takes the positive storage orientation. This creates a complete majority tournament with every gap strictly above $h_{\mathrm{HB}} = 0.02$. The original test starts with one observation on every physical pair, then queries one label at a time. With $R = \binom{n}{2}$ and both global error probabilities set to 0.05, its pairwise level and SPRT boundary are

$$\eta_{\mathrm{HB}} = \min\left\{\frac{0.05}{n}, \frac{0.05}{R - \lfloor (n+1)/3 \rfloor}\right\}, \qquad b_{\mathrm{HB}} = \left\lceil \frac{\log\{(1-\eta_{\mathrm{HB}})/\eta_{\mathrm{HB}}\}}{\log\{(0.5+h_{\mathrm{HB}})/(0.5-h_{\mathrm{HB}})\}} \right\rceil.$$

A pair resolves at its first strict crossing $|2S_m - m| > b_{\mathrm{HB}}$. The direction is the sign of $2S_m - m$. The lower boundary is $0.5 - b_{\mathrm{HB}}/(2m)$, using the symmetric SPRT construction in their Section 7. The active pool removes resolved and negligible pairs. The original test stops when its resolved graph contains a cycle of any length, or is acyclic with every remaining pair negligible, which establishes transitivity in expansion.

With $h_{\mathrm{HB}} = 0.02$, mean first-discovery costs for targeting and the original test are respectively 901/5446 in Sparse, 972/5271 in Block, and 1804/4754 in Ordered-cycle, with 60/60 successes for both throughout. In Low-rank they are 480 and a censored 12000: targeting succeeds 60/60 times and the original test 0/60. The original test correctly accepts the transitive control in every run at mean actual cost 5010, with pruning in every run. Pruning also occurs in 18/60 original Ordered-cycle runs.

The fixed bound is conservative when all gaps are large, especially in Ordered-cycle, whose minimum majority gap is 0.15. We therefore additionally provide the original test with $h_{\mathrm{HB}} = 0.99 \min_e |p_e - 0.5|$, determined from the matrix, on the same outcomes. The oracle comparison changes no targeting parameter. The bounds are approximately 0.03090, 0.02475, 0.02475, 0.14850 and 0.02475 in setting order. Table 3 reports these stronger original baselines. Their paired cost differences from targeting are 2724 [2621,2827], 3227 [3171,3283], and -1109 [-1268, -950] in Sparse, Block and Ordered-cycle. Low-rank remains fully censored for the original test. On the transitive control,

the oracle-gap test accepts in all runs at mean cost 4048. The reversal in Ordered-cycle shows the value of an informative separation bound for the original global decision. The matched-evidence comparison instead evaluates continued margin-positive triangle discovery, with different thresholds and statistical inputs; its policy ranking is a separate result.

Table 3: Original global test on separated matrices, 60 paired repetitions. Costs are mean total labels to first true cycle, including all initial observations; lower is better. Targeting uses $\theta_0 = 0.5$ without a gap input. The original test receives fixed $h_{\mathrm{HB}} = 0.02$ or oracle $h_{\mathrm{HB}} = 0.99 \min_e |p_e - 0.5|$. A dagger indicates censoring at 12000 with 0/60 successes; all other entries have 60/60.

Setting	Targeting	Original, fixed gap	Original, oracle gap
Sparse RPS (9)	901	5446	3625
Block RPS (9)	972	5271	4199
Low-rank (18)	480	$12000^{\dagger}$	$12000^{\dagger}$
Ordered-cycle (9)	1804	4754	695

B.4 Independent probability matrices

We examine environment variation using four independent matrices from each of the 18-object Low-rank and Misspecified families in Appendix A.3. Each matrix draws independent phase noise with standard deviation 0.12; Misspecified also draws independent edge-logit perturbations with standard deviation 1.3. Strengths are equally spaced in $[-0.2, 0.2]$. Sampling parameters are fixed at the values in Section 5 for all eight matrices.

For family index $f \in \{0, 1\}$ (Low-rank, Misspecified) and matrix index $m \in \{0, 1, 2, 3\}$, the matrix seed is $202609123000 + 100f + m$; Table 2 numbers these matrices $m + 1$. Repetition $r \in \{0, \ldots, 19\}$ uses label seed $202609124000 + 10000f + 100m + r$ and independent initial-data seed offset 70000. The sparse initial sampling design is the same 49-pair design as in the direct comparison. Empirical targeting pays its 294 initial labels; adapted π^* has no initial sample. Both reveal prefixes of the same per-edge potential streams and continue reporting under threshold 0.55, pairwise error $0.05/R$, batches at most 25, and total budget 12000. This gives 160 paired repetitions and 320 policy trajectories.

Table 2 reports each matrix separately. Within a matrix, the cost-saving interval is the mean paired difference plus or minus 1.96 standard errors over its 20 repetitions. Differences across rows describe environment variation; these intervals quantify label randomness conditional on the matrix. First cost is capped at 12000 if no true report occurs, and report counts include only distinct true oriented triangles.

The paired output differences (targeting minus adapted π^*) also quantify continued discovery. For Low-rank matrices 1-4, they are respectively 162.5 [158.4,166.6], 159.1 [154.0,164.1], 165.6 [161.2,170.0] and 162.3 [152.7,171.8] triangles. For Misspecified matrices 1-4, they are 33.0 [27.4,38.6], 18.9 [6.9,30.9], 12.8 [2.0,23.6] and 31.4 [25.9,36.8]. These 95% Monte Carlo intervals use the same 20 paired label repetitions per matrix as the first-cost intervals; the overlapping reported triangles are counted within a run, not treated as independent observations.

B.5 Computational cost

On one existing trajectory per independent matrix, median CPU time over three timing repetitions is 2.6-7.4 ms to targeting's first report versus 17.5-56.9 ms for adapted π^*. Full-budget times are 0.231-0.254 versus 0.134-0.162 seconds, excluding label generation and candidate construction. Targeting reaches a first report in fewer batches but spends more computation per batch. The timing paths also include a reversal in output: on one misspecified trajectory, targeting reports 23 triangles and adapted π^* reports 50.

Appendix C.5 compares confidence boundaries on fixed sampling paths.

C Additional sampling and calibration comparisons

C.1 Natural sampling controls

The additional comparison uses the fixed probability matrices and paired sampling design in Appendix A.3. All policies have beta-mixture intervals with $\beta = 1/2$, $\theta_0 = 0.55$, $\alpha = 0.05$, edge errors $\alpha/\binom{n}{2}$ and batches of at most 25. Each nine-object setting has 80 repetitions and a 720-label initial sample; each 18-object setting has 40 repetitions and a 294-label initial sample; the 30-object setting has 20 repetitions and a 492-label initial sample. The budgets are 12000 *confirmation* labels for 9 and 18 objects, and 18000 for 30 objects. Initial sampling cost is added to the reported total. These budgets differ from the 12000 *total*-label literature comparison and are not pooled with it.

Table 4: Sampling under matched confidence bounds. Entries are mean total labels to a first true triangle, including initial labels; lower is better. All displayed rows find a first cycle in every repetition. Uniform-unresolved samples unique physical pairs of unresolved candidate directions. Greedy-margin targets the candidate with largest minimum predicted probability.

Policy	Sparse (9)	Block (9)	Low-rank (18)	Misspec. (18)	Low-rank (30)
Uniform-unresolved	4942	3314	5123	2700	13447
Greedy-margin	961	1207	620	408	881
Empirical targeting	961	1107	573	405	817

Empirical targeting saves 3980 total labels relative to Uniform-unresolved in Sparse RPS (paired approximate 95% interval [3663,4297]) and 2208 [2064,2351] in Block RPS. Greedy-margin's Block-RPS cost is 100 [63,138] above targeting's. Under misspecification its first cost is 407.8 versus 404.6, while its mean continued output is 13.7 triangles versus 46.4. Thus selecting an apparently favorable candidate captures much of the early benefit, but accounting for remaining evidence can materially change later output. The BT-null control yields no reports for any policy; its capped total first cost is 12720.

C.2 Fixed priority and adaptive targeting

The fixed-priority and adaptive policies use the same mixture confidence bounds on 340 paired repetitions with common initial and confirmation data across the six environments. Table 5 contains the five cyclic environments. The fixed ordering rule is specified in Appendix A; both policies skip resolved shared directions. Repetition counts, initial datasets and confirmation budgets are those above.

Table 5: Fixed priority versus adaptive empirical targeting. Costs are mean total labels to a first report, including initial labels; lower is better. Reports are mean true triangles at the confirmation budget plus initial labels; higher is better. Both policies use the same mixture boundary and observations.

	First cost (labels)		Reports at budget	
Setting	Fixed	Adaptive	Fixed	Adaptive
9, Sparse RPS	964.7	961.3	1.0	1.0
9, Block RPS	1080.9	1106.6	27.0	27.0
18, Low-rank	563.4	572.8	120.9	216.0
18, Misspecified	438.4	404.6	11.4	46.4
30, Low-rank	799.5	817.0	142.9	438.2

At 18-object Low-rank the first costs differ by 9.4 labels, while mean output increases from 120.9 to 216.0 triangles. Under misspecification output increases from 11.4 to 46.4. Fixed priority instead has lower first cost in Block RPS, 1080.9 versus 1106.6, with identical output of 27. This comparison isolates when revising the order matters: the largest benefit is continued discovery after an initially successful priority.

C.3 SCORE INCREMENTS AND VERTEX LABELS

To distinguish remaining-cost scoring from an initially favorable guide, we compare empirical targeting, Greedy-margin and fixed priority on the first matrix of each 18-object family in Table 2. All three receive the same 294-label initial sample and potential outcomes in each of the original 20 repetitions, with the same confidence bounds, batches and 12000-total-label budget. Greedy-margin and fixed priority follow the rules in Appendix A; no score parameter changes. This comparison uses the total-budget convention of the independent-matrix study, separately from the confirmation-budget comparison in Table 5.

Table 6: Shared-information score comparison, 20 paired repetitions per matrix. Cost saving is control minus targeting in total labels; output gain is targeting minus control in true triangles at budget. Brackets give 95% paired Monte Carlo intervals; positive values favor targeting.

Matrix	Control	First-cost saving	Output gain
Low-rank 1	Greedy-margin	56.3 [21.4,91.1]	2.4 [−1.2,6.0]
Low-rank 1	Fixed priority	−7.5 [−24.2,9.2]	107.0 [92.9,121.0]
Misspecified 1	Greedy-margin	30.0 [0.8,59.2]	52.7 [48.4,57.0]
Misspecified 1	Fixed priority	37.5 [−36.9,111.9]	50.0 [43.7,56.3]

The scoring increment is most apparent in continued output under misspecification (Table 6). Targeting produces 66.85 triangles on average, versus 14.15 for Greedy-margin and 16.85 for fixed priority, and exceeds each control in every paired run. Their mean first costs are much closer: 477.75, 507.75 and 515.25 labels, respectively. In Low-rank, targeting and Greedy-margin both approach the full family, with mean output 216.0 and 213.6; their output difference is inconclusive. Fixed priority reports 109.05 despite a first cost of 554.0, close to targeting's 561.5 and below Greedy-margin's 617.75. This separates quickly finding one candidate from allocating evidence to complete later candidates.

We also examine sensitivity to canonical tie-breaking. For Sparse RPS, Block RPS and the two matrices above, we take the first 10 paired initial and confirmation samples and apply four independently drawn vertex permutations per matrix. Each permutation jointly relabels the probability matrix, the initially sampled pairs and their outcomes, and every potential confirmation outcome. Reversing a stored pair complements its probability and all binary wins. The comparison therefore preserves the sampled information and initial sampling coverage while changing the vertex indices used for ties. Algorithm parameters and budgets remain fixed.

Table 7: Equivalent vertex relabelings, 10 label repetitions per matrix and four fixed permutations. The permuted columns give the range of four means, not confidence intervals or independent environment variation. Costs include initial labels; reports count true triangles at total budget 12000.

	Targeting first cost		Targeting reports	
Matrix	Original	Permuted range	Original	Permuted range
Sparse RPS	947.5	947.5-947.5	1.0	1.0-1.0
Block RPS	1092.5	1087.5-1095.0	27.0	27.0-27.0
Low-rank 1	594.0	596.5-621.5	216.0	216.0-216.0
Misspecified 1	484.0	471.5-494.0	65.1	64.1-67.8

First-cost means move by at most 27.5 labels under these relabelings (Table 7). The two shared-information controls are also run on every relabeled 18-object sample. In Misspecified, targeting's mean output gains range from 46.8 to 58.5 over Greedy-margin and from 43.8 to 48.9 over fixed priority; every corresponding paired interval is positive. In Low-rank the gains are 0-0.6 and 68.2-91.0, respectively. First-cost differences can change sign across permutations, while the substantial continued-output gains persist. Thus fixed canonical labels do not account for those gains in this check.

For reproducibility, permutation $j \in \{0, 1, 2, 3\}$ in setting $s \in \{0, 1, 2, 3\}$ uses NumPy's default generator with seed $202609190000 + 100s + j$, in the table's setting order. Initial sample and

potential-stream seeds are those of the corresponding literature or independent-matrix design. The same permutation is applied to every repetition of its matrix. Each interval is the mean within-repetition difference plus or minus 1.96 standard errors, conditional on its matrix and permutation. No false reports or empty intersections occur in the 460 trajectories comprising the original-label comparison and relabeling check.

C.4 Near-threshold dynamic calibration and a four-cycle

This experiment measures the probability of any false report during an adaptive sequence of proposals. It uses four objects with $p_{01} = p_{12} = 0.8$, $p_{02} = 1 - q$, and all other probabilities $1/2$. For each $q \in \{0.55, 0.545\}$, there are 2500 independent runs. The initial candidate traverses vertices $(0, 1, 2)$; candidates $(0, 1, 3)$, $(0, 2, 3)$ and $(1, 2, 3)$ enter after rounds 20, 40 and 60. Both orientations are checked. At each of 100 rounds, the least-observed active pair is queried, with fixed-index ties and batch sizes cycling through 10, 25 and 40. Each of the six physical pairs has error $0.05/6$, and its interval is retained across proposals. For setting index $s = 0, 1$ and repetition $r = 0, \ldots, 2499$, the random seed is $202609270000 + 10000s + r$.

Each setting produces one false-report run in 2500; the one-sided exact 95% upper bound is 0.00190 in each case. One threshold-setting run encounters an empty intersection and suspends reporting, as specified by the algorithm. The error indicator records whether any invalid cycle was reported before suspension or budget exhaustion. This matches the run-level event in (1); Theorem 4.1 gives its general control.

The four-cycle example sets $p_{01} = p_{12} = p_{23} = 0.8$, $p_{03} = 0.2$, and $p_{02} = p_{13} = 0.5$. It has no 0.55-positive triangle. With seed 59151000, batches of 25 follow $0 \to 1 \to 2 \to 3 \to 0$ cyclically, using error $0.05/6$ per pair. The first report uses counts $(150, 150, 150, 125)$, totaling 575 labels. Its directed intervals are approximately

$$[0.6562, 0.9049],\ [0.6718, 0.9143],\ [0.5667, 0.8280],\ [0.6109, 0.8780].$$

The lower bound on minimum advantage is 0.0667. Proposing the longer cycle exposes cyclic structure that a triangle-only search cannot report.

C.5 Confidence boundaries on a fixed sampling path

This study isolates how the confidence construction affects the evidence needed to support a positive triangle. Its three directional probabilities all equal $0.55 + d$, for $d \in \{0.20, 0.10, 0.05, 0.02\}$. Each setting has 200 independent repetitions of three independent Bernoulli streams, shared by the Hoeffding, Jeffreys and stitched constructions. The reporting threshold is 0.55 and each physical pair receives error $\delta = 0.05/36$, matching the allocation in a nine-object registry; unused error remains unspent. The stitched parameters are fixed at $(\eta, \nu) = (2, 1.4)$.

The schedule visits edges 0, 1 and 2 cyclically in batches of 25, continuing to visit an edge even after its direction is supported. This keeps the observation path identical for all boundaries. There is no initial sample or adaptive sampling. Each construction intersects its own intervals over the inspected counts and stops when all three lower bounds exceed the threshold, or if any intersection is empty. Cost counts all scheduled labels through that first report. A cap of 100000 observations per edge gives 300000 total labels; a run without a report would receive this capped cost. Every construction reports the cycle in all 200 repetitions of every setting, and no empty intersection occurs before reporting. These positive alternatives assess efficiency, rather than false-report calibration.

Both likelihood mixing and stitching reduce cost relative to the Hoeffding union bound (Table 8). Jeffreys is faster than stitching in every setting, with positive paired difference intervals. Its advantage relative to the stitched mean narrows from 21.3% at gap 0.20 to 6.5% at gap 0.02. The iterated-logarithm rate describes sufficiently small gaps; finite-count constants and the likelihood's Bernoulli information determine the comparison over this grid. These observations support retaining the Jeffreys construction for the main sampling study while using profiles to compare their sufficient costs.

The study indexes gaps by $g = 0, \ldots, 3$, repetitions by $r = 0, \ldots, 199$ and edges by $e = 0, 1, 2$. Independent NumPy PCG64 streams use `SeedSequence([20260915,g,r,e])`; binomial increments with 25 trials generate the inspected Bernoulli prefixes. Mixture threshold crossings are

Table 8: Fixed-path first-report cost, with 200 paired repetitions per gap. Values are mean total labels; lower is better. The last column is stitched minus Jeffreys cost with a 95% paired Monte Carlo interval. All 2400 construction trajectories report before the cap.

Gap d	Hoeffding	Jeffreys	Stitched	Difference [95% interval]
0.20	902	474	603	129 [117, 141]
0.10	4228	2041	2472	432 [390, 473]
0.05	19121	8761	10170	1409 [1215, 1604]
0.02	133529	57341	61318	3977 [3350, 4605]

evaluated with the exact log-beta likelihood ratio; the condition is $S_m/m > \theta_0$ and $\log M_m(\theta_0) > \log(1/\delta)$. Agreement with root-inverted endpoints was checked on 340 count/win pairs, including endpoint counts. True-parameter inclusion at every inspected count also verified nonempty mixture intersections on these paths. Intervals in the last column use the paired sample standard error and the t_{199} quantile; rounding is applied only after computing means and differences.

D Application protocol and reproducibility

Section 3 gives a sampling protocol for fixed-model evaluation. It uses binary judgments, with an independent fair coin available to resolve ties. An application retaining fractional tie scores instead requires a confidence process for nonbinary observations.

A report identifies the cycle and comparison population, threshold and error level, initial sampling protocol, edge counts and intervals, total label cost and stopping reason. It qualifies a scalar ordering with supported matchup advantages. Resampling an existing log targets its empirical comparison population; inference for future tasks requires observations from that population. This protocol specifies the data needed to extend the synthetic experiments to model evaluation.

The supplied code implements the sampling policies, confidence bounds and reporting rules, and includes the comparison matrices, random seeds and numerical records. Separate scripts reproduce the literature comparison, shared-information controls, independent-matrix results, boundary comparison and dynamic reporting experiments.

D.1 Strength and style in the schematic

In the schematic in Figure 1, let q_i be version v_i's scalar strength and ψ_i its style angle. We combine a strength difference with an antisymmetric style interaction, a decomposition also used in the multidimensional Elo (mElo) model of Balduzzi et al. (2018). A two-dimensional interaction between vectors of fixed norm gives a sinusoidal offset after choosing the angular orientation. For amplitude $A > 0$ and logistic function $\sigma(x) = (1 + e^{-x})^{-1}$, the comparison probability is

$$\mathbb{P}(v_i \succ v_j) = \sigma\big(q_i - q_j + A\sin(\psi_i - \psi_j)\big).$$

The strength difference and style offset are in log-odds units; antisymmetry gives complementary probabilities on reversal.

For the illustrated descending path, choose $q_t = q_0 - td$ and $\psi_t = t\omega$ modulo 2π, with $0 < \omega < \pi$ and $0 < d < A\sin\omega$. The log-odds for $v_{t+1} \succ v_t$ are $A\sin\omega - d > 0$: a favorable style change offsets the loss of scalar strength at every local update. Against a fixed reference population with common strength and uniform style angle, rotation leaves mean win probability unchanged, while decreasing q_t strictly lowers it by monotonicity of σ. The cylinder thus separates local matchup success from strength within this constructed model. Comparisons restricted to a short recent arc can miss losses to older versions. The neural self-play experiments of Hernandez et al. (2022) provide the separate empirical example discussed in the introduction; the schematic itself specifies a possible mechanism rather than an observed training trajectory.

E FURTHER CONSEQUENCES AND CALIBRATION ALTERNATIVES

The simultaneous guarantee in Theorem 4.1 also supports alternative error allocations, initial sampling protocols and quantitative reports. The following statements describe these uses. Their proofs are collected in Appendix K, after the confidence-sequence construction and the main proofs.

E.1 ERROR ALLOCATION FOR A FIXED CANDIDATE FAMILY

Pairwise error allocation can account for the difficulty of resolving each required direction. Consider a fixed candidate family with R edges and gaps $d_e > 0$. The allocation-dependent part of its Hoeffding sample cap is proportional to

$$H(\delta) = \sum_e a_e \log(1/\delta_e), \qquad a_e = d_e^{-2}.$$

The ceiling and the terms $a_e \log(16/d_e^2)$ are separate from this objective.

Proposition E.1 (Error allocation for a fixed family)**.** *Let $A = \sum_e a_e$ and $w_e^* = a_e/A$. Among positive allocations summing to α, H is minimized by $\delta_e^* = \alpha w_e^*$. For any other $\delta_e = \alpha w_e$,*

$$H(\delta) - H(\delta^*) = A \sum_e w_e^* \log \frac{w_e^*}{w_e}.$$

The objective rewards granting harder edges a larger share of error: reducing a logarithmic penalty saves more observations when its inverse-square-gap coefficient is large. The relative-entropy identity quantifies the loss from any different allocation.

For data-derived weights $\widetilde{w}_e > 0$ with $\sum_e \widetilde{w}_e = 1$ and a fixed mixing constant $\xi \in (0, 1)$, the allocation $\delta_e = \alpha\{(1-\xi)\widetilde{w}_e + \xi/R\}$ sums to α and gives every edge error at least $\alpha\xi/R$. Conditional on the initial sample, the allocations are fixed. The validity theorem therefore applies even if the estimated gaps are biased or ordered incorrectly. The relative-entropy identity describes their effect on the deterministic leading bound when compared with the unknown oracle weights. An expanding edge family instead requires predictable spending, such as the registration-rank rule, or an explicitly maintained residual error budget.

E.2 EVIDENCE SHARING AND THE SUPPORTED GRAPH

Proposition E.2 (Overlapping cycles)**.** *Partition the objects into three sets A, B, C of size g. Suppose every cross-group edge has probability $\theta_0 + \Delta < 1$, with $\Delta > 0$, in the direction $A \to B \to C \to A$. There are g^3 positive triangles supported by $3g^2$ distinct edges. Supporting those edges supports every triangle. A procedure that collects separate, nonreused evidence for each triangle instead maintains $3g^3$ edge streams at the same individual precision.*

Proposition E.3 (A compact representation of valid reports)**.** *Let D_t contain the registered directed edges whose lower bounds exceed θ_0. On the simultaneous coverage event of Theorem 4.1, every directed cycle in D_t is positive. Given adjacency lists, one such cycle, or an acyclicity conclusion for D_t, can be obtained in $O(n + R)$ time and storage.*

The graph D_t compactly represents supported cycles while preserving their simultaneous guarantee. It is useful when listing every cycle would be expensive, since enumeration also depends on output size. An acyclic D_t means that no cycle is currently supported; the population may still contain one. The sampling implementation separately scans its explicit triangle candidates.

E.3 REUSING INITIAL OBSERVATIONS AND THE CONDITIONING LEVEL

Proposition E.4 (Pre-registration before an adaptive initial phase)**.** *For a known finite edge set, choose errors summing to at most α before observing any labels and register the confidence sequences at that time. An arbitrary first phase may then fit a predictor or choose candidates using revealed stream prefixes. Continuing the same edge processes with all subsequent observations satisfies* (1) *unconditionally from the start, with all initial observations included in the evidence.*

Pre-registration gives an unconditional guarantee across both phases. Fresh confirmation in Assumption 3.1 instead allows errors chosen from the initial sample and gives coverage conditional on that initial sample. These choices determine which labels enter the evidence: initial-data-dependent allocations require new confirmation outcomes, while including initial outcomes requires prior registration. Both protocols preserve the same edge processes across all later queries.

E.4 Cyclewise intersection-union calibration

Proposition E.5 (Shared-data cycle IUT)**.** *Fix a finite family $\mathcal{C}$ before confirmation, conditionally on any independent initial data. Assign $\alpha_C > 0$ with $\sum_C \alpha_C \le \alpha$. For each direction in C, use an anytime level-α_C one-sided test of $H_{0,a} : \mu_a \le \theta_0$. Report C only after every direction has rejected. Then the probability of any false cycle report, at any time, is at most α. Tests for different cycles may use the same observations.*

For $K = 2\binom{n}{3}$ predetermined oriented triangles, one choice is equal $\alpha_C = \alpha/K$. A two-sided edge confidence sequence at error α/K supplies each one-sided test through its directional lower bound. Equal cycle levels permit computing the endpoints once per edge and sharing all observations across candidates. The witness argument in Appendix K calibrates this Cycle-IUT implementation. Pairwise allocation instead controls a simultaneous edge event, supporting both later-proposed cycles and quantitative summaries.

E.5 Effect summaries from the common coverage event

For the minimum advantage defined in Section 4, write its interval endpoints as

$$[\underline{\gamma}_C(t), \overline{\gamma}_C(t)] = [\min_{a\in C} L_a(t) - \tfrac{1}{2}, \min_{a\in C} U_a(t) - \tfrac{1}{2}].$$

Taking minima preserves the componentwise inequalities, so

$$\underline{\gamma}_C(t) \le \gamma_C \le \overline{\gamma}_C(t) \quad \text{for every } C, t \tag{9}$$

on the simultaneous coverage event. Monotonicity of the logit gives a corresponding interval for a cycle's sum of log-odds.

Proposition E.6 (Simultaneous margins and log-odds)**.** *On the event in Theorem 4.1, for every time and every adaptively chosen directed simple cycle, both* (9) *and*

$$\sum_{a\in C} \operatorname{logit} L_a(t) \;\le\; \sum_{a\in C} \operatorname{logit} \mu_a \;\le\; \sum_{a\in C} \operatorname{logit} U_a(t) \tag{10}$$

hold, with extended-real endpoints. These statements and all cycle reports share the same error event.

The endpoint sums are well-defined on the common coverage event for interior probabilities: lower endpoints are below one and upper endpoints are above zero. A log-odds sum measures departure from additive strengths along a cycle; minimum advantage measures whether every direction favors its traversal. Both summaries use the accumulated adaptive evidence. Inference stops if a confidence intersection becomes empty.

Fixed-design conditional inference would use a separate sample: choose the target edges, counts and errors before collecting fresh outcomes. Adaptive stopping changes the distribution conditional on a realized count. For example, stopping at the first win fixes that count of wins at one. A joint guarantee across discovery and a later analysis therefore needs either the common coverage event or a preallocated error budget for the two procedures.

F Construction of the confidence sequences

The construction starts with a single Bernoulli(p) stream. Write Y_r for its rth binary outcome, $S_m = \sum_{r=1}^{m} Y_r$ for the win count, and $\mathcal{G}_m = \sigma(Y_1, \ldots, Y_m)$ for the information revealed by the first m outcomes. Appendix G then combines edge processes along an adaptive history. The subsequent proofs bound completion, first discovery and prescribed output.

F.1 THE HOEFFDING CONSTRUCTION

For m outcomes with S_m wins, the elementary Hoeffding interval has radius

$$r(m, \delta) = \sqrt{\frac{\log\{\pi^2 m^2/(3\delta)\}}{2m}}. \tag{11}$$

Its fixed-m failure bound is $6\delta/(\pi^2 m^2)$, whose sum is δ. For any fixed $m \geq 1$, Hoeffding's inequality and the radius in (11) give

$$\mathbb{P}_p\{|S_m/m - p| > r(m, \delta)\} \leq 2\exp\{-2mr(m, \delta)^2\} = \frac{6\delta}{\pi^2 m^2}.$$

Summing over m proves that every raw interval contains p with probability at least $1 - \delta$. Intersecting the intervals over observed own-edge sample sizes preserves this event. Inspection only at batch endpoints also preserves it, because those endpoints form a subset of the covered sample sizes.

F.2 THE BETA-MIXTURE CONSTRUCTION

For a fixed alternative $q \in (0, 1)$, define the likelihood ratio

$$L_m(q, p) = \left(\frac{q}{p}\right)^{S_m} \left(\frac{1-q}{1-p}\right)^{m-S_m}, \qquad L_0(q, p) = 1.$$

For $p \in (0, 1)$,

$$\mathbb{E}_p\{L_{m+1}(q, p) \mid \mathcal{G}_m\} = L_m(q, p)\left[p\frac{q}{p} + (1-p)\frac{1-q}{1-p}\right] = L_m(q, p).$$

Integrating against the Beta(β, β) density, with $\beta > 0$, gives

$$\begin{aligned} M_m(p) &= \int_0^1 L_m(q, p)\frac{q^{\beta-1}(1-q)^{\beta-1}}{B(\beta, \beta)}\, dq \\ &= \frac{B(S_m + \beta, m - S_m + \beta)}{B(\beta, \beta)p^{S_m}(1-p)^{m-S_m}}. \end{aligned}$$

Tonelli's theorem and the conditional identity show that $M_m(p)$ is a nonnegative mean-one martingale under the tested parameter p. Ville's inequality yields

$$\mathbb{P}_p\{\exists m \geq 0 : M_m(p) \geq 1/\delta\} \leq \delta.$$

Consequently, the running intersection of the sets in (3) is a $(1 - \delta)$ confidence sequence. Keeping equality in the set gives a closed interval and can only reduce its exclusion event.

For computation, write

$$f_m(p) = \log B(S_m + \beta, m - S_m + \beta) - \log B(\beta, \beta) - S_m \log p - (m - S_m)\log(1 - p) + \log \delta.$$

Its second derivative is $S_m/p^2 + (m - S_m)/(1 - p)^2 \geq 0$. For $0 < S_m < m$, its minimum is at S_m/m. At that point the integrated likelihood is no greater than the maximum likelihood, so $f_m(S_m/m) \leq \log \delta < 0$. The lower and upper roots can therefore be bracketed on $[0, S_m/m]$ and $[S_m/m, 1]$. If $S_m = 0$, the lower endpoint is zero and only the upper root is needed. If $S_m = m$, the upper endpoint is one. At $m = 0$ the interval is $[0, 1]$.

The implementation uses log-beta and stable $x \log y$ evaluations with $0 \log 0 = 0$. It solves each root to absolute tolerance 10^{-12}, then intersects with the previous interval. This avoids numerical underflow from multiplying a long sequence of likelihoods.

F.3 INTERIOR EXPANSION OF THE JEFFREYS MIXTURE

Write $\mathrm{kl}(x, y) = x \log(x/y) + (1 - x)\log\{(1 - x)/(1 - y)\}$ for Bernoulli relative entropy. Let $\widehat{p}_m = S_m/m \in [\iota, 1 - \iota]$ for fixed $0 < \iota < 1/2$, and set $\beta = 1/2$. The empirical win frequency stays in this interior interval in the following expansion:

$$\log M_m(p) = m\,\mathrm{kl}(\widehat{p}_m, p) - \tfrac{1}{2}\log m + \tfrac{1}{2}\log(2/\pi) + O(m^{-1}). \tag{12}$$

Uniform Stirling expansion of the gamma functions gives

$$\log B(m\widehat{p}_m + \tfrac{1}{2}, m(1-\widehat{p}_m) + \tfrac{1}{2}) - \log B(\tfrac{1}{2}, \tfrac{1}{2})$$
$$= m\{\widehat{p}_m \log \widehat{p}_m + (1-\widehat{p}_m)\log(1-\widehat{p}_m)\} - \tfrac{1}{2}\log m + \tfrac{1}{2}\log(2/\pi) + O(m^{-1}).$$

Subtracting $m\widehat{p}_m \log p + m(1-\widehat{p}_m)\log(1-p)$ proves (12). The remainder is uniform in $\widehat{p}_m$ on the stated interval and is independent of the tested p. For fixed δ, Taylor expansion of binary relative entropy around $p = \widehat{p}_m$ gives the raw endpoints

$$\widehat{p}_m \pm \sqrt{\frac{2\widehat{p}_m(1-\widehat{p}_m)}{m}\left\{\log(1/\delta) + \tfrac{1}{2}\log m + \tfrac{1}{2}\log(\pi/2)\right\}} + O\left(\frac{\log m}{m}\right).$$

The mixture's $\log m$ penalty is smaller than that of the stated Hoeffding union bound. Running intersections can further narrow the raw intervals. The expansion describes their interior width; endpoint computation uses the finite-sample beta-function likelihood ratio.

F.4 Polynomial stitching on the own-edge clock

For a sharper resolution profile, specialize standard polynomial stitching (Howard et al., 2021). Fix $\eta, \nu > 1$ before confirmation and write ζ for the Riemann zeta function. The empirical-mean radius is

$$r_{\mathrm{st}}(m,\delta) = \frac{\eta^{1/4} + \eta^{-1/4}}{2}\sqrt{\frac{\log\{2\zeta(\nu)/\delta\} + \nu\log(1+\log_\eta m)}{2m}}.$$

The argument below specifies its two-sided error and constants. Appendix H.1 shows that the first count after which $2r_{\mathrm{st}}(m,\delta) < d$ always holds is a profile of order $d^{-2}\{\log(1/\delta) + \log\log(\mathrm{e}/d)\}$. Substitution in (7) replaces $\log(1/d_e)$ in (8) by $\log\log(\mathrm{e}/d_e)$, with constants depending only on η, ν; the same improvement applies to (6). For epochs $\eta^k \le m < \eta^{k+1}$, $k \ge 0$, put

$$v_k = \eta^{k+1/2}, \qquad A_k = \log\frac{2\zeta(\nu)(k+1)^\nu}{\delta}, \qquad \lambda_k = \sqrt{8A_k/v_k}.$$

Writing $X_m = S_m - mp$, Hoeffding's exponential bound makes $\exp\{\lambda X_m - \lambda^2 m/8\}$ a nonnegative supermartingale for each fixed real λ. Ville's inequality therefore gives

$$\mathbb{P}_p\left\{\exists m \ge 1 : X_m \ge \frac{A_k}{\lambda_k} + \frac{\lambda_k m}{8}\right\} \le e^{-A_k}.$$

In epoch k, $m/v_k \in [\eta^{-1/2}, \eta^{1/2})$ and $k \le \log_\eta m$. Consequently, with $\kappa_\eta = (\eta^{1/4} + \eta^{-1/4})/2$,

$$\frac{A_k}{\lambda_k} + \frac{\lambda_k m}{8} = \sqrt{\frac{A_k m}{2}}\,\frac{\sqrt{v_k/m} + \sqrt{m/v_k}}{2}$$
$$\le \kappa_\eta\sqrt{\frac{A_k m}{2}} \le m r_{\mathrm{st}}(m,\delta).$$

Crossing the curved upper boundary in this epoch entails crossing the displayed linear boundary. Repeat for $-X_m$ and sum over both tails and all epochs. Since $2\sum_{k\ge 0} e^{-A_k} = \delta$,

$$\mathbb{P}_p\{\exists m \ge 1 : |S_m/m - p| > r_{\mathrm{st}}(m,\delta)\} \le \delta.$$

Clipping to $[0,1]$ and running intersection preserve coverage; at zero observations use $[0,1]$. The count m is the number of this edge's outcomes. Adaptive global rounds and batch endpoints only select prefixes of this event. Fixed $(\eta,\nu) = (2, 1.4)$ are used in Appendix C.5; selecting parameters from the same edge's outcomes would require accounting for that selection.

G Predictable registration and simultaneous validity

Let $\mathcal{F}_t$ contain the initial sample, decisions, registrations, internal randomization, and confirmation outcomes revealed through round t. A batch decision is $\mathcal{F}_{t-1}$-measurable. For the jth registered edge, let $\mathcal{H}_j$ denote the information immediately after choosing its identity and error allocation, but before revealing its first confirmation outcome. The index and allocation may depend on earlier observations on other edges.

Lemma G.1 (Validity at a random registration time)**.** *Under Assumption 3.1, the Hoeffding, beta-mixture and stitched constructions satisfy*

$$\mathbb{P}_p\{p_{e_j} \notin [L_{e_j}(t), U_{e_j}(t)] \text{ for some subsequent } t \mid \mathcal{H}_j\} \le \delta_j,$$

where $0 < \delta_j < 1$ *is measurable with respect to* $\mathcal{H}_j$*.*

Proof. Conditionally on registration information, the unrevealed stream has the fixed selected edge's Bernoulli distribution. For the Hoeffding construction, the countable union bound in Appendix F applies to its entire own-edge stream. For the mixture construction, the same conditional distribution gives the likelihood-ratio martingale and Ville bound. For stitching, the exponential-supermartingale proof in Appendix F.4 applies conditionally on registration. Adaptive queries reveal selected prefixes of these simultaneously covered streams. Running intersections and orientation reversal preserve coverage. □

Proof of Theorem 4.1. Let A_j be the event that the jth edge is registered and its confidence sequence subsequently excludes its parameter. Assign zero error to indices never registered. Lemma G.1, the tower property, and countable subadditivity give

$$\mathbb{P}_p\left(\bigcup_{j\ge1} A_j \,\middle|\, \mathcal{F}_0\right) \le \mathbb{E}_p\left(\sum_{j\ge1} \delta_j \,\middle|\, \mathcal{F}_0\right) \le \alpha.$$

Let $\mathcal{H}$ be the complementary event. At every time on $\mathcal{H}$, each stored confidence interval contains its physical edge probability. For a reverse direction, $[1 - U_e, 1 - L_e]$ contains $1 - p_e$. Every reported cycle therefore satisfies $\mu_a \ge L_a > \theta_0$ on each of its directed edges. This implication is simultaneous over the full observation history and every candidate the procedure considers. Thus any false report belongs to $\mathcal{H}^c$, proving (1) conditionally on $\mathcal{F}_0$.

An empty confidence intersection cannot occur on $\mathcal{H}$. Suspending further reports when it occurs is therefore consistent with the same error bound and prevents conflicting claims on an exceptional path. □

The role of fresh observations. An edge may be registered after the procedure observes many outcomes on other edges. Its own confirmation outcomes must still be unrevealed at registration. Selecting an edge using its unregistered outcomes and then treating those outcomes as a fresh confirmation prefix does not satisfy the conditional argument. Such labels can be used in the initial sample, followed by a new confirmation stream. Alternatively, an application can register every potentially queried edge before collecting any labels and retain their complete histories. In either implementation, returning to an edge reuses its original process and error allocation.

H FINITE-FAMILY RESOLUTION

Lemma H.1 (A sufficient edge count)**.** *For* $0 < d \le 1$ *and* $0 < \delta \le 1/2$*, the integer* $h = h(d, \delta)$ *in* (5) *satisfies* $2r(m, \delta) < d$ *for every* $m \ge h$*.*

Proof. Set $A = \log\{16/(\delta d^2)\}$. Since $A \ge \log 32 > 1$, $16A/d^2 \le h \le 17A/d^2$. Therefore

$$\begin{aligned}\log\frac{\pi^2 h^2}{3\delta} &\le \log\frac{\pi^2 17^2}{3} + 2\log A + 4\log(1/d) + \log(1/\delta)\\ &\le 2A + 2\log A + \log\frac{\pi^2 17^2}{3\cdot 16^2} \le 4A.\end{aligned}$$

The last inequality follows because $2A - 2\log A$ is increasing for $A > 1$ and exceeds the remaining constant at $A = \log 32$. Consequently,

$$\{2r(h,\delta)\}^2 = \frac{2\log\{\pi^2 h^2/(3\delta)\}}{h} \le \frac{8A}{16A/d^2} = \frac{d^2}{2} < d^2.$$

The function $r(m, \delta)^2$ is decreasing once $\log\{\pi^2 m^2/(3\delta)\} > 2$, which holds for $m \ge h$. The inequality thus holds at all later counts. □

Lemma H.2 (A conservative Jeffreys-mixture profile)**.** *For $\beta = 1/2$ and integer $m \geq 1$, every raw mixture interval is contained in the empirical-mean interval of radius*

$$\overline{r}_{\mathrm{mix}}(m, \delta) = \sqrt{\frac{\log\{\pi(m+1)/(2\delta)\}}{2m}} \leq r(m, \delta).$$

Consequently, (5) *is a resolution profile for the Jeffreys mixture.*

Proof. For integer wins s, write $\widehat{p} = s/m$ and $\mathcal{L}(q) = q^s(1-q)^{m-s}$. The Jeffreys density is at least $2/\pi$, so its integrated likelihood satisfies

$$J_m \geq \frac{2}{\pi}\int_0^1 \mathcal{L}(q)\, dq = \frac{2}{\pi(m+1)\binom{m}{s}} \geq \frac{2\mathcal{L}(\widehat{p})}{\pi(m+1)}.$$

The last inequality uses $\binom{m}{s}\mathcal{L}(\widehat{p}) \leq 1$, including endpoint counts with $0^0 = 1$. For p in the raw mixture interval, $J_m/\mathcal{L}(p) \leq 1/\delta$, implying

$$m\,\mathrm{kl}(\widehat{p}, p) = \log\frac{\mathcal{L}(\widehat{p})}{\mathcal{L}(p)} \leq \log\frac{\pi(m+1)}{2\delta}.$$

The inequality $\mathrm{kl}(x, y) \geq 2(x-y)^2$ gives the containment. The radius comparison follows from $3(m+1) \leq 2\pi m^2$ for integer $m \geq 1$. Lemma H.1 then gives the strict diameter bound at every $m \geq h(d, \delta)$. This envelope is used only in the cost proof; the experimental mixture uses the exact likelihood endpoints. □

Proof of Theorem 5.2. Work on the event $\mathcal{H}$ from the validity proof. At an inspected count $m \geq h_{\mathsf{B}}(d_e, \delta_e)$, the raw interval contains the true probability and has diameter less than d_e. A positive direction therefore has lower endpoint strictly above $\mu_a - d_e \geq \theta_0$, and a negative direction has upper endpoint strictly below $\mu_a + d_e \leq \theta_0$. Reversal preserves diameter, and running intersection can only strengthen these bounds.

Nestedness keeps every relevant direction resolved thereafter. An unresolved-edge policy cannot query the edge again. Its last batch overshoots the sufficient count by at most $b-1$, giving the individual cap $h_{\mathsf{B}}(d_e, \delta_e) + b - 1$.

Every unresolved candidate contains an unresolved direction. Otherwise all its directions would already be resolved, so either all would be positive and the candidate supported, or one would be negative and the candidate refuted. Hence the procedure always has an admissible query until it terminates. Each query advances an edge subject to a finite cap, proving both termination and the sum in (7).

For (8), suppose only edges in $U(\mathcal{C})$ are registered. Their ranks are then at most R, assigning unqueried edges the remaining ranks for evaluating the bound. Under (2), each error is at least $6\alpha/(\pi^2 R^2)$. Profile (5) decreases in δ, so substitution of this lower bound gives (8). Fixed errors α/R give the same order directly. The general bound (7) uses actual errors and also allows earlier registrations outside the family. □

Finite per-edge sample bounds ensure progress without a separate fairness condition: repeated queries resolve an edge, after which the policy must choose another unresolved direction. Positive gaps and a fixed family make the sum finite. At the threshold an edge need not resolve, and a growing family can keep introducing work. These upper bounds hold on the coverage event, of conditional probability at least $1-\alpha$; they do not bound the cost on exceptional paths.

H.1 The stitched resolution profile

Define $h_{\mathrm{st}}(d, \delta)$ as the smallest positive integer n such that $2r_{\mathrm{st}}(m, \delta) < d$ for every integer $m \geq n$. The raw diameter is at most $2r_{\mathrm{st}}$, so this is a resolution profile whenever finite. The tail condition avoids requiring monotonicity at small counts.

Lemma H.3 (Stitched sufficient counts)**.** *For fixed $\eta, \nu > 1$, h_{st} is finite and nonincreasing in both d and δ. Uniformly over $0 < d \leq 1$ and $0 < \delta \leq 1/2$,*

$$h_{\mathrm{st}}(d, \delta) = O_{\eta,\nu}\left(d^{-2}\{\log(1/\delta) + \log\log(\mathrm{e}/d)\}\right).$$

Proof. Put $A = \log\{2\zeta(\nu)/\delta\} > 1$ and $F(x) = A + \nu \log(1 + \log x/\log \eta)$ for $x \geq 1$. With $\kappa_\eta = (\eta^{1/4} + \eta^{-1/4})/2$, the squared diameter envelope is $2\kappa_\eta^2 F(x)/x$, which tends to zero. The profile is therefore finite, and its defining tail sets give both monotonicities.

For its order, set $D = \log(\mathrm{e}/d) \geq 1$, $L = A + \nu \log D > 1$, and $n = \lceil Cd^{-2}L \rceil$ for a constant $C \geq 1$ depending only on η, ν. Since $n \leq 2Cd^{-2}L$,

$$\log n \leq \log(2C) + 2(D-1) + \log L.$$

Using $D, L \geq 1$ and $\log L \leq L - 1$, there is a constant c_η such that

$$\log(1 + \log n/\log \eta) \leq \log D + \log L + c_\eta + \log(1 + \log(2C)),$$
$$F(n) \leq (1+\nu)L + \nu c_\eta + \nu \log(1 + \log(2C)).$$

A sufficiently large fixed C makes $2\kappa_\eta^2 F(n) < nd^2$, since C grows faster than $\log \log C$. Increase C if needed so $n \geq e^\nu$. For $x \geq e^\nu$, $xF'(x) = \nu/(\log \eta + \log x) < 1 < F(x)$, so $F(x)/x$ decreases. The strict diameter bound thus holds for every $m \geq n$. Finally $A = \log(1/\delta) + O_\nu(1)$, proving the rate. □

Substituting this profile in (7) gives the stitched family rate in the main text. Fixed errors α/R yield $\log(R/\alpha)$; registration-rank errors with no outside registrations are at least $6\alpha/(\pi^2 R^2)$ and yield the same logarithmic order. Substitution in (6) gives the corresponding prefix bound.

I First discovery under a initially fixed priority order

Proof of Theorem 5.1. Condition on the initial sample and work on the simultaneous event $\mathcal{H}$. Every supported direction is then truly above θ_0, and every refuted direction is truly below it. Before the first report, a candidate preceding C_{j_*} can therefore end only by refutation. In its predetermined direction order, a refutation occurs no later than resolving the first true negative direction. Earlier refutation from a previously shared edge can only shorten that visit. Hence all directions queried while visiting candidate $j < j_*$ lie in F_j. Directions needed for C_{j_*} lie in F_{j_*} by definition. If shared evidence supports any other cycle sooner, stopping at that report only reduces the count.

No edge outside E_* is queried before the first report. At an inspected count of at least $h_{\mathsf{B}}(d_e^*, \delta_e)$, the raw interval on edge $e \in E_*$ has diameter less than d_e^* and contains the truth. Hence all its directions occurring in these prefixes are resolved. Nestedness prevents further queries on those directions. The last batch overshoots by at most $b - 1$, giving the cap $h_{\mathsf{B}}(d_e^*, \delta_e) + b - 1$. Each incomplete visit contains an unresolved direction, so the finite sum prevents stalling before a report. Summing proves (6). □

The sets F_j and E_* describe the true environment in the bound. The policy obtains its order from the initial sample. Appendix A specifies the initially fixed ordering and the use of settled shared directions. The bound starts with zero confirmation counts, while pre-registered reuse of initial observations follows the protocol in Appendix E.3.

J An information bound for a prescribed positive family

Write $\mathrm{kl}(x, y) = x \log(x/y) + (1-x)\log\{(1-x)/(1-y)\}$ for Bernoulli relative entropy.

Theorem J.1 (Cost of a prescribed positive family)**.** *Fix an environment p and a nonempty finite family $\mathcal{C}_+$ of its positive cycles. Consider a procedure satisfying* (1) *uniformly over unrestricted edge probabilities and reporting every cycle in $\mathcal{C}_+$ with probability at least $1 - \alpha$ under p. Let N_e^{tot} count all observed labels on edge e, including informative initial labels. Write μ_e for the win probability in its required direction. If non-label side information has the same law under the single-edge alternatives below, then*

$$\mathbb{E}_p \sum_e N_e^{\mathrm{tot}} \geq \sum_{e \in U(\mathcal{C}_+)} \frac{\mathrm{kl}(1-\alpha, \alpha)}{\mathrm{kl}(\mu_e, \theta_0)}. \tag{13}$$

Lemma J.2 (Adaptive change of measure)**.** *Let p and q be two Bernoulli edge environments. Suppose the algorithm uses the same decision kernels in both and its initial non-label side information has the same distribution. For a finite transcript with T label opportunities, let $N_e^{\mathrm{tot}}(T)$ count labels observed on edge e. For every transcript-measurable event A_T,*

$$\sum_e \mathbb{E}_p N_e^{\mathrm{tot}}(T)\ \mathrm{kl}(p_e, q_e) \geq \mathrm{kl}\{\mathbb{P}_p(A_T), \mathbb{P}_q(A_T)\}. \tag{14}$$

Informative initial labels are included among the observations.

Proof. Append a no-observation symbol after stopping, so both environments define transcripts of the same length. Conditional on the preceding transcript, an edge choice, stopping decision, or randomization has the same kernel under p and q and contributes zero conditional relative entropy. A label from edge e contributes $\mathrm{kl}(p_e, q_e)$. The chain rule therefore identifies the left-hand side as transcript relative entropy. Applying data processing to the indicator of A_T proves (14). Predetermined batches can be expanded into their individual labels. □

Proof of Theorem J.1. For a required physical edge e, choose a cycle $C_e \in \mathcal{C}_+$ containing it. All positive cycles containing e use the same direction: opposing directions would require both $p_e > \theta_0$ and $1 - p_e > \theta_0$, impossible for $\theta_0 \geq 1/2$. Write μ_e for the required directional probability.

Construct $q^{(e)}$ by changing only this directional probability to θ_0. Under $q^{(e)}$, the cycle C_e is invalid. Let A_T be the event that C_e has been reported by the Tth label opportunity, and let $A = \bigcup_T A_T$. Completeness under p implies $\mathbb{P}_p(A) \geq 1-\alpha$. Uniform soundness implies $\mathbb{P}_{q^{(e)}}(A) \leq \alpha$. Lemma J.2 gives

$$\mathbb{E}_p N_e^{\mathrm{tot}}(T)\ \mathrm{kl}(\mu_e, \theta_0) \geq \mathrm{kl}\{\mathbb{P}_p(A_T), \mathbb{P}_{q^{(e)}}(A_T)\}.$$

Binary relative entropy is unchanged when both arguments are reversed, so this expression is valid in either storage orientation. Letting $T \to \infty$, monotone convergence of counts and lower semicontinuity of relative entropy yield

$$\mathbb{E}_p N_e^{\mathrm{tot}}\ \mathrm{kl}(\mu_e, \theta_0) \geq \mathrm{kl}(1-\alpha, \alpha).$$

For the last step, binary relative entropy is minimized over $x \geq 1-\alpha$ and $y \leq \alpha$ at $(1-\alpha, \alpha)$, since $\alpha < 1/2$. Summing over the distinct edges of $\mathcal{C}_+$ proves (13). If an expected count is infinite, the inequality is immediate. □

The alternatives identify the evidence needed on each required pair, even when a model guides the queries. This conclusion uses the unrestricted probability class in Theorem J.1. A structural model can exclude these alternatives, and an objective requiring only one of several possible discoveries can choose an easier cycle. Both changes alter the lower-bound problem.

K Proofs of the secondary propositions

Proof of Proposition E.1. For $\delta_e > 0$ with $\sum_e \delta_e = \alpha$, write $w_e = \delta_e/\alpha$. Since $a_e = A w_e^*$,

$$\begin{aligned} H(\delta) - H(\delta^*) &= \sum_e a_e \{\log(1/(\alpha w_e)) - \log(1/(\alpha w_e^*))\} \\ &= A \sum_e w_e^* \log(w_e^*/w_e) \geq 0. \end{aligned}$$

Relative entropy vanishes exactly when $w = w^*$, proving the minimum and the identity. Alternatively, the multiplier equation $-a_e/\delta_e + \lambda = 0$ gives $\delta_e = \alpha a_e/A$. The result concerns the smooth allocation-dependent term of the bound. Integer ceilings can change an exact finite-count optimizer. □

Proof of Proposition E.2. Choosing one vertex from each group gives g^3 triangles. Each of the three ordered group pairs contributes g^2 edges. Once all $3g^2$ directional lower bounds exceed θ_0, every triangle is supported. Separate collection for each triangle uses three streams per triangle, totaling $3g^3$. The ratio is g at equal precision per stream; calibrated sample costs also depend on the error allocations. □

Proof of Proposition E.3. On the simultaneous coverage event, each edge included in D_t has true directional probability above θ_0, so every cycle satisfies the required conjunction. At most one orientation of a physical edge can be included: simultaneous inclusion would imply both $p_e > \theta_0$ and $1 - p_e > \theta_0$. Thus D_t has at most R arcs.

Run depth-first search, marking each vertex unvisited, active, or completed, and retain its parent in the search forest. Encountering an arc to an active ancestor closes a directed cycle, recovered by tracing the parent chain. If no such arc is encountered, the graph is acyclic. Each vertex is entered once and each arc examined once, giving $O(n + R)$ time. Adjacency lists, the vertex states, parents, and search stack use $O(n + R)$ storage. □

Proof of Proposition E.4. Apply the simultaneous validity argument with the initial sigma field containing only the fixed design and parameter-independent randomization, rather than the subsequently observed initial sample. Each edge is valid on its entire own-edge sequence. Adaptive first-phase sampling and later sampling reveal prefixes of the same sequences, so the union event covers both phases and all later reports. No reset or additional error allocation occurs at the phase transition. □

Proof of Proposition E.5. Fix the parameter vector. For any invalid C, choose a deterministic witness direction a_C with $\mu_{a_C} \leq \theta_0$. The event of ever reporting C is contained in the event that the test on a_C ever rejects, whose probability is at most α_C. Summing over invalid cycles bounds the probability of any false report by $\sum_C \alpha_C \leq \alpha$. This argument requires neither independence among directions nor independence among cyclewise tests. It also shows why there is no additional division by $|C|$. □

Proof of Proposition E.6. For each direction, $L_a \leq \mu_a \leq U_a$ holds simultaneously under physical-edge reversal. Minima preserve componentwise inequalities, giving the margin result. The logit is increasing on $(0, 1)$ and has limits $-\infty, +\infty$ at $0, 1$; summing gives (10). Since every edge statement is already simultaneous in time, the argument applies pathwise to any chosen cycle or collection. No additional union bound is required. □